e-ISSN: 2636-8579

JHSM

Journal of Health Sciences and Medicine

VOLUME: 8 ISSUE: 5 YEAR: 2025

MHA
MEDIHEALTH ACADEMY

FRANCHISE OWNER

MediHealth Academy Publishing

LAYOUT EDITOR

Şerife KUTLU
Engineer, MediHealth Academy Publishing, Ankara, Turkiye

CORRESPONDENCE ADDRESS

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

ARTICLE SUBMISSION ADDRESS

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

CONTENTS

Original Articles

CONTENTS

Original Articles

CONTENTS

Original Articles

Erratum

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1704036
J Health Sci Med. 2025;8(5):739-746

A two-decade bibliometric exploration of AI applications in obstetrics and gynecology (2005-2024)

Fatma Kılıç Hamzaoğlu
Department of Obstetrics and Gynecology, Faculty of Medicine, Necmettin Erbakan University, Konya, Turkiye

Cite this article as: Kılıç Hamzaoğlu F. A two-decade bibliometric exploration of AI applications in obstetrics and gynecology (2005-2024). *J Health Sci Med.* 2025;8(5):739-746.

Received: 22.05.2025 ◆ **Accepted**: 14.07.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: Artificial Intelligence (AI) technologies have significantly impacted obstetrics and gynecology (Obs&Gyn), particularly in diagnostics, treatment, and patient care. This study aims to conduct a bibliometric analysis of AI-related research in Obs&Gyn published between January 1, 2005, and December 31, 2024. The main objectives are to explore publication trends, leading contributors, research themes, collaboration patterns, and emerging technologies.

Methods: A total of 959 publications were retrieved from the Web of Science Core Collection using the keyword "artificial intelligence" and filtered by the Obs&Gyn category. VOSviewer software was used to map co-authorship, keyword co-occurrence, bibliographic coupling, and geographic collaborations.

Results: AI research in Obs&Gyn increased notably after 2017. North America and Europe led in publication output, with Asia also showing strong contributions. Prominent themes included machine learning, deep learning, IVF, diagnostic imaging, and robotic surgery. Collaboration networks revealed strong institutional and international partnerships.

Conclusion: This study underscores the transformative potential of AI in Obs&Gyn and highlights critical research areas, key contributors, and collaboration dynamics. Findings provide a foundation for future research, emphasizing the need for inclusivity, ethical AI adoption, and addressing global healthcare disparities in Obs&Gyn.

Keywords: Artificial Intelligence, obstetrics, gynecology, machine learning, Deep Learning

INTRODUCTION

The rapid evolution of artificial intelligence (AI) technologies has transformed various fields, including healthcare. In obstetrics and gynecology (Obs&Gyn), AI applications have emerged as powerful tools to enhance diagnostic precision, optimize treatment strategies, and address pressing challenges in maternal and fetal health. From predictive models for embryo viability to advancements in robotic surgery, the integration of AI has opened new avenues for improving outcomes and streamlining clinical workflows.[1,2]

Bibliometric analyses provide a systematic method to evaluate the progression and impact of research within a field, offering insights into publication trends, influential contributors, and emerging hotspots. These analyses are particularly relevant in Obs&Gyn, where the intersection of AI with clinical practice has gained substantial momentum in recent years. For instance, the proliferation of research on robotic surgery, as highlighted by Xiao et al.[1] and Levin et al.,[2] underscores the transformative potential of AI in enhancing surgical precision and reducing patient recovery times. Additionally, the integration of machine learning techniques in predictive models for obstetric complications has demonstrated promising results in improving maternal and neonatal care.[3]

A critical aspect of AI's application in Obs&Gyn is its ability to address resource disparities and promote equitable access to healthcare. Industry 4.0 technologies, as explored by Sibanda et al.,[4] have been pivotal in optimizing maternal health care systems in resource-limited settings. Furthermore, the use of natural language processing tools, such as ChatGPT, has sparked discussions around AI's potential to support decision-making and patient communication in clinical settings.[5,6]

Despite these advancements, the global distribution of AI research in Obs&Gyn remains uneven, with contributions concentrated in certain regions, particularly North America and Europe.[5,7] Emerging players such as China and other Asian countries are progressively expanding their influence in this domain, as evidenced by recent bibliometric studies.[8] However, contributions from Africa and other underrepresented regions, while growing, remain limited, highlighting the need for a more inclusive research landscape.[4]

Existing bibliometric studies have also emphasized the interdisciplinary nature of AI research in Obs&Gyn. Keywords such as "machine learning," "deep learning," "IVF," and "ultrasound" consistently appear as dominant themes, reflecting the diverse applications of AI across reproductive

Corresponding Author: Fatma Kılıç Hamzaoğlu, dr.fatmakilich@gmail.com

health and diagnostics.[9,10] Moreover, AI's potential to transform clinical workflows through emerging technologies, such as radiomics and digital pathology, underscores its growing relevance in Obs&Gyn.[8,11]

The primary aim of this study is to conduct a comprehensive bibliometric analysis of research trends in AI within the domain of Obs&Gyn. This study seeks to identify the key research areas, influential contributors, emerging technologies, and collaboration networks that are shaping the integration of AI in Obs&Gyn. By examining the temporal distribution of publications, most-cited works, prominent institutions, and keyword trends, this research provides a detailed understanding of how AI is transforming clinical practices, diagnostics, and research in this critical healthcare field.

The significance of this study is underscored by the increasing reliance on AI technologies in healthcare. As maternal and fetal health are central to population well-being, the application of AI in Obs&Gyn has the potential to revolutionize patient care through predictive analytics, personalized treatment plans, and enhanced diagnostic accuracy. Furthermore, as global challenges such as resource limitations and unequal access to healthcare persist, understanding the role of AI in addressing these disparities is vital. This study also highlights the interdisciplinary and collaborative nature of AI research, offering insights into how advancements in computer science, medicine, and engineering converge to solve complex healthcare challenges.

By providing a granular view of the state of AI research in Obs&Gyn, this study contributes to the literature by identifying opportunities for future research, fostering collaboration among stakeholders, and supporting evidence-based decision-making in healthcare policy and practice.

METHODS

Ethics

Our study is a bibliometric analysis based solely on previously published data obtained from public databases. It does not involve human participants, personal data, or any clinical interventions. Therefore, ethical approval is not required for this type of research. All procedures were carried out in accordance with the ethical rules and the principles.

Study Design and Data Collection

This study conducted a bibliometric analysis of the literature on "artificial intelligence" (AI) published between January 1, 2005, and December 31, 2024. A total of 959 publications were evaluated, with data retrieved from the Web of Science (WoS) Core Collection database. The keyword "artificial intelligence" was used as the primary search term, applying the "topic" filter to include titles, abstracts, author keywords, and Keywords Plus®. Publications categorized under "obstetrics gynecology" in the Web of Science Categories were selected to ensure relevance to the intersection of AI and this specific medical domain.

The decision to use "artificial intelligence" as the sole keyword and focus on the "obstetrics gynecology" category was based on the need for specificity and precision in capturing relevant literature. This approach ensured a focused dataset while eliminating irrelevant results, maximizing both the relevance and comprehensiveness of the study.

Data export: Data from the selected papers were exported and saved in Microsoft Excel 2019 for quantitative analysis and EndNote Desktop for citation management. The exported dataset included various parameters such as authors, organizations, countries/regions, keywords, titles, abstracts, publication years, source journals along with their impact factors (as per the Journal Citation Report, 2022), and citation counts. This structured dataset enabled qualitative and quantitative analyses to uncover trends, patterns, and key contributions in AI research within obstetrics and gynecology.

Data analysis: The bibliometric data were systematically analyzed to identify trends in publication volume over time, citation patterns, and the impact of specific journals, authors, and institutions. Authorship and collaboration patterns were also examined, focusing on the affiliations and contributions of leading researchers as well as international collaboration networks. Keyword analysis was performed to investigate co-occurrence patterns, revealing dominant themes and emerging trends. Additionally, the geographical distribution of research was mapped to understand contributions from different countries and regions. Bibliographic coupling and co-citation networks were analyzed to uncover relationships between publications, authors, and references.

Visualization: To visualize relationships and trends within the dataset, VOSviewer software (version 1.6.11, Leiden University, The Hague, Netherlands) was utilized. Co-authorship networks were generated to map collaborative relationships among authors, countries, and organizations, where node size indicated the frequency of co-authorship, and the thickness of connecting lines reflected the strength of collaboration. Keyword co-occurrence networks were created to highlight research hotspots, with larger nodes representing higher keyword frequency and clusters depicted in distinct colors to illustrate thematic groupings. Bibliographic coupling maps revealed connections based on shared references, while co-citation networks identified foundational works cited together.

Visualization maps were further refined using three distinct types: network visualization maps, density visualization maps, and overlay visualization maps. The network visualization maps illustrated clusters of keywords and their interrelationships, with larger nodes indicating higher frequency and thicker lines representing stronger connections. Density visualization maps used a gradient color scheme to depict the frequency of keyword occurrences, with red denoting the highest frequency. Overlay visualization maps captured the temporal evolution of keywords, using colors to indicate average publication years (APY), enabling an assessment of the novelty and emergence of specific research themes. The parameters for visualization in VOSviewer were set to optimize clarity and interpretability, with the method set to association strength, attraction set to 2, repulsion set to 0, resolution at 1, and minimum cluster size of 1.

Research Questions

- What are the publication trends over time in AI research within Obs&Gyn?
- Which journals, authors, and institutions are the leading contributors to AI research in Obs&Gyn, and what is their impact?
- What are the most frequently cited articles in AI research for Obs&Gyn, and what are their significant contributions?
- What are the dominant themes and keywords in AI research within Obs&Gyn, and what do they reveal about emerging trends?
- What are the patterns of collaboration among authors, institutions, and countries in AI research for Obs&Gyn?
- What is the geographic distribution of AI research in Obs&Gyn, and which regions are emerging as leaders or contributors?
- How can AI address current challenges in Obs&Gyn, including resource limitations, diagnostic inaccuracies, and healthcare inequities?

RESULTS

Analysis of Articles by Year

Figure 1 presents the annual distribution of studies in the dataset created using Web of Science data.

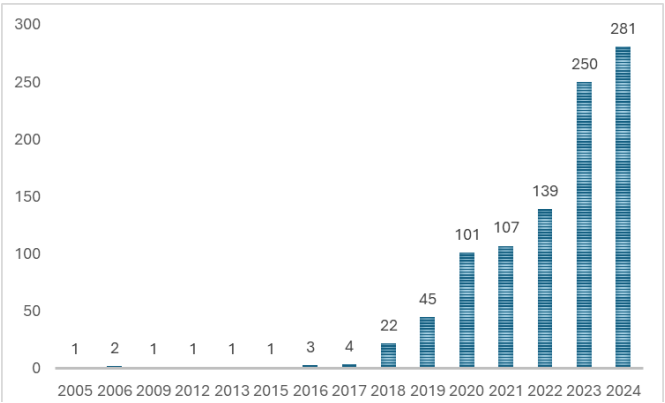

Figure 1. Annual distribution of articles on "artificial intelligence"

When examining the annual distribution of articles on "artificial intelligence" shown in Figure 1, it is evident that research activities in this field were initially limited, with some periods showing no publications. After 2017, academic interest in this area began to rise, and the number of articles increased steadily. This growth can be directly linked to the application of AI technologies across various disciplines, the development of new algorithms, and the popularization of subfields such as machine learning. A consistent upward trend was recorded between 2018 and 2024, with 2024 standing out as the year with the highest number of publications. This reflects the rapid technological advancements in AI, supported by the development of big data and computational power, which have driven academic work in this direction.

Journals with the Most Publications on the Topic

The number of articles related to "artificial intelligence" published in various journals, as recorded in Web of Science, is presented in Figure 2.

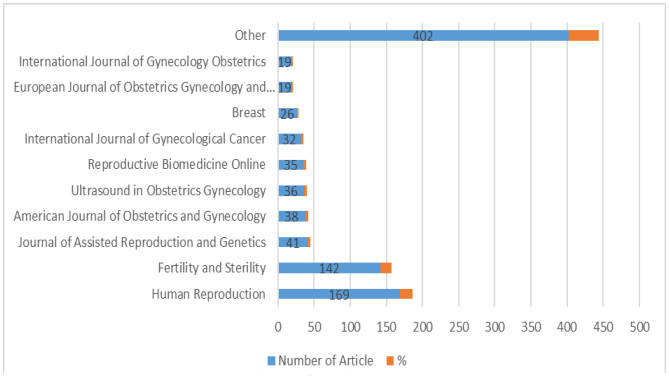

Figure 2. Journal names, number of publications, and percentage distribution of articles

Figure 2 ranks journals publishing on "Artificial Intelligence" by the number of articles. Human Reproduction leads with 169 articles, demonstrating the significant role of AI applications in reproductive health. This is followed by Fertility and Sterility with 142 articles. These two journals have made substantial contributions to the literature by publishing a large portion of the studies in this field. The Journal of Assisted Reproduction and Genetics ranks third with 41 articles, while American Journal of Obstetrics and Gynecology and Ultrasound in Obstetrics and Gynecology follow with 38 and 36 articles, respectively. These publication counts highlight the widespread use of AI-based methods in critical areas such as gynecology and reproductive health.

Additionally, Reproductive Biomedicine Online published 35 articles, International Journal of Gynecological Cancer 32 articles, and Breast 26 articles, showcasing the focus on AI applications in various medical subfields. The European Journal of Obstetrics Gynecology and Reproductive Biology and International Journal of Gynecology Obstetrics each published 19 articles, appearing at the lower end of the rankings. Lastly, the "other" category includes 402 articles, indicating that studies on AI have been published across a broader range of journals. This high number of articles in other journals underscores the interdisciplinary nature of AI research and its wide range of applications.

Authors, Article Titles, Journals, Publication Dates, and Citation Counts of the Most Cited Publications

Table 1 presents the authors, article titles, journals, publication dates, and citation counts for the most-cited publications on "artificial intelligence" in the Web of Science database.

The most-cited article, authored by Tran et al.[12] in 2019, focuses on predictive AI use in reproductive health and has received 186 citations. The second-ranked article by Tagliafico et al.[13] discusses the use of AI in breast cancer diagnosis, with 167 citations. VerMilyea et al.'s[14] 2020 article, which developed an AI-based model for predicting embryo viability, ranks third with 158 citations. Carter et al.'s[15] article from the same year

Table 1. Details of the most cited articles on "artificial intelligence"

No	Authors	Article title	Journal name	Year	Citation count
1	Tran et al.[12]	Deep learning as a predictive tool for fetal heart pregnancy following time-lapse incubation and blastocyst transfer	Human Reproduction	2019	186
2	Tagliafico et al.[13]	Overview of radiomics in breast cancer diagnosis and prognostication	Breast	2020	167
3	VerMilyea et al.[14]	Development of an artificial intelligence-based assessment model for prediction of embryo viability using static images captured by optical light microscopy during IVF	Human Reproduction	2020	158
4	Carter et al.[15]	The ethical, legal and social implications of using artificial intelligence systems in breast cancer care	Breast	2020	123
5	Grünebaum et al.[16]	The exciting potential for ChatGPT in obstetrics and gynecology	American Journal of Obstetrics and Gynecology	2023	112

addresses the ethical and social implications of AI in breast cancer care, with 123 citations. Lastly, Grünebaum et al.'s[16] 2023 article, which explores the use of ChatGPT in obstetrics and gynecology, received 112 citations. These highly cited studies highlight the significant interest in AI's diverse applications in the medical field.

Publication Statistics for the Most Cited Institutions

The institutions where the authors of articles on "artificial intelligence" in the Web of Science database are affiliated, along with the number of articles published and their citation counts, are presented in Table 2.

Table 2. Distribution of the most cited institutions and publications based on WoS data

No	Institution	Number of publications	Citation count
1	The University of Sydney	11	409
2	The University of Adelaide	17	369
3	Harvard Medical School	34	355
4	University of Oxford	17	340
5	Weill Cornell Medicine	20	280

Table 2 provides information on the institutions with the highest citation counts for articles related to "artificial intelligence" in the Web of Science database, along with the number of articles published by these institutions and their corresponding citation counts. Ranked first, the University of Sydney has published 11 articles, receiving a total of 409 citations, making it the institution with the highest citation count. This indicates that the university's research in the field of artificial intelligence has had a significant academic impact.

In second place, the University of Adelaide has published 17 articles with 369 citations, demonstrating a notable citation performance alongside a high number of publications. Harvard Medical School, which holds the highest number of publications with 34 articles, ranks third in citations with a total of 355. This suggests that while the institution has a broad publication output, its citation-per-article ratio is comparatively lower than other institutions.

The University of Oxford, with 17 articles and 340 citations, is ranked fourth, highlighting its significant academic contribution to the field of artificial intelligence. In fifth place, Weill Cornell Medicine has published 20 articles, accumulating a total of 280 citations. This result indicates that the institution's publications in this field have been recognized and valued by the academic community, making meaningful contributions to the area.

Overall, the table reveals that research on artificial intelligence is predominantly conducted by certain universities and medical schools, whose studies have had substantial academic influence and resonated across various disciplines.

Trends in the Use of Keywords

The most frequently used keywords and their frequencies in articles on "Artificial Intelligence" in the Web of Science database are presented in Figure 3.

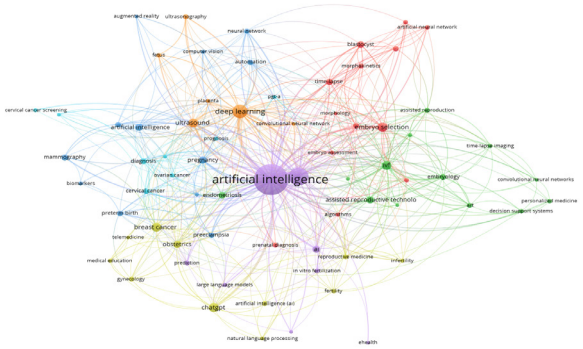

Figure 3. Co-occurring keywords and their frequencies

A bibliometric analysis using Vosviewer software was conducted, setting a minimum threshold of five occurrences for keyword inclusion. This threshold ensures that only more frequently used and semantically significant terms are analyzed. While a total of 1,445 unique keywords were identified during the analysis, only 71 met the threshold and were included in the study. This method demonstrates that only these 71 keywords and their relationships were examined in the analysis.

As a result of the analysis, the most frequently used and strongly interconnected keywords in a specific research field were identified. Additionally, a total of 502 connections and four distinct clusters among the keywords were detected. The findings provide a detailed understanding of the relationships between keywords in the research field. This type of analysis helps to better understand the terminology of the field and offers valuable insights for identifying key concepts for future research.

Figure 1 shows the most frequently used keywords and their frequencies in academic publications on "artificial intelligence." According to the data, the most frequently used keyword is "artificial intelligence," which appears 337 times. This demonstrates that the primary focus of the research area is artificial intelligence and emphasizes the critical role of this term in the literature. The second most frequently used keyword is "machine learning," appearing 114 times, indicating that machine learning is an important subdiscipline within AI research. The third-ranked keyword, "deep learning," appears 67 times, highlighting its significance as a major research topic in studies related to AI.

Other keywords used less frequently but still notable in the literature include "embryo selection" (30 occurrences), "ChatGPT" (30 occurrences), "IVF" (29 occurrences), and "ultrasound" (28 occurrences). These terms indicate a strong interest in various applications of AI in the medical field. Notably, areas such as embryo selection and IVF appear to be among the priority topics in AI-based decision support systems. This data clearly reflects the main research areas focused on in the literature on "Artificial Intelligence" and their interrelationships. Additionally, the strategic selection of keywords is shown to increase research visibility and contribute to a better understanding of the relationships within the literature.

Inter-Institutional Collaboration Analysis

The affiliations of authors contributing to articles on "artificial intelligence" in the Web of Science database and the collaboration networks between these institutions were analyzed. The findings are visualized in Figure 4.

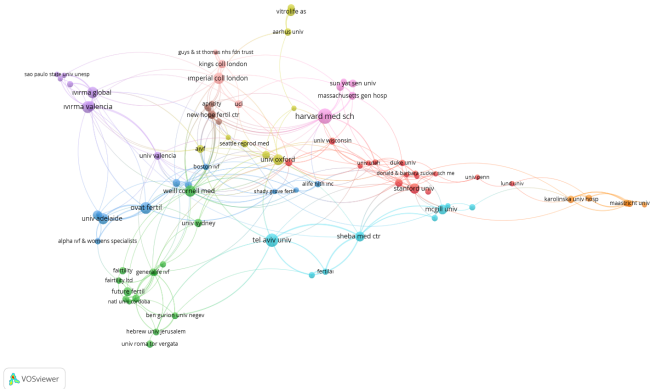

Figure 4. Bibliometric network of institutional collaboration

The bibliometric analysis using Vosviewer software illustrates the collaborations among institutions. In the visualization, the size of the circles represents the publication productivity of the institutions, and the thickness of the lines connecting them indicates the strength of their collaborative relationships. Different colors distinguish thematic or regional groups, while the connections symbolize the collaborations between institutions. The thickness of these lines reflects the intensity of collaboration.

In the network, Harvard Medical School is positioned centrally with 16 collaborations, highlighting its pivotal role in the global network of AI research. This finding indicates that Harvard Medical School has a broad collaboration network and has established strong relationships with other institutions in the field. Stanford University, with 13 collaborations, ranks second, followed by University of Valencia, which has 12 collaborations and demonstrates significant collaborative activity. Other notable institutions include Sheba Medical Center (9 collaborations) and University of Oxford (8 collaborations), both of which play important roles within the network.

Overall, the observed connections and clustering structures in the visualization reveal that AI research in this field is conducted through intensive collaborations among specific institutions. Such analyses not only evaluate the collaboration potential of institutions but also provide strategic guidance for more effective academic research direction and the identification of new partnership opportunities.

Author Collaboration Analysis

The author collaboration network in the field of "artificial intelligence" was analyzed based on articles in the Web of Science database, and the findings are presented in Figure 5.

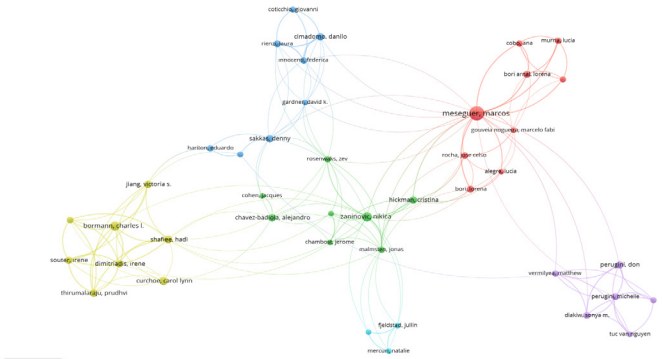

Figure 5. Bibliometric network of author collaborations (circle size represents key authors, and lines between circles indicate collaboration relationships.)

Figure 5 analyzes the bibliographic connections and collaboration networks among authors who have published at least five articles. In total, 4,508 authors were assessed, but only 101 authors who met the specified criteria were included in the analysis. The requirement for authors to have published at least five articles ensures the selection of those who have made a significant contribution to the literature and demonstrated sufficient productivity. This threshold supports the inclusion of only active and influential authors, making the results more meaningful and reliable.

The map visualizes the bibliographic matches and collaboration networks among authors. Each author is represented by a circle, and the size of the circle reflects the extent of their contribution to the literature. The colors of the circles represent the clusters to which the authors belong, indicating thematic or research-oriented groups. The lines connecting the circles signify the strength of bibliographic connections, i.e., the frequency of shared references, providing a detailed view of the scientific relationships among authors.

The analysis identifies authors such as Meseguer-Marcos, Gouveia-Rodrigues, and Rocha-Cristina as central figures

within densely connected clusters. The red cluster represents authors who focus on specific subfields, with Meseguer-Marcos playing a leading role. Similarly, the blue cluster, represented by Coticchio Giovanni and Cimado Danilo, demonstrates extensive collaboration within and beyond their cluster. The green cluster, centered on Nikita Zaninovic and Cristina Hickman, shows strong internal connections, reflecting a thematic focus. Additionally, the yellow cluster represents a smaller group, including Charles Bormann and Irene Dimitriadis, who are involved in tightly-knit collaborations with specific research objectives. The orange and purple clusters, while smaller, exhibit strong internal links and concentrate on distinct research topics.

Overall, the dense connections and clear clustering structures in the visualization demonstrate active collaboration among authors and provide valuable insights for identifying prominent researchers in the field.

Geographic Distribution of Citations

The geographic distribution of citations for articles on "artificial intelligence" in the Web of Science database was analyzed, and the findings are visualized in Figure 6.

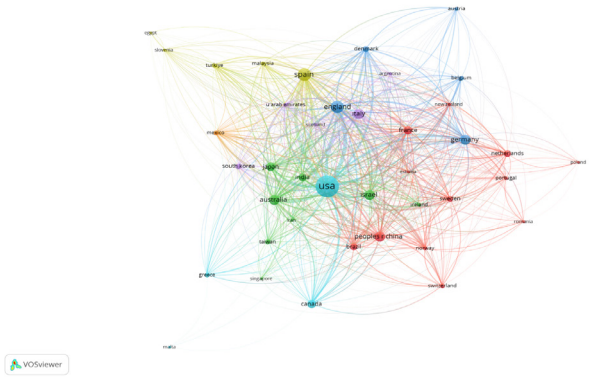

Figure 6. Geographic distribution of citations and collaborations

Figure 6 presents the geographic distribution of citations for articles on "artificial intelligence" and the academic collaboration networks among countries. In the analysis, countries with at least five articles were considered, and of the 76 countries meeting this criterion, 40 were included in the evaluation. Seven distinct clusters were identified, each corresponding to a specific research focus or collaboration network.

In the visualization, circles represent countries, and their size indicates the country's contribution to the literature. The lines connecting the circles represent collaborations between countries, and the thickness of the lines reflects the strength and frequency of these collaborations. The colors of the clusters indicate thematic or geographic groupings.

The United States is positioned at the center of the map as the leading contributor, with strong collaborations across other countries. The country plays a dominant role in global AI research and has established dense academic networks with many nations. European countries, including the United Kingdom, Germany, France, and Italy, also stand out as significant contributors, forming a regional powerhouse through strong collaborations with each other. The United Kingdom, in particular, has developed strong ties with the United States, playing a pivotal role in the global academic network.

In Asia, countries such as China, Japan, and South Korea are notable contributors. These nations have established significant collaborations both regionally and internationally, particularly with the United States. China, positioned near the center, highlights its growing influence in the literature and its role in fostering international collaborations.

Turkiye, positioned in the middle of the map, acts as a bridge between Europe, Asia, and the Middle East, highlighting its strategic importance in connecting various regions. Smaller circles represent countries contributing less extensively but still participating in collaborations within niche areas.

Overall, the visualization of geographic distribution underscores the global nature of AI research. The leading roles of the United States, the United Kingdom, Germany, and China are evident, while Turkey and other Asian countries stand out for their regional collaborations. This analysis highlights the importance of fostering international academic collaborations to advance research in this field.

DISCUSSION

The findings of this study provide a comprehensive overview of research trends in AI within Obs&Gyn, highlighting both the progression of the field and its current research hotspots. These findings are compared with those from prominent bibliometric analyses, including studies by Xiao et al.,[1] Levin et al.[4] and Sibanda et al.,[5] to identify overlaps and divergences.

Our study, consistent with Xiao et al.[1] and Levin et al.[5] demonstrates a marked increase in AI-related research in Obs&Gyn after 2017. This period coincides with technological advancements in machine learning and deep learning, coupled with increased computational capacity and data availability. Both studies similarly identified 2018-2024 as a phase of accelerated growth, reflecting heightened interest in AI applications for reproductive health and diagnostic tools.

However, while Xiao et al.[1] emphasized a surge in publications related specifically to robotic surgery, our findings indicate a broader diversification of research topics, including embryo selection, IVF, and diagnostic imaging. This broader scope suggests that AI applications in Obs&Gyn extend beyond surgical innovations to encompass various facets of clinical care.

This study identified Human Reproduction and Fertility and Sterility as leading journals in publishing AI-related research in Obs&Gyn. Levin et al.[2] similarly highlighted these journals but noted that articles on robotic surgery received the highest citations, underscoring their significant impact. Our analysis aligns with this observation but also brings attention to journals like Ultrasound in Obstetrics & Gynecology and Reproductive Biomedicine Online, which play key roles in disseminating AI research across diverse subfields.

Consistent with Levin et al.[4] and Sibanda et al.,[5] this study revealed that institutions in the United States and Europe dominate AI research in Obs&Gyn. Notably, our findings show a growing contribution from Asian countries, particularly China, reflecting the global expansion of research in this domain. Sibanda et al.[4] also emphasized the emerging contributions from African institutions, particularly in maternal health, which was less apparent in our dataset. This discrepancy highlights the need for more inclusive bibliometric analyses to capture diverse regional contributions.

Our findings confirm the central role of institutions like Harvard Medical School and the University of Oxford in fostering global collaboration networks, as reported by Levin et al.[2] These institutions act as hubs for multi-institutional research efforts. Additionally, Vosviewer analyses from all studies consistently demonstrate the dominance of North American and European institutions, with developing regions beginning to make noticeable contributions.

The keyword analysis in this study closely aligns with findings from Xiao et al.[1] and Levin et al.[2] Terms such as "machine learning," "deep learning," and "IVF" emerged as prominent themes, reflecting the interdisciplinary nature of AI research in Obs&Gyn. Interestingly, Sibanda et al.[4] highlighted the role of Industry 4.0 technologies, which was less emphasized in our dataset but represents an emerging frontier in maternal healthcare research.[17-19]

The overlaps and divergences across studies point to several critical directions for future research. First, expanding bibliometric datasets by integrating data from multiple platforms such as PubMed and Scopus can help capture more diverse contributions, particularly from underrepresented regions. This approach would provide a more comprehensive understanding of global research activities and their impact. Second, focusing on the ethical and social implications of AI in obstetrics and gynecology, as noted by Sibanda et al.[4] is crucial for ensuring equitable and responsible deployment of these technologies. Addressing these implications would promote the fair use of AI and mitigate potential disparities in healthcare access and outcomes.[18]

Additionally, strengthening collaborations with institutions in developing regions can diversify perspectives and foster innovation in resource-limited settings. Such partnerships would not only enhance global research networks but also empower regions with limited resources to contribute to the evolving AI landscape. Lastly, the integration of AI with emerging technologies like ChatGPT offers promising avenues for enhancing clinical decision-making. Exploring these innovative applications further could pave the way for more efficient and personalized healthcare solutions. These future directions collectively underscore the need for inclusive, ethical, and technologically adaptive approaches to advance AI research in obstetrics and gynecology.

Limitations

This study has several limitations that should be considered when interpreting the findings. First, the analysis was confined to publications indexed in the Web of Science Core Collection database, potentially excluding relevant articles from other databases such as Scopus or PubMed. Second, the study focused only on articles published in English, which may overlook contributions in other languages, particularly from non-English-speaking regions. Lastly, the use of "artificial intelligence" as the sole keyword may have excluded studies using alternative terminology, limiting the scope of the dataset.

CONCLUSION

This bibliometric analysis highlights the rapid growth and interdisciplinary nature of AI research within obstetrics and gynecology over the past two decades. The results demonstrate the increasing global interest in this field, with significant contributions from North America, Europe, and emerging regions such as Asia. Key themes, including machine learning, deep learning, and diagnostic imaging, emphasize the transformative potential of AI in enhancing maternal and fetal healthcare. While the findings underscore the importance of collaboration and innovation, they also reveal disparities in research output and collaboration across regions. Addressing these disparities and fostering inclusive research networks will be crucial for maximizing the global impact of AI in Obs&Gyn. Future research should explore underrepresented regions, integrate data from multiple databases, and address ethical implications to ensure equitable and responsible AI adoption. By offering a detailed understanding of current trends and gaps, this study provides a roadmap for advancing AI research and its application in obstetrics and gynecology.

ETHICAL DECLARATIONS

Ethics Committee Approval

Since our study is a bibliometric analysis based solely on previously published data obtained from public databases, ethical approval is not required.

Informed Consent

Since this research is a bibliometric study, it did not require informed consent.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

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

Author Contributions

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

REFERENCES

1. Xiao P, Li L, Qu J, Wang G. Global research hotspots and trends on robotic surgery in obstetrics and gynecology: a bibliometric analysis based on VOSviewer. *Front Surg.* 2024;11:1308489. doi:10.3389/fsurg.2024.1308489

2. Levin G, Siedhoff M, Wright KN, et al. Robotic surgery in obstetrics and gynecology: a bibliometric study. *J Robot Surg.* 2023;17(5):2387-2397. doi:10.1007/s11701-023-01672-1
3. Lestari D, Maulana FI, Adi PDP, Rahayu A, Nadlifatin R, Widartha VP. Bibliometric analysis of intelligent techniques for obstetric complication prediction in the last 20 years. In: 2024 4th International Conference on Emerging Smart Technologies and Applications (eSmarTA). *IEEE*. 2024: 1-8. doi:10.1109/eSmarTA62850.2024.10638994
4. Sibanda K, Ndayizigamiye P, Twinomurinzi H. Industry 4.0 technologies in maternal health care: bibliometric analysis and research agenda. *JMIR Pediatr Parent.* 2024;7:e47848. doi:10.2196/47848
5. Levin G, Brezinov Y, Meyer R. Exploring the use of ChatGPT in OBGYN: a bibliometric analysis of the first ChatGPT-related publications. *Arch Gynecol Obstet.* 2023;308(6):1785-1789. doi:10.1007/s00404-023-07081-x
6. Ray PP. Bridging the gap: integrating ChatGPT into obstetrics and gynecology research-a call to action. *Arch Gynecol Obstet.* 2024;309(3): 1111-1113. doi:10.1007/s00404-023-07129-y
7. Sreedharan S, Mian M, Robertson RA, Yang N. The top 100 most cited articles in medical artificial intelligence: a bibliometric analysis. *J Med Artif Intell.* 2020;3. doi:10.21037/jmai.2019.11.04
8. Xiong DD, He RQ, Huang ZG, et al. Global bibliometric mapping of the research trends in artificial intelligence-based digital pathology for lung cancer over the past two decades. *Digit Health.* 2024;10: 205520762412 77735. doi:10.1177/20552076241277735
9. Fu Q, Lu Z, Sui J, Chang Y, Zhang M. Trends in artificial intelligence and ultrasound medicine: a bibliometric and visualized analysis. Available at SSRN 4333609. 2023.
10. Oprescu AM, Miro-Amarante G, García-Díaz L, Beltrán LM, Rey VE, Romero-Ternero M. Artificial intelligence in pregnancy: a scoping review. *IEEE Access.* 2020;8:181450-181484.
11. Triantafyllopoulos L, Paxinou E, Feretzakis G, Kalles D, Verykios VS. Mapping how artificial intelligence blends with healthcare: insights from a bibliometric analysis. *Future Internet.* 2024;16(7):221. doi:10. 3390/fi16070221
12. Tran T, Do TT, Reid I, Carneiro G. Bayesian generative active deep learning. In International conference on machine learning. 2019, pp. 6295-6304). PMLR.
13. Tagliafico AS, Piana M, Schenone D, Lai R, Massone AM, Houssami N. Overview of radiomics in breast cancer diagnosis and prognostication. *Breast.* 2020;49:74-80. doi:10.1016/j.breast.2019.10.018
14. VerMilyea M, Hall JMM, Diakiw SM, et al. Development of an artificial intelligence-based assessment model for prediction of embryo viability using static images captured by optical light microscopy during IVF. *Hum Reprod.* 2020;35(4):770-784. doi:10.1093/humrep/deaa013
15. Carter SM, Rogers W, Win KT, Frazer H, Richards B, Houssami N. The ethical, legal and social implications of using artificial intelligence systems in breast cancer care. *Breast.* 2020;49:25-32. doi:10.1016/j. breast.2019.10.001
16. Grünebaum A, Chervenak J, Pollet SL, Katz A, Chervenak FA. The exciting potential for ChatGPT in obstetrics and gynecology. *Am J Obstet Gynecol.* 2023;228(6):696-705. doi:10.1016/j.ajog.2023.03.009
17. Brandt JS, Hadaya O, Schuster M, Rosen T, Sauer MV, Ananth CV. A bibliometric analysis of top-cited journal articles in obstetrics and gynecology. *JAMA Netw Open.* 2019;2(12):e1918007. doi:10.1001/jama networkopen.2019.18007
18. Maulana FI, Lestari D, Adi PDP, Hamim M, Rahayu A, Widartha VP. Direction of machine learning research in obstetric care in the last 10 years. In: 2024 4th International Conference on Emerging Smart Technologies and Applications (eSmarTA). IEEE; 2024. p. 1-7.
19. Xie Y, Zhai YS, Lu G. Evolution of artificial intelligence in healthcare: a 30-year bibliometric study. *Front Med.* 2024;11:1505692. doi:10.3389/ fmed.2024.1505692

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1720074
J Health Sci Med. 2025;8(5):747-753

Prognostic value of Systemic Immune-Inflammation Index in head and neck cancer

Engin Eren Kavak[1], İsmail Dilli[1], Ömer Bayır[2], Hüseyin Demirci[3], Samet Özlügedik[2]

[1]Department of Medical Oncology, Ankara Etlik City Hospital, Ankara, Turkiye
[2]Department of Ear, Nose, Throat and Head and Neck Surgery, Ankara Etlik City Hospital, Ankara, Turkiye
[3]Department of Endocrinology and Metabolism, Ankara Etlik City Hospital, Ankara, Turkiye

Cite this article as: Kavak EE, Dilli İ, Bayır Ö, Demirci H, Özlügedik S. Prognostic value of Systemic Immune-Inflammation Index in head and neck cancer. *J Health Sci Med.* 2025;8(5):747-753.

Received: 15.06.2025 ◆ **Accepted**: 14.07.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: This study aimed to evaluate the prognostic value of the Systemic Immune-Inflammation Index (SII) in patients with head and neck cancer and its association with survival outcomes including disease-free survival (DFS) and overall survival (OS).
Methods: The patients diagnosed with head and neck cancer were retrospectively analyzed. Patients were stratified into two groups based on the SII cut-off value (796): low SII (L-SII) and high SII (H-SII). Clinical, demographic, and treatment-related parameters were compared between the groups. Kaplan-Meier survival analysis and Cox regression were used for univariate and multivariate analyses of DFS and OS.
Results: Of the total number of patients included in the study (n=184), 67 with high SII (≥796) exhibited significantly higher recurrence rates (43.3% vs. 8.5%, p<0.001) and higher mortality (26.9% vs. 11.1%, p=0.006) compared to those with low SII. Median DFS was shorter in the H-SII group (13.7 vs. 18.7 months), although the difference was not statistically significant (p=0.25). In multivariate Cox analysis, advanced stage (HR: 3.00, 95% CI: 1.38-6.50, p=0.005), ECOG ≥2 (HR: 3.72, 95% CI: 1.35-10.22, p=0.01), and high SII (HR: 1.86, 95% CI: 1.01-3.16, p=0.05) were independently associated with worse OS. Although high SII was not an independent predictor for DFS, it showed a clear trend toward worse outcomes (HR: 1.56, 95% CI: 0.72-3.34, p=0.25).
Conclusion: High SII levels were associated with worse clinical outcomes and significantly higher rates of recurrence and mortality. While SII was an independent prognostic factor for OS, its effect on DFS did not reach statistical significance. These findings support the potential utility of SII as a simple, inflammation-based prognostic biomarker in head and neck cancers.
Keywords: Head and neck cancer, Systemic Immune-Inflammation Index, overall survival, progression-free survival, disease-free survival

INTRODUCTION

Head and neck cancer (HNC) encompasses a range of malignancies originating in the nasopharynx, larynx, oropharynx, hypopharynx, oral cavity, salivary glands, and paranasal sinuses and is a significant global health issue.[1] The treatment approach for HNC varies based on stage and location. While early-stage disease is typically treated with surgery or radiotherapy, induction chemotherapy is generally reserved for selected cases with locally advanced tumors. In recurrent or metastatic settings, systemic therapy aims to prolong survival and manage symptoms, although curative treatment is generally not feasible.

Optimal decision-making, treatment planning, and posttreatment response assessment for HNC patients require a multidisciplinary approach involving surgeons, medical oncologists, and radiation oncologists, as well as dentists, speech/swallowing pathologists, dietitians, psychosocial oncologists, prosthodontists, and rehabilitation therapists. Multidisciplinary tumor boards significantly impact diagnostic and treatment decisions for many patients with newly diagnosed HNC.[2]

Despite advancements in multidisciplinary care, prognostic standards still fall short, leading to divergent survival rates among patients with identical tumor-node-metastasis (TNM) stages. Research has focused on prognostic factors aiding clinicians in identifying individuals with elevated susceptibility to HNC recurrence and mortality.[3] Recognized prognostic factors in HNC include TNM staging, extranodal spread, HPV status, and patient attributes such as age, performance status, and history of smoking and alcohol consumption.[4]

Patient immunity and systemic inflammation play pivotal roles in angiogenesis and cancer progression, high neutrophil and platelet counts promote angiogenesis and tumor progression, while lymphopenia indicates impaired antitumor immunity.[5] Studies have revealed associations between inflammation markers-such as the neutrophil-to-lymphocyte ratio,

Corresponding Author: Engin Eren Kavak, engineren2000@yahoo.com

lymphocyte-to-monocyte ratio, platelet-to-lymphocyte ratio, and C-reactive protein/albumin ratio-and cancer prognosis and survival.[6-9] These findings underscore the significance of considering inflammatory markers in prognostic evaluation and treatment strategies for cancer patients.

A systematic review and meta-analysis encompassing 100 studies with 40,559 patients investigated the Systemic Immune-Inflammation Index (SII) as a prognostic determinant in various malignant solid tumors and revealed that elevated SII levels detrimentally impacted overall survival.[10]

While TNM staging and HPV status remain important prognostic indicators, additional biomarkers may help refine risk stratification in heterogeneous HNC populations. The present study investigates the impact of pretreatment SII values on survival outcomes in patients diagnosed with HNC. The study hypothesizes that high pretreatment SII is associated with poor disease-free survival (DFS) and overall survival (OS) in patients with HNC.

METHODS

The study was approved by the Ankara Etlik City Hospital Scientific Researches Evaluation and Ethics Committee (Date: 15.05.2024, Decision No: AEŞH-BADEK-2024-432). The study was conducted in accordance with the ethical principles set forth in the 1975 Declaration of Helsinki for the conduct of research.

We retrospectively analyzed patient records and electronic health data from individuals who were diagnosed with HNC and admitted to the Medical Oncology Clinic of Ankara Dışkapı Training and Research Hospital and the Medical Oncology Clinic of Ankara Etlik City Hospital between January 2017 and March 2024.

Although head and neck cancer is often considered a single entity, it encompasses a wide spectrum of malignancies with distinct biological behaviors, etiologies, and treatment responses. Nasopharyngeal carcinomas are strongly associated with Epstein-Barr Virus and exhibit unique epidemiologic and therapeutic characteristics, including a high sensitivity to radiotherapy. Similarly, salivary gland tumors comprise a heterogeneous group of histological subtypes, such as adenoid cystic and mucoepidermoid carcinomas, which differ significantly from squamous cell carcinomas in terms of progression patterns, treatment strategies, and prognosis. To ensure a homogeneous study population and to accurately evaluate the prognostic significance of the SII, patients with nasopharyngeal and salivary gland malignancies were excluded from the analysis. The following criteria were utilised for the exclusion of patients: the presence of active infections, autoimmune disorders (e.g. Behcet's disease or Hashimoto's thyroiditis), haemoglobinopathies, haematological conditions such as sickle cell anaemia, coagulation disorders, liver disorders, renal diseases, the use of corticosteroids during treatment, and incomplete baseline blood test results (**Figure 1**).

Standardized protocols were used across participating centers for Eastern Cooperative Oncology Group (ECOG) performance scoring, laboratory testing, and staging. We examined various parameters, including smoking habits, sex distribution, tumor localization, tumor stage [according to the American Joint Committee on Cancer TNM Staging Guidelines, 8th Edition (AJCC 2017)], treatment modalities, and the impact of radiotherapy, chemotherapy, surgery, and their combination on patient survival. We also explored the relationship between pretreatment SII values and survival outcomes. Performance status was assessed using the ECOG Performance Scale.[11]

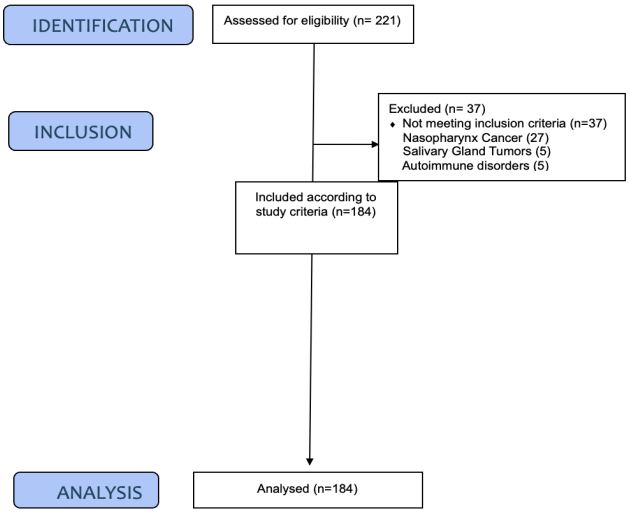

Figure 1. STROBE flow diagram

STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

SII Evaluation

SII was calculated using blood samples collected within seven days prior to the initiation of any treatment. The index was calculated using granulocytes as a proxy for platelet (P), neutrophil (N), and lymphocyte (L) counts with the following formula: SII=absolute P×absolute N/absolute L. The receiver operating characteristic (ROC) analysis was used to determine the optimal cut-off values for classifying the SII as low (LSII) or high (HSII) for predicting overall survival. The LSII and HSII were defined as values below and above the cut-off point, respectively.

Statistical Analysis

The primary endpoint was OS defined as the interval between the start of treatment and the date of death or last visit. The secondary endpoint was DFS, defined as the time from treatment initiation to recurrence, death, or last follow-up, reflecting curative-intent outcomes. Quantitative variables are expressed as the means and ranges, while categorical variables are expressed as percentage frequency distributions. Pearson's X^2 test was used to compare demographic characteristics between groups.

Kaplan-Meier survival curves were used to estimate survival outcomes, and log-rank tests were performed for intergroup comparisons. The Cox regression model identified independent risk factors associated with DFS and OS in univariate analyses. Significant variables from univariate analysis were included in multivariate Cox analysis. A two-tailed $p<0.05$ was considered to indicate statistical significance. Statistical analyses were performed using the Bluesky statistical (Version 10.3.2) program.

RESULTS

Of the 184 patients, 156 (84.8%) were male, and 28 (15.2%) were female. The median age was 63 years (range 32-91). The tumor locations included the larynx (51.6%), oral cavity (23.4%), oropharynx (10.9%), hypopharynx (9.8%), and paranasal sinuses (3.8%). Staging classified 1.6% as stage I, 13.6% as stage II, 33.7% as stage III, 32.6% as stage IVA, 13.6% as stage IVB, and 4.9% as stage IVC. There was no significant difference in demographic characteristics between the LSII and HSII groups (Table 1).

The ROC analysis was conducted to ascertain the most suitable cut-off values for the classification of SII as low (LSII) or high (HSII). The cut-off point was determined to be 796. The area under the curve (AUC) was determined to be 0.808 (95% CI: 0.697-0.952), p<0.001, with a specificity of 74.4% and sensitivity of 74.5% (Figure 2).

In the comparison of treatment and survival outcomes according to SII groups, no statistically significant differences were observed in the distribution of primary treatment modalities, including surgery, concurrent chemoradiotherapy

Table 1. Clinical and demographical parameters

Parameters	Whole cohort (n=184)	L-SII (<796) (n=117)	H-SII (≥796) (n=67)	p-value
Sex n (%)				
Male	156 (84.8)	100 (85.5)	56 (83.6)	0.73
Female	28 (15.2)	17 (14.5)	11 (16.4)	
Age (years)				
Median (min-max)	63.0 (32-91)	63 (25-90)	61 (32-91)	0.27*
Age groups n (%)				
<65 years	107 (58.2)	65 (55.6)	42 (62.7)	0.34
≥65 years	77 (41.8)	52 (44.4)	25 (37.3)	
Smoking n (%)				
Current smoker	156 (84.8)	101 (86.3)	55 (82.1)	0.44
Never smoke	28 (15.2)	16 (33.7)	12 (17.9)	
ECOG PS n (%)				
0	100 (54.3)	67 (57.3)	33 (49.3)	0.22
1	73 (39.7)	45 (38.5)	28 (41.8)	
2≤	11 (6.0)	5 (4.2)	6 (8.9)	
Primer n (%)				
Larynx	95 (51.6)	68 (58.1)	27 (40.3)	0.16
Oral cavity	43 (23.4)	23 (19.7)	20 (29.9)	
Oropharynx	20 (10.9)	13 (11.1)	7 (10.4)	
Hypopharynx	18 (9.8)	9 (7.7)	9 (13.4)	
Paranasal sinuses	7 (3.8)	3 (2.6)	4 (6.0)	
Stage n (%)				
I	3 (1.6)	2 (1.7)	1 (1.5)	0.08
II	25 (13.6)	13 (11.1)	12 (17.9)	
III	62 (33.7)	47 (40.2)	15 (22.4)	
IVA	60 (32.6)	38 (32.5)	22 (32.8)	
IVB	25 (13.6)	11 (9.4)	14 (20.9)	
IVC	9 (4.9)	6 (5.1)	3 (4.5)	
Extranodal extension n (%)				
Yes	37 (20.1)	18 (15.4)	19 (28.4)	0.03
No	147 (79.9)	99 (84.6)	48 (71.6)	
HPV status n (%)				
Not evaluated	159 (86.4)	102 (87.2)	57 (85.1)	0.80
Positive	9 (4.9)	6 (5.1)	3 (4.5)	
Negative	16 (8.7)	9 (7.7)	7 (10.4)	

Pearson X^2 test, *: Mann-Whitney U test, SII: Systemic Immune-Inflammation Index, min: Minimum, max: Maximum, ECOG: Eastern Cooperative Oncology Group, PS: Performance status, HPV: Human papillomavirus

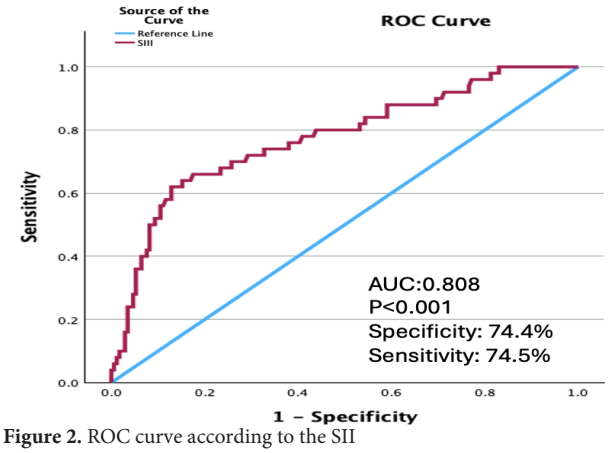

Figure 2. ROC curve according to the SII

ROC: Receiver operating characteristic, SII: Systemic Immune-Inflammation Index

(CCRT), induction chemotherapy followed by CCRT, palliative chemotherapy, radiotherapy alone, or no treatment (p=0.21). Although the proportion of patients receiving adjuvant chemotherapy was higher in the high SII group (29.9%) compared to the low SII group (17.9%), this difference did not reach statistical significance (p=0.06). Notably, disease recurrence was significantly more common in the high SII group (43.3%) than in the low SII group (8.5%) (p<0.001). Similarly, mortality was higher in the high SII group (26.9% vs. 11.1%, p=0.006) (Table 2).

Local/regional recurrence or metastasis occurred in 21.2% of patients: 10 in the LSII arm and 29 in the HSII arm. The median DFS was 18.7 months in the LSII arm and 13.7 months in the HSII arm (p=0.25) (Figure 3). During follow-up, 31 patients died (13 in the LSII arm and 18 in the HSII arm). The estimated median OS was 103.8 months in the LSII arm and 80.3 months in the HSII arm (p=0.035) (Figure 4).

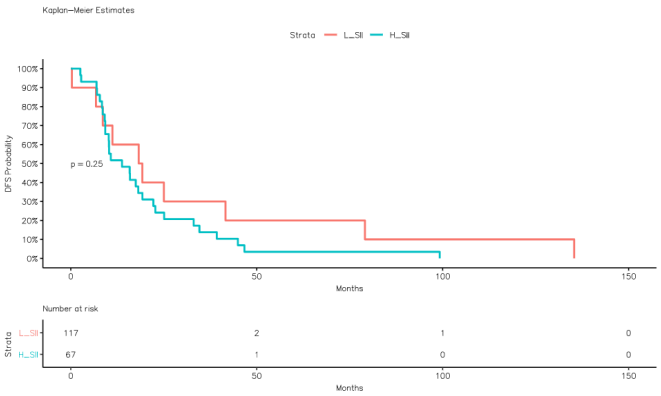

Figure 3. Kaplan-Meier estimates of disease-free survival according to the SII

SII: Systemic Immune-Inflammation Index

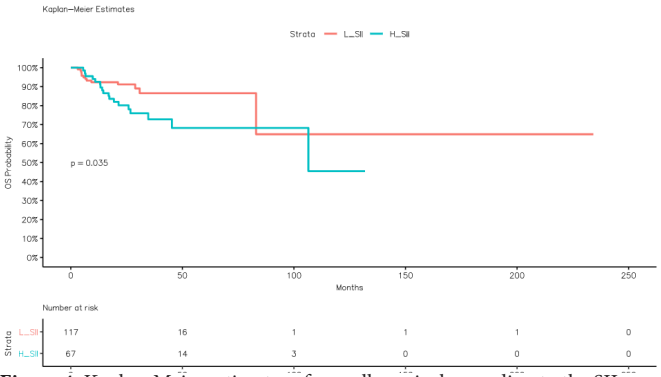

Figure 4. Kaplan-Meier estimates of overall survival according to the SII

SII: Systemic Immune-Inflammation Index

Table 2. Treatment and survival parameters

Parameters	Whole cohort (n=184)	L-SII (<796) (n=117)	H-SII (≥796) (n=67)	p-value
Primary treatment n (%)				
Surgery	57 (31.0)	29 (24.8)	28 (41.8)	0.21
Concurrent	70 (38.0)	47 (40.2)	23 (34.3)	
Chemoradiotherapy (CC)	46 (25.0)	34 (29.1)	12 (17.9)	
Induction chemotherapy+CC	4 (2.2)	3 (2.6)	1 (1.5)	
Palliative chemotherapy	3 (1.6)	2 (66.7)	1 (1.5)	
Only radiotherapy	4 (2.2)	2 (1.7)	2 (3.0)	
None				
Adjuvant chemotherapy n (%)				
Yes	41 (22.3)	21 (17.9)	20 (29.9)	0.06
No	73 (39.7)	96 (82.1)	47 (70.1)	
Recurrence n (%)				
Yes	39 (21.2)	10 (8.5)	29 (43.3)	<0.001
No	145 (78.8)	107 (91.5)	38 (56.7)	
Median DFS (range) (months)	15.8 (8.6-NR)	18.7 (8.6-NR)	13.7 (10.2-22.2)	0.25*
Exitus n (%)				
Yes	31 (16.8)	13 (11.1)	18 (26.9)	0.006
No	153 (83.2)	104 (88.9)	49 (73.1)	
Median OS (range) (months)	103.8 (38.7-168.8)	103.8 (16.8-190.8)	80.3 (7.8-153.1)	0.14*
Median follow up) (months)	21.6	19.5	22.8	

Pearson X^2 test, *: Log-rank test, SII: Systemic Immune-Inflammation Index, PS: Performance status, DFS: Disease-free survival, OS: Overall survival

In the univariate Cox regression analysis for DFS, older age (≥65), smoking status, advanced stage (stage IV), and higher ECOG performance score (≥2) were associated with poorer outcomes. However, in the multivariate model, age ≥65 (HR: 0.41, 95% CI: 0.18-0.94, p=0.03), current smoking (HR: 3.44, 95% CI: 1.00-11.78, p=0.04), and advanced stage (HR: 3.00, 95% CI: 1.38-6.50, p=0.005) remained independent predictors of shorter DFS. ECOG performance status and SII were not statistically significant in the multivariate analysis for DFS (Table 3).

Table 3. Univariate and multivariate analyses of the prognostic factors affecting DFS in head and neck cancer

Parameters	Univariate analyses		Multivariate analyses	
	HR (95% CI)	p	HR (95% CI)	p
Sex				
Female	Ref	0.10		
Male	1.93 (0.87-4.29)			
Age				
<65 age	Ref	0.03	0.41 (0.18-0.94)	0.03
≥65 age	0.45 (0.21-0.94)			
Smoking				
Never	Ref	0.01	3.44 (1.00-11.78)	0.04
Current	2.05 (0.95-4.41)			
Stage				
≤III	Ref	0.02	3.00 (1.38-6.50)	0.005
IV	2.18 (1.12-4.25)			
Extranodal extension				
No	Ref	0.02	0.89 (0.28-2.77)	0.84
Yes	2.40 (1.13-5.11)			
ECOG PS				
0-1	Ref	0.14		
2≤	2.46 (0.72-8.38)			
SII				
<796	Ref	0.25		
≥796	1.56 (0.72-3.34)			

Cox regression analyse, HR and 95% CI, DFS: Disease-free survival, HR: Hazard ratios, CI: Confidence intervals, ECOG: Eastern Cooperative Oncology Group, PS: Performance status, SII: Systemic Immune-Inflammation Index

For OS, univariate analysis revealed that advanced stage, ECOG ≥2, and high SII (>796) were significantly associated with worse outcomes. These associations persisted in multivariate analysis, where advanced stage (HR: 2.77, 95% CI: 1.24-6.15, p=0.01), ECOG ≥2 (HR: 3.72, 95% CI: 1.35-10.22, p=0.01), and high SII (HR: 1.86, 95% CI: 1.01-3.16, p=0.05) were identified as independent negative prognostic factors for OS (Table 4).

DISCUSSION

Many studies have investigated the relationships between stage, performance status, treatment protocol, and prognosis in HNC patients, identifying these parameters as important prognostic markers for overall survival.[12-15] Similar to the literature, our study revealed significant relationships between stage, ECOG performance status, and overall survival.

Table 4. Univariate and multivariate analyses of the prognostic factors affecting OS in head and neck cancer

Parameters	Univariate analyses		Multivariate analyses	
	HR (95% CI)	p	HR (95% CI)	p
Sex				
Female	Ref	0.92		
Male	1.04 (0.40-2.73)			
Age				
<65 age	Ref	0.19		
≥65 age	0.60 (0.28-1.29)			
Smoking				
Never	Ref	0.66		
Current	0.79 (0.27-2.26)			
Stage				
≤III	Ref	0.01	2.77 (1.24-6.15)	0.01
IV	2.67 (1.22-5.83)			
Extranodal extension				
No	Ref	<0.001	3.42 (1.47-7.93)	0.004
Yes	4.86 (2.38-9.90)			
ECOG PS				
0-1	Ref	0.004	3.72 (1.35-10.22)	0.01
2≤	4.10(1.55-10.80)			
SII				
<796	Ref	0.03	1.86 (1.01-3.16)	0.05
≥796	2.13 (1.03-4.39)			

Cox regression analyse, HR and 95% CI, OS: Overall survival, HR: Hazard ratios, CI: Confidence intervals, ECOG: Eastern Cooperative Oncology Group, PS: Performance status, SII: Systemic Immune-Inflammation Index

The five-year OS for patients with stage I-II disease is typically 70-90%, while patients with advanced disease (stage III-IV) have a worse prognosis.[16] Our patient group included 84.8% with stage III-IV disease, with a 21.2% recurrence rate.

Smoking, a marker accepted as a prognostic factor in head and neck cancers, was evaluated in one study. According to the post hoc analysis, the 2-year PFS was significantly greater for patients who smoked less than 10 pack-years than for those who smoked 10 pack-years or more (92% vs. 57%; p=.0014). There was also a statistically significant difference in 2-year OS, although not as markedly as PFS (93% vs. 86%; p=.040).[17] In our study, we observed a significant prognostic effect of smoking on DFS (p=0.01).

Many studies emphasize that extranodal (or extracapsular) extension is a negative prognostic factor in patients with HNC undergoing primary surgery.[18,19] In our study, extranodal extension was positive in 20.1% of the patients. A significant negative prognostic effect on DFS and OS was observed in univariate analysis (p=0.02 and p<0.001, respectively). Multivariate analysis (p=0.004) revealed that this effect was maintained on OS.

Treatment of patients with HPV-associated oropharyngeal carcinoma is similar to that of HPV-negative patients, except in the context of clinical trials. Although testing for HPV associations provides prognostic information, there are

insufficient phase III data to modify treatment according to HPV status. At the same time, despite this excellent prognosis, positive margins and extracapsular extension may be associated with worse oncologic outcomes, including the risk of developing systemic disease.[20] In the present retrospective study, the effect of HPV on prognosis was not evaluated due to its absence from the pathology reports of the majority of patients.

High neutrophil or monocyte counts have been linked to poorer oncological outcomes not only in HNC but also in various other tumor types.[21] Previous meta-analyses and studies have underscored the prognostic significance of associations between absolute neutrophil, lymphocyte, monocyte, and platelet counts. Consequently, there has been a growing emphasis on amalgamating these parameters and examining their combinations as potential biomarkers.[22-24]

While numerous studies have highlighted the association between markers of systemic inflammation and DFS and OS across various tumor types, there remains a lack of consensus regarding cut-off values specific to each cancer type. In a study by Rizzo et al.[25] comprising 925 HPV-negative HNSCC patients, the SII cut-off values for OS and DFS were <602 and >754, respectively. In our study, we determined the SII cut-off value for predicting survival to be 796. These varying cut-off values underscore the complexity of systemic inflammatory markers and highlight the need for further research to establish standardized thresholds tailored to specific tumor types.

In 2022, Wang et al.[26] conducted a meta-analysis involving 12 studies comprising 4369 HNC patients and revealed that elevated pretreatment SII values were correlated with worse OS, DFS. In a retrospective study by Zhou et al.[27] focusing on HNSCC patients, a high SII was identified as a prognostic factor associated with both OS and DFS in univariate analysis, although it did not emerge as an independent prognostic factor in multivariate analysis. In the present study, high SII was identified as an independent poor prognostic factor for OS in both univariate and multivariate analyses.

This finding aligns with previous studies suggesting that systemic inflammatory markers such as the SII are significant for cancer prognosis. An elevated SII likely reflects an enhanced inflammatory response and an immunosuppressive microenvironment, contributing to cancer progression.

Clinical Implications

The identification of a high SII as a prognostic marker for OS has important clinical implications. These findings suggest that patients with elevated SII values might benefit from more aggressive or tailored therapeutic strategies. For instance, these patients might be considered for closer surveillance or adjunctive treatments aimed at modulating the immune response. Additionally, the SII could be integrated into existing prognostic models to improve their accuracy and utility in clinical decision-making.

Our study also underscores the need for a multidisciplinary approach to the management of HNC. Given the complex interplay between inflammation, immunity, and cancer, collaboration among oncologists, immunologists, and other specialists is crucial to optimize treatment outcomes. Further research into the underlying mechanisms by which the SII influences cancer progression could lead to novel therapeutic interventions targeting inflammatory pathways.

Limitations

While our study provides valuable insights, it is not without limitations. First, the retrospective design may introduce selection bias and limit the generalizability of our findings. Additionally, the single-center nature of the study may not reflect the broader population of HNC patients. Prospective, multicenter studies are needed to validate our results and confirm the prognostic value of the SII in diverse patient cohorts.

Due to the small number of patients receiving palliative chemotherapy (n=4), subgroup analysis by treatment intent was not feasible. Although a small number of patients received non-curative treatment, DFS was used as the secondary endpoint since most patients were treated with curative intent.

Another limitation is the lack of data on other potential confounding factors, such as HPV status, which is known to influence prognosis in HNC patients. HPV status, an important confounder especially for oropharyngeal tumors, was unavailable in most cases and therefore could not be included in the analysis. Future studies should consider including a comprehensive range of clinical and biological variables to provide a more nuanced understanding of the factors affecting prognosis.

CONCLUSION

As a result, the present study identified a high pretreatment SII value as an independent poor prognostic factor for OS in HNC patients. Although SII did not have a significant effect on DFS, its capacity to predict recurrence underscores its potential clinical utility. The incorporation of SII into routine prognostic assessments holds the potential to enhance risk stratification and to inform treatment strategies, thereby improving patient outcomes in HNC patients.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was carried out with the permission of the Ankara Etlik City Hospital Scientific Researches Evaluation and Ethics Committee (Date: 15.05.2024, Decision No: AEŞH-BADEK-2024-432).

Informed Consent

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

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

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

Author Contributions

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

Availability of Data and Materials

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

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. *Cancer J Clin*. 2021;71(3):209-249. doi:10.3322/caac.21660
2. Wheless SA, McKinney KA, Zanation AM. A prospective study of the clinical impact of a multidisciplinary head and neck tumor board. *Otolaryngol Head Neck Surg*. 2010;143(5):650-654. doi:10.1016/j.otohns.2010.07.020
3. O-charoenrat P, Pillai G, Patel S, et al. Tumour thickness predicts cervical nodal metastases and survival in early oral tongue cancer. *Oral Oncol*. 2003;39(4):386-390. doi:10.1016/s1368-8375(02)00142-2
4. Huang SH, O'Sullivan B. Overview of the 8th Edition TNM Classification for Head and Neck Cancer. *Curr Treat Options Oncol*. 2017;18(7):40. doi:10.1007/s11864-017-0484-y
5. Diakos CI, Charles KA, McMillan DC, Clarke SJ. Cancer-related inflammation and treatment effectiveness. *Lancet Oncol*. 2014;15(11):e493-e503. doi:10.1016/s1470-2045(14)70263-3
6. Ye M, Huang A, Yuan B, Tan G, Ai J, Liu H. Neutrophil-to-lymphocyte ratio and monocyte-to-eosinophil ratio as prognostic indicators for advanced nasopharyngeal carcinoma. *Eur Arch Oto Rhino Laryngol*. 2024;281(4):1971-1989. doi:10.1007/s00405-024-08474-7
7. Juesas Iglesias L, Sánchez-Canteli M, Pedregal Mallo D, et al. Prognostic value of preoperative inflammatory ratios in early glottic cancer treated with transoral laser surgery. *Head Neck*. 2024;46(4):819-830. doi:10.1002/hed.27631
8. Iwasa YI, Kitoh R, Yokota Y, et al. Post-treatment neutrophil/lymphocyte ratio is a prognostic factor in head and neck cancers treated with nivolumab. *Cancer Diagn Prog*. 2024;4(2):182-188. doi:10.21873/cdp.10305
9. Zhang S, Chen Z, Ling J, et al. Nomograms based on the lymphocyte-albumin-neutrophil ratio (LANR) for predicting the prognosis of nasopharyngeal carcinoma patients after definitive radiotherapy. *Scientific Reports*. 2024;14(1):5388. doi:10.1038/s41598-024-56043-z
10. Templeton AJ, McNamara MG, Šeruga B, et al. Prognostic role of neutrophil-to-lymphocyte ratio in solid tumors: a systematic review and meta-analysis. *J Nat Cancer Inst*. 2014;106(6):dju124. doi:10.1093/jnci/dju124
11. Oken MM, Creech RH, Tormey DC, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. *Am J Clin Oncol*. 1982;5(6):649-655.
12. Forastiere AA, Metch B, Schuller DE, et al. Randomized comparison of cisplatin plus fluorouracil and carboplatin plus fluorouracil versus methotrexate in advanced squamous-cell carcinoma of the head and neck: a Southwest Oncology Group study. *J Clin Oncol*. 1992;10(8):1245-1251. doi:10.1200/jco.1992.10.8.1245
13. Jacobs C, Lyman G, Velez-García E, et al. A phase III randomized study comparing cisplatin and fluorouracil as single agents and in combination for advanced squamous cell carcinoma of the head and neck. *J Clin Oncol*. 1992;10(2):257-263. doi:10.1200/jco.1992.10.2.257
14. Recondo G, Armand JP, Tellez-Bernal E, et al. Recurrent and/or metastatic head and neck squamous cell carcinoma: a clinical, univariate and multivariate analysis of response and survival with cisplatin-based chemotherapy. *Laryngoscope*. 1991;101(5):494-501. doi:10.1288/00005537-199105000-00009
15. Argiris A, Li Y, Forastiere A. Prognostic factors and long-term survivorship in patients with recurrent or metastatic carcinoma of the head and neck. *Cancer*. 2004;101(10):2222-2229. doi:10.1002/cncr.20640
16. Cadoni G, Giraldi L, Petrelli L, et al. Prognostic factors in head and neck cancer: a 10-year retrospective analysis in a single-institution in Italy. *Acta Otorhinolaryngol Italica*. 2017;37(6):458-466. doi:10.14639/0392-100x-1246
17. Marur S, Li S, Cmelak AJ, et al. E1308: Phase II trial of induction chemotherapy followed by reduced-dose radiation and weekly cetuximab in patients with HPV-associated resectable squamous cell carcinoma of the oropharynx-ECOG-ACRIN Cancer Research Group. *J Clin Oncol*. 2017;35(5):490-497. doi:10.1200/jco.2016.68.3300
18. An Y, Park HS, Kelly JR, et al. The prognostic value of extranodal extension in human papillomavirus-associated oropharyngeal squamous cell carcinoma. *Cancer*. 2017;123(14):2762-2772. doi:10.1002/cncr.30598
19. Bauer E, Mazul A, Chernock R, et al. Extranodal extension is a strong prognosticator in HPV-positive oropharyngeal squamous cell carcinoma. *Laryngoscope*. 2020;130(4):939-945. doi:10.1002/lary.28059
20. Vokes EE, Agrawal N, Seiwert TY. HPV-associated head and neck cancer. *J Nat Cancer Instit*. 2015;107(12):djv344. doi:10.1093/jnci/djv344
21. Hsueh C, Tao L, Zhang M, et al. The prognostic value of preoperative neutrophils, platelets, lymphocytes, monocytes and calculated ratios in patients with laryngeal squamous cell cancer. *Oncotarget*. 2017;8(36):60514-60527. doi:10.18632/oncotarget.16234
22. Tham T, Bardash Y, Herman SW, Costantino PD. Neutrophil-to-lymphocyte ratio as a prognostic indicator in head and neck cancer: a systematic review and meta-analysis. *Head Neck*. 2018;40(11):2546-2557. doi:10.1002/hed.25324
23. Wei D, Liu J, Ma J. The value of lymphocyte to monocyte ratio in the prognosis of head and neck squamous cell carcinoma: a meta-analysis. *PeerJ*. 2023;11:e16014. doi:10.7717/peerj.16014
24. Wang Y, Xu C, Zhang Z. Prognostic value of pretreatment lymphocyte-to-monocyte ratio in patients with glioma: a meta-analysis. *BMC Med*. 2023;21(1):486. doi:10.1186/s12916-023-03199-6
25. Boscolo-Rizzo P, D'Alessandro A, Polesel J, et al. Different inflammatory blood markers correlate with specific outcomes in incident HPV-negative head and neck squamous cell carcinoma: a retrospective cohort study. BMC Cancer. 2022;22(1):243. doi:10.1186/s12885-022-09327-4
26. Wang YT, Kuo LT, Weng HH, et al. Systemic Immun e-Inflammation Index as a predictor for head and neck cancer prognosis: a meta-analysis. *Front Oncol*. 2022;12:899518. doi:10.3389/fonc.2022.899518
27. Zhou S, Yuan H, Wang J, et al. Prognostic value of systemic inflammatory marker in patients with head and neck squamous cell carcinoma undergoing surgical resection. *Future Oncol*. 2020;16(10):559-571. doi:10.2217/fon-2020-0010

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1706182
J Health Sci Med. 2025;8(5):754-758

Diagnostic adequacy of thyroid fine needle aspiration biopsy: a comparative analysis by age and gender

Erkan Bilgin, Ahmet Bayrak

Department of Radiology, Dr. Abdurrahman Yurtaslan Ankara Oncology Training and Research Hospital, University of Health Sciences, Ankara, Turkiye

Cite this article as: Bilgin E, Bayrak A. Diagnostic adequacy of thyroid fine needle aspiration biopsy: a comparative analysis by age and gender. *J Health Sci Med.* 2025;8(5):754-758.

Received: 28.05.2025 ◆ **Accepted**: 16.07.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: To evaluate the diagnostic success of thyroid fine needle aspiration biopsy (FNAB) across different age and gender groups, and to investigate the effect of aging and sex-related variables on biopsy adequacy.

Methods: This retrospective study included 486 patients who underwent thyroid FNAB between 2022 and 2024. Patients were stratified into two age groups (<65 and ≥65 years) and analyzed by gender. Diagnostic outcomes were classified as diagnostic (benign, malignant, atypia) or non-diagnostic. Statistical analyses, including Fisher's exact and chi-square tests, were applied to determine the significance of differences.

Results: The overall diagnostic yield was 81.7%. Patients under 65 years had a significantly higher diagnostic success rate (86.4%) compared to those aged 65 and older (65.5%, p<0.01). Among older patients, diagnostic adequacy was notably lower in males (53.1%) than females (70.5%). The distribution of pathological diagnoses also varied: the atypia rate was higher in the ≥65 group (8.33%), whereas malignancy was slightly more prevalent in younger patients. Gender alone did not significantly influence diagnostic outcomes (p>0.05), but advanced age was associated with reduced adequacy, potentially due to structural tissue changes or technical limitations.

Conclusion: The diagnostic performance of thyroid FNAB diminishes significantly with age, especially in elderly male patients. Gender does not appear to be an independent determinant. These findings highlight the need for optimized biopsy strategies in older populations to enhance diagnostic yield.

Keywords: Thyroid nodule, biopsy, fine-needle, age factors, gender factors, diagnostic techniques and procedures, cytological techniques

INTRODUCTION

Thyroid nodules are frequently encountered in clinical practice, with a palpation-based prevalence of 5-7%, while ultrasonographic screening reveals nodules in up to 50% of the general adult population.[1] Although the majority of these nodules are benign, approximately 5-10% may harbor malignancy, making accurate evaluation critical in determining further management.[2] Thyroid nodules are more frequent in elderly patients, approximately 50% of individuals aged 65 years have thyroid nodules detected by ultrasonography.[3] A cross-sectional survey of asymptomatic adults using ultrasonography to detect thyroid nodules demonstrated an even higher prevalence of 80% in women and 74% in men over 60 years old.[4] With advancing age, the prevalence of clinically relevant thyroid nodules increases, whereas the risk that such nodules are malignant decreases.[5] Nonetheless, when thyroid cancer is detected in older individuals, a higher-risk histological phenotype is more likely.[6] It is of clinical importance to recognize potentially aggressive thyroid cancers that are more likely to occur in older adults.[7]

Fine needle aspiration biopsy (FNAB) has long been recognized as the first-line diagnostic modality for the assessment of thyroid nodules. It is favored for its minimally invasive nature, low complication risk, and high diagnostic accuracy in differentiating benign from suspicious or malignant lesions.[8] The Bethesda system for reporting thyroid cytopathology (BSRTC) has further standardized cytological interpretation and reporting, enhancing consistency and clinical decision-making.[9]

Despite the widespread use of FNAB, its diagnostic performance is subject to several influencing factors. These include nodule characteristics (e.g., size, location, cystic content), operator expertise, aspiration technique, and patient-specific variables such as age and gender.[10] Several studies have suggested that aging may negatively impact FNAB adequacy, potentially due to increased fibrosis, decreased cellularity, or altered tissue compliance in older patients.[11] Gender differences have also been explored in relation to FNAB adequacy and thyroid pathology. Although thyroid nodules

Corresponding Author: Erkan Bilgin, erk314@yahoo.com

are more prevalent in females, and women are more likely to undergo FNAB, the influence of gender on diagnostic success remains controversial. While some data suggest a marginally higher diagnostic yield in women, likely due to anatomical or hormonal factors, these differences are often not statistically significant.[12] Moreover, the interpretation of gender effects may be confounded by age distribution, comorbidities, and differences in thyroid disease epidemiology between sexes.

In addition to adequacy, the cytological classification of biopsy outcomes-including benign, malignant, and atypia of undetermined significance/ follicular lesion of undetermined significance (AUS/FLUS) plays a critical role in patient management pathways. Notably, the frequency of AUS/FLUS results and nondiagnostic samples may vary with age, which could influence follow-up recommendations, repeat biopsy rates, and surgical decisions.[13]

Thyroid FNAB shows high feasibility in elderly patients. Given the aging global population and increasing prevalence of thyroid nodular disease among the elderly, it is imperative to understand how age and gender influence FNAB performance. Older adults may present unique diagnostic challenges, and understanding the limits of FNAB in this subgroup could lead to protocol adjustments such as using adjunctive imaging guidance, repeat aspiration strategies, or more selective surgical referrals.[14,15] Core needle biopsy of thyroid nodules is a highly sensitive, specific and reliable method with no major complications.[16-18]

The present study aims to provide a comprehensive evaluation of thyroid FNAB outcomes by age and gender. Specifically, it investigates whether patients aged 65 and older exhibit lower diagnostic success compared to younger individuals, and whether sex-based disparities influence biopsy adequacy. By retrospectively analyzing a substantial cohort using robust statistical techniques, this study contributes to the ongoing discussion on optimizing FNAB protocols and refining diagnostic accuracy across diverse patient populations.

METHODS

Ethics

The study was carried out with the permission of the Dr. Abdurrahman Yurtaslan Ankara Oncology SUAM Non-interventional Clinical Researches Ethics Committee (Date: 22.05.2025, Decision No: 2025-05/74). The study was conducted in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki Declaration and its later amendments.

Study Design and Setting

This retrospective observational study was conducted at a tertiary care hospital and included patients who underwent thyroid FNAB between January 2022 and December 2024. The study protocol was approved by the institutional ethics committee, and all data were anonymized prior to analysis.

Patient Selection

A total of 486 patients were evaluated. Patient demographic data (age, gender) and cytopathological FNAB results were obtained from the hospital's electronic medical record system and pathology database. Two radiologists performed all FNAB procedures with at least 5 years of experience in thyroid ultrasonography and interventional procedures.

Inclusion criteria:

- Patients aged 18 years and older.
- Patients who underwent FNAB for thyroid nodules with available cytological results.
- First-time FNAB procedures (i.e., not repeated or follow-up biopsies).
- Adequate documentation of age and gender.

Exclusion criteria:

- Patients with incomplete or missing biopsy reports.
- Repeat biopsies of the same lesion.
- History of thyroid surgery or radioiodine therapy prior to biopsy.
- Inadequate documentation of clinical data (e.g., unknown age or gender).

Data Collection and Classification

Patients were stratified into two age groups;

- **Group 1:** <65 years
- **Group 2:** ≥65 years

Gender was categorized as male or female. FNAB results were grouped based on the Bethesda System into the following categories;

- **Diagnostic:** Benign, malignant, AUS/FLUS
- **Non-Diagnostic:** Specimens that failed to meet adequacy criteria as defined by BSRTC guidelines.

Statistical Analysis

Data analyses were performed using SPSS version 26.0 (IBM Corp., Armonk, NY, USA). Categorical variables were expressed as counts and percentages. The chi-square test (X^2) or Fisher's exact test was used to compare diagnostic rates across age and gender groups. A p-value less than 0.05 was considered statistically significant.

Subgroup analyses were also conducted to explore diagnostic adequacy specifically within age-gender intersections (e.g., ≥65 years female vs. ≥65 years male). Additional comparisons were made regarding the distribution of cytological outcomes (benign, malignant, AUS/FLUS) across the study groups.

RESULTS

A total of 486 patients [384 females (79.01%) and 102 males (20.99%)] who underwent thyroid FNAB were included in the final analysis. The mean age of the cohort was 52.3 years (range: 18-85). Of these, 376 patients (77.37%) were under the age of 65, and 110 patients (22.63%) were 65 years or older.

Diagnostic Yield by Age Group

The overall diagnostic adequacy rate across the cohort was 81.69% (n=397), while 18.31% (n=89) of the FNAB samples were non-diagnostic. When stratified by age, the diagnostic

success rate was significantly higher in the <65 group (86.44%) compared to the ≥65 group (65.45%) (p<0.01). This finding indicates a clear decline in cytological adequacy with increasing age (Table 1).

Table 1. Diagnostic success rates by age and gender

Group	Diagnostic	Non-diagnostic	Diagnostic yield (%)	p-value
<65 years	325	51	86.44%	<0.01
≥65 years	72	38	65.45%	
Female	320	64	83.33%	0.083
Male	77	25	75.49%	

To further evaluate the independent effects of age and gender on diagnostic adequacy, a binary logistic regression analysis was performed using diagnostic yield (diagnostic vs. non-diagnostic) as the dependent variable. The model included age (continuous) and gender (male=1, female=2) as covariates. The overall model was statistically significant (X^2=18.59, df=2, p<0.001), with a pseudo-R^2 value of 0.040, indicating that the model explained approximately 4% of the variance in diagnostic success. Increasing age was significantly associated with a reduced likelihood of diagnostic adequacy (β=–0.0376, p<0.001; OR=0.963), suggesting that each additional year of age decreased the odds of a diagnostic result by approximately 3.7%. Gender, however, was not found to be a statistically significant predictor in the model (β=0.308, p=0.265; OR=1.36), indicating that the apparent differences in diagnostic success between males and females may be confounded by age.

Diagnostic Yield by Gender

When evaluating diagnostic yield based on gender, 83.3% of female patients had diagnostic results compared to 75.5% of male patients. Although diagnostic adequacy was higher in females, this difference did not reach statistical significance (p=0.083). However, within the ≥65 age group, diagnostic success was notably lower in males (53.12%) than in females (70.51%), suggesting a potential compounded effect of age and gender.

Distribution of Cytopathological Results

Among the diagnostic cases (n=397), the majority were benign (n=354, 89.17%), while 22 samples (5.54%) were malignant, and 21 (5.29%) were classified as AUS/FLUS. A detailed analysis showed that:

- In the <65 group, benign diagnoses accounted for 89.23%, malignancy for 6.15%, and AUS/FLUS for 4.62%.
- In the ≥65 group, benign results were 88.89%, malignancy 2.78%, and AUS/FLUS 8.33%, indicating a higher rate of indeterminate results in older patients.

These findings are summarized in Table 2.

Table 2. Cytopathology distribution by age group

Age group	Benign	Malignant	AUS/FLUS	Total diagnostic samples
<65 years	290 (89.23%)	20 (6.15%)	15 (4.62%)	325
≥65 years	64 (88.89%)	2 (2.78%)	6 (8.33%)	72

AUS/FLUS: Atypia of undetermined significance/follicular lesion of undetermined significance

These results suggest that both advanced age and male gender-particularly in combination-may be associated with reduced diagnostic adequacy in thyroid FNAB. In contrast, cytopathological distribution patterns (benign, malignant, AUS/FLUS) appear largely consistent across age groups, with the notable exception of a higher atypia rate in elderly patients.

DISCUSSION

This study aimed to evaluate the diagnostic adequacy of thyroid FNAB in different age and gender groups. Our results revealed that patients aged 65 years and older had significantly lower diagnostic adequacy compared to those under 65, a finding that is consistent with prior literature suggesting age-related challenges in FNAB performance.[11] The diagnostic yield among elderly patients was 65.45%, which represents a substantial reduction compared to the 86.44% yield in younger individuals. Several factors may explain this discrepancy, including increased stromal fibrosis, reduced cellularity, or degenerative changes in thyroid tissue with aging, which make aspiration and cytological interpretation more difficult.[10]

Although gender alone did not show a statistically significant effect on diagnostic adequacy, female patients had a numerically higher diagnostic yield than males (83.3% vs. 75.5%). Interestingly, in the subgroup analysis of patients aged 65 and above, the diagnostic yield was considerably lower in males (53.12%) than in females (70.51%). These findings align with previous studies suggesting that male patients, especially in older age groups, may present additional technical challenges in FNAB, or may have different nodule characteristics influencing biopsy outcomes.[12]

Regarding cytopathological distribution, benign results were predominant across all age groups, in line with epidemiological data indicating that the majority of thyroid nodules are non-neoplastic.[2] Interestingly, the rate of AUS/FLUS was higher in elderly patients (8.33% vs. 4.62%), which may reflect either increased diagnostic uncertainty in this population or intrinsic biological differences in nodule behavior. In contrast, the malignancy rate was slightly lower in older adults (2.78% vs. 6.15%), although this may also reflect the lower overall diagnostic yield and higher rate of non-diagnostic results in this subgroup.

Previous studies have reported similar trends, with age-related stromal and cellular changes contributing to reduced FNAB efficacy.[14,15] Furthermore, the increase in AUS/FLUS rates among elderly patients observed in this study parallels findings in earlier reports, suggesting an increased likelihood of indeterminate cytology in this group.

Our logistic regression analysis confirmed that age is an independent and statistically significant predictor of diagnostic adequacy in thyroid FNAB. Specifically, increasing age was associated with a progressive decline in the odds of obtaining a diagnostic result, even after controlling for gender. These findings support the hypothesis that age-related changes in thyroid tissue-such as increased fibrosis, decreased cellularity, and altered architecture-contribute to reduced sample adequacy in elderly individuals. In contrast,

gender did not independently affect diagnostic performance, aligning with previous reports that any observed differences by sex may be mediated by age distribution rather than intrinsic biological or anatomical differences. This reinforces the need for adjusted biopsy strategies in elderly populations, irrespective of gender, to improve diagnostic yield.

Our findings have important clinical implications. First, the reduced diagnostic adequacy in elderly patients highlights the need for tailored strategies to improve FNAB success in this group. These may include the use of ultrasound guidance by experienced operators, repeated aspirations, or considering core needle biopsy in selected cases. Second, the observed trends underscore the importance of interpreting cytological findings within the context of patient age and gender, particularly when determining the need for surgery or additional diagnostic workup.[14]

The second FNAB does not increase the likelihood of diagnosing malignancy in old aged patients.[19] Evidences revealed the effectiveness of core needle biopsy in reducing inconclusive results and improving the diagnostic performance of thyroid nodules initially diagnosed as AUS/FLUS by FNAB.[20] In patients over 65 years of age, core biopsy may be considered as an option instead of repeat FNAB.

Limitations

This study has several limitations. First, its retrospective design inherently limits the ability to control for potential confounding variables such as nodule size, echogenicity, and operator variability. Second, the study was conducted at a single tertiary center, which may affect the generalizability of the results to other clinical settings. Third, we did not include long-term follow-up or histopathological confirmation for all cases, which limits our ability to assess the true false-negative or false-positive rates. Additionally, although data were stratified by age and gender, we did not account for potential interaction effects between other comorbidities (e.g., diabetes, autoimmune thyroiditis) and biopsy outcomes. Additionally, the lack of data regarding nodule-specific features such as ultrasound characteristics (e.g., echogenicity, composition, margins) and thyroid imaging reporting and data system (TI-RADS) scores represents a significant limitation. These variables are known to influence FNAB outcomes and may have provided more granular insight into diagnostic adequacy.

Despite these limitations, the relatively large sample size and systematic evaluation across clinically relevant subgroups strengthen the validity of our conclusions.

CONCLUSION

This study demonstrates that the diagnostic adequacy of thyroid FNAB significantly decreases in patients aged 65 years and older. While gender alone does not appear to be a statistically significant determinant, elderly males show the lowest diagnostic success rates, suggesting a compounded effect of age and sex on biopsy performance. Additionally, the higher prevalence of atypia in the elderly population may contribute to increased diagnostic ambiguity and influence clinical decision-making. These findings emphasize the need for age-adapted strategies to enhance diagnostic yield in older adults, including optimized sampling techniques and operator training. Future prospective and multi-center studies are warranted to confirm these results and explore underlying mechanisms.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was carried out with the permission of the Dr. Abdurrahman Yurtaslan Ankara Oncology SUAM Non-interventional Clinical Researches Ethics Committee (Date: 22.05.2025, Decision No: 2025-05/74).

Informed Consent

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

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

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

Author Contributions

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

REFERENCES

1. Gharib H, Papini E, Garber JR, et al. American Association of Clinical Endocrinologists, American College of Endocrinology, and Associazione Medici Endocrinologi Medical guidelines for clinical practice for the diagnosis and management of thyroid nodules-2016 update. *Endocr Pract.* 2016;22(5):622-639. doi:10.4158/EP161208.GL
2. 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. doi:10.1089/thy.2015.0020
3. Mazzaferri EL. Management of a solitary thyroid nodule. *N Engl J Med.* 1993;328(8):553-559. doi:10.1056/NEJM199302253280807
4. Guth S, Theune U, Aberle J, Galach A, Bamberger CM. Very high prevalence of thyroid nodules detected by high frequency (13 MHz) ultrasound examination. *Eur J Clin Invest.* 2009;39(8):699-706. doi:10.1111/j.1365-2362.2009.02162.x
5. Poma AM, Macerola E, Basolo A, et al. Fine-needle aspiration cytology and histological types of thyroid cancer in the elderly: evaluation of 9070 patients from a single referral centre. *Cancers (Basel).* 2021;13:4. doi:10.3390/cancers13040907
6. Kwong N, Medici M, Angell TE, et al. The influence of patient age on thyroid nodule formation, multinodularity, and thyroid cancer risk. *J Clin Endocrinol Metab.* 2015;100(12):4434-4440. doi:10.1210/jc.2015-3100
7. Ospina NS, Papaleontiou M. Thyroid Nodule evaluation and management in older adults: a review of practical considerations for clinical endocrinologists. *Endocr Pract.* 2021;27(3):261-268. doi:10.1016/j.eprac.2021.02.003
8. Cibas ES, Ali SZ, Conference NCITFSotS. The Bethesda system for reporting thyroid cytopathology. *Am J Clin Pathol.* 2009;132(5):658-665. doi:10.1309/AJCPPHLWMI3JV4LA

9. Baloch ZW, LiVolsi VA, Asa SL, et al. Diagnostic terminology and morphologic criteria for cytologic diagnosis of thyroid lesions: a synopsis of the National Cancer Institute Thyroid Fine-Needle Aspiration State of the Science Conference. *Diagn Cytopathol*. 2008;36(6):425-437. doi: 10.1002/dc.20830
10. Frates MC, Benson CB, Charboneau JW, et al. Management of thyroid nodules detected at US: Society of Radiologists in Ultrasound consensus conference statement. *Radiology*. 2005;237(3):794-800. doi:10.1148/radiol.2373050220
11. Belfiore A, Giuffrida D, La Rosa GL, et al. High frequency of cancer in cold thyroid nodules occurring at young age. *Acta Endocrinol (Copenh)*. 1989;121(2):197-202. doi:10.1530/acta.0.1210197
12. Yoon JH, Lee HS, Kim EK, Moon HJ, Kwak JY. Malignancy risk stratification of thyroid nodules: comparison between the thyroid imaging reporting and data system and the 2014 American Thyroid Association management guidelines. *Radiology*. 2016;278(3):917-924. doi:10.1148/radiol.2015150056
13. Bongiovanni M, Spitale A, Faquin WC, Mazzucchelli L, Baloch ZW. The Bethesda system for reporting thyroid cytopathology: a meta-analysis. *Acta Cytol*. 2012;56(4):333-339. doi:10.1159/000339959
14. Trimboli P, Giovanella L. Reliability of core needle biopsy as a second-line procedure in thyroid nodules with an indeterminate fine-needle aspiration report: a systematic review and meta-analysis. *Ultrasonography*. 2018;37(2):121-128. doi:10.14366/usg.17066
15. Rossi ED, Bizzarro T, Martini M, et al. The role of fine-needle aspiration in the thyroid nodules of elderly patients. *Oncotarget*. 2016;7(11):11850-11859. doi:10.18632/oncotarget.7643
16. Jeong EJ, Chung SR, Baek JH, Choi YJ, Kim JK, Lee JH. A comparison of ultrasound-guided fine needle aspiration versus core needle biopsy for thyroid nodules: pain, tolerability, and complications. *Endocrinol Metab (Seoul)*. 2018;33(1):114-120. doi:10.3803/EnM.2018.33.1.114
17. Shin JH. Strategies for safe and effective core needle biopsy of thyroid nodules with macrocalcification. *Int J Thyroidol*. 2023;16(2):195-199.
18. Aysan E, Guler B, Kiran T, Idiz UO. Core needle biopsy in the diagnosis of thyroid nodules. *Am Surg*. 2023;89(12):5170-5174. doi:10.1177/00031348221142570
19. Atmış VEÖ, B. Comparison of the results of thyroidectomy and second fine-needle aspiration biopsy of the old age group with the previous Bethesda III group: is the second biopsy necessary. *Erciyes Med J*. 2020;42(1):25-29. doi:10.14744/etd.2019.80378
20. Yim Y, Baek JH. Core needle biopsy in the management of thyroid nodules with an indeterminate fine-needle aspiration report. *Gland Surg*. 2019;8(Suppl 2):S77-S85. doi:10.21037/gs.2018.09.07

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1690010
J Health Sci Med. 2025;8(5):759-764

Evaluating the Turkish validity and reliability of the Mini Z 2.0 Clinician Worklife Survey among physicians

Nazmiye Kaçmaz Ersü[1], Batuhan Bakırarar[2]

[1]Department of Family Medicine, Efeler Family Health Center, Ministry of Health Aydın Provincial Health Directorate, Aydın, Turkiye
[2]Department of Process Development, ADO Group, Antalya, Turkiye

Cite this article as: Kaçmaz Ersü N, Bakırarar B. Evaluating the Turkish validity and reliability of the Mini Z 2.0 Clinician Worklife Survey among physicians. *J Health Sci Med.* 2025;8(5):759-764.

Received: 03.05.2025 ◆ **Accepted**: 19.07.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: Physicians face stress, job dissatisfaction, and burnout in their careers. There are many scales that assess burnout. Among them, Maslach Burnout Inventory is a standard tool used to assess burnout. The Mini Z 2.0 Clinician Worklife Survey was developed as a new, easy-to-use tool for identifying the factors which cause burnout, job dissatisfaction, and stress. Maslach Burnout Inventory can be utilized in the general population whereas the Mini Z 2.0 survey assesses the worklife of physicians. In Turkiye, there are no scales which can assess all those factors in the worklife practically and collectively among physicians. This study therefore aimed to evaluate the Turkish validity and reliability of the Mini Z 2.0 Clinician Worklife Survey and contribute to the literature.

Methods: The study was conducted with 221 physicians. The construct validity was determined using confirmatory factor analysis. Test-retest reliability and internal consistency were assessed using Spearman's correlation coefficient and Cronbach's alpha, respectively. The concurrent validity was determined by testing correlations between Mini Z 2.0 and Maslach Burnout Inventory subscales were met using Spearman's correlation coefficient.

Results: The confirmatory factor analysis showed that the scale has two factors. The Spearman's correlation coefficient results varied between 0.753 and 0.858. Cronbach's alpha values of the scale and subscales were found to be between 0.881 and 0.942 in the internal consistency analysis. Floor/ceiling effects were considered not to be present.

Conclusion: It was found that the Turkish version of the Mini Z 2.0 Clinician Worklife Survey is valid and reliable.

Keywords: Burnout, job dissatisfaction, stress, Mini Z 2.0

INTRODUCTION

Today, several research studies address the challenges of worklife. Work overload, long working hours, problems with fellow employees and administrators, and conditions of the workplace cause stress among employees.[1,2] When individuals come up against work demands that are not compatible with their knowledge and abilities and challenge their coping skills, their reactions lead to job stress.[3] Such circumstances reduce job satisfaction, therefore increasing the likelihood of physical and psychological problems, and burnout among employees.[4,5]

Physicians represent one of the occupational groups that frequently face stress, job dissatisfaction, and burnout in their careers.[5] Multitude of their responsibilities, work overload, long and irregular working hours, financial problems, lack of time they make for themselves and their families, conflicts with administrators, fellow employees, and patients, medical errors and litigations, and challenges with the electronic health records lead to stress among physicians.[6] Job dissatisfaction and burnout impact physicians' approach toward their patients. The burnout syndrome causes longer recovery durations for patients, increased number of medical errors, and decreased patient satisfaction.[7] It may increase the risk of making medical errors and reduce job satisfaction, consequently causing physicians to quit their jobs. The syndrome may also hinder safe, accessible, and low-cost healthcare services and decrease the quality of healthcare.[8] Impacts of this situation on patients have alarming repercussions on patients and the society, and it is addressed as a public health concern.[9]

There are many scales that assess burnout. Among them, Maslach Burnout Inventory (MBI) is a standard tool used to assess burnout.[10] The Mini Z 2.0 Clinician Worklife Survey was developed as a new, easy-to-use tool for identifying the factors which cause burnout, job dissatisfaction, and stress among physicians.[11] MBI can be utilized in the general population whereas the Mini Z 2.0 survey assesses the worklife of physicians.

With the Mini Z 2.0 Clinician Worklife Survey, factors leading to stress among physicians in their work lives, their levels of job satisfaction and burnout can be evaluated rapidly and with fewer questions. It is stated that the results of the Mini Z 2.0 Clinician Worklife Survey can be utilized to guide

Corresponding Author: Batuhan Bakırarar, batuhan_bakirarar@hotmail.com

interventions that aim to ensure physician wellbeing and improve performance of the healthcare system.[9,12]

In Turkiye, there are no scales which can assess all those factors in the worklife practically and collectively among physicians. This study therefore aimed to evaluate the Turkish validity and reliability of the Mini Z 2.0 Clinician Worklife Survey and contribute to the literature.

METHODS

Translation and Cultural Adaptation

For the validity and reliability study of the Mini Z 2.0 Clinician Worklife Survey, Dr. M. Linzer who developed the survey and C. Ergin who conducted the Turkish validity and reliability study of the MBI, which is the other scale utilized in the study, were contacted via e-mail. Required permissions were received for using the scales in the present study. The study was carried out with the permission of the Ankara University Human Researches Ethics Committee (Date: 13.04.2022, Decision No: İ04-191-22). This investigation was carried out in compliance with the principles outlined in the Declaration of Helsinki. Each participant accepted informed consent for the use of their data in the research.

The Turkish adaptation of the Mini Z 2.0 Clinician Worklife Survey was performed by three linguists and two subject matter experts to ensure linguistic validity. Two forward translators worked independently (double-blind) and translated the original Mini Z 2.0 Clinician Worklife Survey into Turkish. The Turkish version was back translated into English by two independent translators and compared to the original scale. These two independent translators are the independent individuals unfamiliar with the original scale. The Mini Z 2.0 Clinician Worklife Survey was finalized based on the feedback of a Turkish linguist who reviewed the English and Turkish meanings of the scale items.

After the translation process, the scale was first applied to a group of 40 people and the answers to the questions were analyzed to check the comprehensibility and applicability of the questions in the scale. Since the answers were logical and evenly distributed, it was concluded that the scale was applicable to the target group.

Work Items

Mini Z 2.0 Clinician Worklife Survey: The Mini Z 2.0 Clinician Worklife Survey was developed by Mark Linzer, MD, and his team at Hennepin Healthcare. The Mini Z Survey was derived from the "Z" Clinician Survey (for Zero Burnout Program). The survey was adapted from prior work performed for the Physician Worklife Survey and the minimizing error and maximizing outcomes (MEMO) study.[13,14] It consists of 10 items rated using a 5-point Likert scale. The Mini Z 2.0 Survey has two subscales. Subscale-1 is used to assess "supportive work environment" (satisfaction, burnout, value alignment with leaders, and teamwork) whereas subscale-2 is used to assess "work pace and electronic medical record (EMR) stress" (stress, documentation time pressure, home EMR use and EMR proficiency). Questions 1-5 are rated within subscale-1. Higher scores from this subscale refer to a more supportive work environment. Questions 6-10 are rated within subscale-2, and higher scores from this subscale represent reasonable work pace and manageable EMR stress. Higher total scores obtained from the entire survey indicate the presence of better working conditions.

Maslach Burnout Inventory: Maslach Burnout Inventory which was developed by Maslach and Jackson[15] and of which Turkish validity and reliability study was conducted by Ergin[16] was utilized as the reference scale in the present study. The original form rated using a 7-point Likert Scale was rated again by Ergin using a 5-point Likert Scale. This scale consists of 22 items and three subscales in total. Among those subscales, emotional exhaustion (EE) is composed of nine items, depersonalization (DP) is composed of five items, and personal accomplishment (PA) is composed of eight items. EE and DP subscales are negative, and PA subscale is positive in the inventory used in the present study. Higher levels of burnout are indicated by higher scores from the emotional exhaustion and depersonalization subscales, and lower scores from the personal accomplishment subscale. Lower scores from the emotional exhaustion and depersonalization subscales and higher scores from the personal accomplishment subscale refer to lower levels of burnout.

Data Collection

The first question of the survey is about approval concerning that the participants took part in the study voluntarily. The survey form is comprised of questions about sociodemographic information, occupational data, and items of the Mini Z Clinician Worklife Survey-Turkish Form and the Maslach Burnout Inventory. The third question of the Mini Z Clinical Worklife Survey was added as a control question to prevent participants from responding randomly. The survey form was sent to the physicians working at the departments of Basic Medical Sciences, Internal Medical Sciences, and Surgical Medical Sciences via e-mail. The form was completed online by a total of 233 physicians; 12 physicians who did not provide the desirable response to the control question were excluded from the study, and the study was conducted with 221 physicians. The survey forms were sent to the same physicians again one month later. 189 of 221 physicians completed the survey again.

Statistical Analysis

The data were analyzed on SPSS 11.5 and AMOS 24.0 software. As descriptive statistics, mean±standard deviation and median (minimum-maximum) were utilized for quantitative variables, and number of persons (percentage) were used for qualitative variables. For quantitative variables, Mann-Whitney U test was performed to see whether there was a statistically significant difference between categories of the qualitative variable with two categories. For qualitative variables, Kruskal Wallis H test was used to find out whether there was a statistically significant difference between categories of the qualitative variable with more than two categories since the assumptions of normality could not be met. Confirmatory factor analysis, linear regression, and Spearman's rank correlation coefficient were used for construct validity, predictive validity, and concurrent validity,

respectively. Spearman's rank correlation coefficient was also utilized for the reliability of the test-retest. Cronbach's Alpha was also calculated for the reliability. Mann-Whitney U test was performed for Item Discrimination Index. Statistical significance level was accepted to be 0.05.

RESULTS

Validity

Content validity: Content validity in the study was evaluated by 15 experts categorizing 10 questions with a triple rating system as being "essential," "useful, but not essential," or "not necessary". The table value of the smallest content validity ratio (CVR) for 15 experts is 0.49. CVR is calculated with the equation CVR=[E/(N/2)]-1; where E: number of experts indicating "essential", and N: total number of experts. Based on the CVR values in Table 1, it was concluded that all items should be retained in the item pool since CVR values of all items are greater than 0.49.

Table 1. CVR and CVI values of items

Items	Essential	Useful, but not essential	Not necessary	CVR	CVI
I1	14	1	0	0.867	0.813
I2	13	2	0	0.733	
I3	13	2	0	0.733	
I4	15	0	0	1.000	
I5	13	2	0	0.733	
I6	13	1	1	0.733	
I7	14	1	0	0.867	
I8	14	1	0	0.867	
I9	13	2	0	0.733	
I10	14	1	0	0.867	

CVR: Content validity ratio, CVI: Content Validity Index

Content Validity Index (CVI) for the scale equals to the mean CVR across items retained in the item pool. In the present study, it was found CVR=(0.867+0.733+0.733+…+0.867)/10=0.813. As CVI=0.813>0.67 (cut-off value), the scale was concluded to be statistically significant.

Logical validity: Since the scale measured the most important components accurately and properly and provided the desired information most accurately, its logical validity was established.

Validity as against a reference: How well the scale used in a study assesses the attribute in question compared to the gold standard is tested in the presence of a scale known and used as the gold standard in the literature. To that end, the two most common validity tests conducted with a reference are:

- Concurrent validity, and
- Predictive validity.

Concurrent validity: The Mini Z 2.0 Clinician Worklife Survey used in the present study has two subscales which are "Supportive Work Environment" and "Work Pace and EMR Stress". Maslach Burnout Inventory used as the gold standard has 3 subscales which are "Emotional Exhaustion", "Personal Accomplishment", and "Depersonalization". Results concerning the correlation between the scales used in the study were shown in Table 2.

Table 2. Correlation between scales

Scales		Emotional exhaustion	Personal accomplishment	Depersonalization
Supportive work environment	Correlation coefficient	-0.823	0.789	-0.770
	p-value	<0.001	<0.001	<0.001
Work pace and EMR stress	Correlation coefficient	-0.804	0.753	-0.769
	p-value	<0.001	<0.001	<0.001
Mini Z total score	Correlation coefficient	-0.858	0.816	-0.815
	p-value	<0.001	<0.001	<0.001

EMR: Electronic medical record

Coefficients of the correlation between the supportive work environment subscale and the emotional exhaustion, personal accomplishment, and depersonalization subscales were found to be -0.823, 0.789, and -0.770, respectively. Coefficients of the correlation between the Work Pace and EMR Stress subscale and the Emotional Exhaustion, Personal Accomplishment, and Depersonalization subscales were found to be -0.804, 0.753, and -0.769, respectively. Coefficients of the correlation between the Mini Z total score and the Emotional Exhaustion, Personal Accomplishment, and Depersonalization subscales were found to be -0.858, 0.816, and -0.815, respectively. These results suggest that concurrent validity for the Mini Z 2.0 Clinician Worklife Survey and its subscales is adequate.

Predictive validity: For the present study, the emotional exhaustion, personal accomplishment, and depersonalization subscales were considered dependent variables, and the Supportive Work Environment and work pace and EMR stress subscales, and the Mini Z total score were considered independent variables. Univariate linear regression results are given in Table 3. As can be seen, the results were significant for all variables. The results mean that Mini Z and its subscales can be used instead of the gold standard.

Testing of factorability: Kaiser-Meyer-Olkin (KMO) test was used to establish whether the sample examined in the factor analysis was fit for the analysis. A KMO measure of over 0.80 is expected for a good factor analysis.[17] The KMO value of 0.936 was found in the present study, and the sample was concluded to be adequate for the factor analysis. In addition, Bartlett's test of sphericity was performed to see whether the correlation matrix was fit for the factor analysis, and the result was found to be significant (p<0.001).

Construct validity: A confirmatory factor analysis was used in the present study since the Turkish validity and reliability study was performed for a scale for which validity and reliability was established in its original language. Factor loadings of the items in the scale were shown in Table 4 by subscales. As can be seen in the table, factor loadings of all items were found to be over 0.7, and construct validity was

Table 3. Univariate linear regression results							
						95% CI for β	
Independent variables	**Dependent variables**	**β**	**SE**	**p-value**	**R^2**	**Lower**	**Upper**
Supportive work environment	Emotional exhaustion	-1.383	0.064	<0.001	0.679	-1.509	-1.256
	Personal accomplishment	1.186	0.059	<0.001	0.652	1.070	1.301
	Depersonalization	-0.823	0.042	<0.001	0.633	-0.907	-0.740
Work pace and EMR stress	Emotional exhaustion	-1.411	0.074	<0.001	0.623	-1.557	-1.265
	Personal accomplishment	1.180	0.069	<0.001	0.569	1.043	1.316
	Depersonalization	-0.855	0.047	<0.001	0.602	-0.947	-0.762
Mini Z total score	Emotional exhaustion	-0.768	0.033	<0.001	0.717	-0.833	-0.704
	Personal accomplishment	0.651	0.031	<0.001	0.672	0.591	0.712
	Depersonalization	-0.461	0.021	<0.001	0.680	-0.504	-0.419

SE: Standard error, CI: Confidence interval, EMR: Electronic medical record

established for the subscales of the scale. A Chi-square/degree of freedom (X^2/df) value below 3 is considered adequate; this value was found to be 2.555 in the present study. The acceptable value for GFI, CFI, and TLI is 0.9; the values found in the present study were 0.934 for GFI, 0.971 for CFI, and 0.959 for TLI. The acceptable value for RMSEA is 0.08; this value was found to be 0.08 in the present study. In summary, construct validity was established in terms of the criteria used for validity.[18] The path diagram for construct validity is given in Figure.

Table 4. Item factor loadings by subscales		
Items	**Supportive work environment**	**Work pace and EMR stress**
1	0.885	
2	0.837	
3	0.877	
4	0.812	
5	0.820	
6		0.812
7		0.853
8		0.653
9		0.716
10		0.724

EMR: Electronic medical record

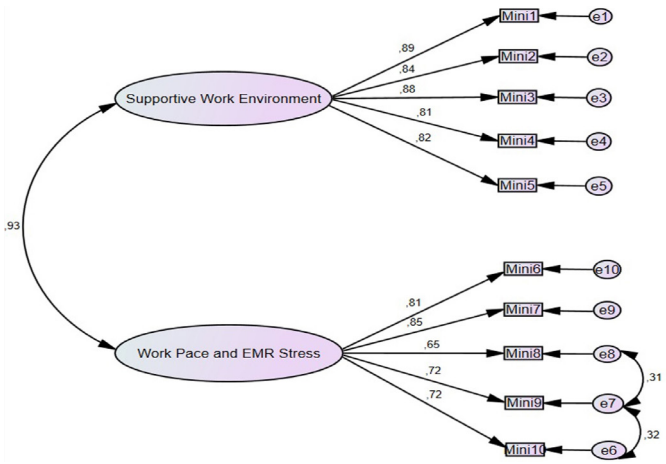

Figure. A confirmatory factor model for Mini Z with factor loadings

Reliability

Test-retest method: Spearman's rank correlation coefficients were found to be 0.988 for the Supportive Work Environment subscale, 0.987 for the Work Pace and EMR Stress subscale, and 0.991 for the Mini Z total score. The scale and its subscales were concluded to be reliable based on these results.

Cronbach's alpha: Cronbach's Alphas were calculated to be 0.926 for the Supportive Work Environment subscale, 0.881 for the Work Pace and EMR Stress subscale, and 0.942 for the Mini Z total score. Therefore, the scale and its subscales were concluded to have high reliability.

Comparison of top-bottom 27% groups (Item Discrimination Index): For the Supportive Work Environment, Work Pace and EMR Stress subscales, and the Mini Z total score, a significant difference was found between the top and bottom 27% groups ($p<0.001$, $p<0.001$, and $p<0.001$, respectively). The scale was therefore concluded to have an adequate item distinction index.

Examination of ceiling/floor effect in the scale: In the study, there were 11 (5.0%) participants who scored 5 and 1 (0.5%) participant who scored 25. The lowest and highest possible scores from the Work Pace and EMR subscale were 5 and 25, respectively. 7 (3.2%) participants scored 5, and 1 (0.5%) participant scored 25 in the study. The lowest and highest possible scores from the entire scale were 10 and 50, respectively. Whereas 6 (2.7%) participants scored 10, none of the participants scored 50 in the study. These results suggest that there is no ceiling/floor effect on the scale and its subscales.

Descriptive Statistics

The scale and its subscales were provided in Table 5 for the physicians who participated in the study. No significant difference was found in terms of scale scores for gender and years of practice. There was a significant difference in at least one scale score for all other variables.

DISCUSSION

The Mini Z 2.0 is a simple, newly developed scale recognized as a powerful tool to assess physician burnout and the

Table 5. Comparisons of descriptive variables for subscale scores and total score

		Scores								
		Supportive work environment			Work pace and EMR stress			Mini Z total score		
Variables		Mean±SD	Median (min-max)	p-value	Mean±SD	Median (min-max)	p-value	Mean±SD	Median (min-max)	p-value
Gender	Female	12.43±4.67	12.00 (5.00-22.00)	0.755[a]	13.30±4.29	13.00 (5.00-25.00)	0.083[a]	25.72±8.38	25.00 (10.00-47.00)	0.232[a]
	Male	12.40±5.19	11.00 (5.00-25.00)		12.38±4.90	12.00 (5.00-24.00)		24.78±9.77	24.00 (10.00-48.00)	
Years of practice	<5 years	11.28±3.98	10.00 (5.00-21.00)	0.137[a]	12.35±3.68	13.00 (6.00-22.00)	0.666[a]	23.63±7.02	23.00 (14.00-43.00)	0.312[a]
	≥5 years	12.69±5.11	12.00 (5.00-25.00)		12.93±4.84	13.00 (5.00-25.00)		25.61±9.54	25.00 (10.00-48.00)	
Department	Basic medical sciences	14.27±5.61	14.00 (5.00-25.00)	<0.001[b]	14.98±5.01	15.00 (5.00-25.00)	<0.001[b]	29.26±10.15	28.00 (10.00-48.00)	<0.001[b]
	Internal medical sciences	12.40±4.62	12.00 (5.00-23.00)		12.55±4.24	13.00 (5.00-22.00)		24.95±8.26	25.00 (10.00-45.00)	
	Surgical medical sciences	10.31±3.69	10.00 (5.00-23.00)		10.79±3.78	10.00 (5.00-24.00)		21.10±7.27	20.00 (10.00-47.00)	
Exposure to violence	No	13.19±4.94	12.00 (5.00-25.00)	<0.001[a]	13.61±4.45	13.00 (5.00-24.00)	<0.001[a]	26.81±8.95	26.00 (10.00-48.00)	<0.001[a]
	Yes	10.36±4.32	9.00 (5.00-22.00)		10.72±4.50	10.00 (5.00-25.00)		21.08±8.30	19.00 (10.00-47.00)	
Making enough time for family	No	11.00±4.06	10.00 (5.00-23.00)	0.001[b]	11.18±3.77	10.00 (6.00-19.00)	<0.001[b]	22.18±7.28	20.00 (13.00-39.00)	<0.001[b]
	Yes	14.26±5.38	13.00 (5.00-25.00)		14.58±5.15	14.00 (5.00-25.00)		28.84±10.12	28.00 (10.00-48.00)	
	Sometimes	11.88±4.70	12.00 (5.00-22.00)		12.53±4.27	13.00 (5.00-23.00)		24.41±8.49	24.00 (10.00-45.00)	
Making time for oneself and one's hobbies	No	11.50±4.20	10.50 (5.00-23.00)	0.059[b]	11.67±3.84	11.50 (5.00-22.00)	0.003[b]	23.17±7.48	23.00 (10.00-43.00)	0.010[b]
	Yes	13.00±5.35	11.00 (5.00-25.00)		13.35±5.07	11.00 (5.00-25.00)		26.35±9.98	23.00 (10.00-48.00)	
	Sometimes	13.41±5.48	13.00 (5.00-23.00)		14.15±5.08	14.00 (5.00-24.00)		27.56±10.20	27.00 (10.00-47.00)	
Problem taking annual leaves	No	14.31±5.64	13.50 (5.00-25.00)	<0.001[b]	14.64±5.08	14.50 (5.00-24.00)	<0.001[b]	28.95±10.23	28.00 (10.00-48.00)	<0.001[b]
	Yes	10.79±4.05	10.00 (5.00-22.00)		11.26±4.01	10.00 (5.00-25.00)		22.04±7.62	20.00 (10.00-47.00)	
	Sometimes	12.51±4.23	12.00 (5.00-22.00)		12.90±3.95	13.00 (5.00-23.00)		25.41±7.66	25.00 (11.00-45.00)	

EMR: Electronic medical record, SD: Standard deviation, Min: Minimum, Max: Maximum, a: Mann-Whitney U test, b: Kruskal Wallis H test

workplace-related stress factors. Validity and reliability of Mini Z 2.0's Turkish version was evaluated in the study.

It is expected that Turkish version of the Mini Z 2.0 will play a key role in assessing physician wellbeing and burnout in Turkiye. Physician burnout is a common issue in Turkiye, and routine assessments are required to prevent it.[19] MBI is recognized as the gold standard in assessing physician burnout;[20,21] however, Turkish version of the Mini Z has several advantages over MBI. Firstly, MBI is a long questionnaire composed of 22 items; on the contrary, the Mini Z has 10 items and is quite easy to apply. Secondly, MBI can be used for a fee whereas the Mini Z is free to use. Lastly, MBI assesses burnout only; on the other hand, the Mini Z can also assess satisfaction and other consequences of stress factors at workplace. Assessment of stress factors would help guide the interventions aiming to improve physician welfare in the healthcare system.[19,21]

It was reported in the original study that construct validity was established with a two-factor structure via exploratory factor analysis (EFA).[22] In the study conducted by Nagasaki et al.,[12] it was reported that the two-factor structure had good fit. They found a CFI value of 0.839, a TLI value of 0.762, and a RMSEA value of 0.148. Although the model with the best fit was reported to be the two-factor model, they stated that construct validity could not be established with the two-factor structure. In the present study, a confirmatory factor analysis was performed based on the two-factor structure in the original study. A CFI value of 0.971, a TLI value of 0.959, and a RMSEA value of 0.08 were found. Moreover, the correlation between the subscales of MBI, which is the gold standard, and the Mini Z 2.0 Survey and its subscales were investigated for concurrent validity, and correlation coefficients were found to be ranging between 0.753-0.858. Based on the totality of findings, it was concluded that the validity of the scale was established.

In assessments, a scale is considered more reliable as its Cronbach's alpha is closer to 1. In the original study of the scale, Cronbach's alphas were found to be 0.74 for the supportive work environment subscale, 0.72 for the Work Pace and EMR Stress subscale, and 0.80 for the Mini Z total score.[22] In the study performed by Nagasaki et al.,[12] only the Cronbach's alpha of the entire scale was calculated and found to be 0.80. In the present study, Cronbach's alphas were calculated to be 0.926 for the Supportive Work Environment subscale, 0.881 for the Work Pace and EMR Stress subscale, and 0.942 for the Mini Z total score. One of the methods used in the reliability analysis is to identify whether a responder will provide the same responses when the scale is repeated. The correlation coefficient evaluated to that end is expected to be positive and high. Test-retest correlation coefficients in the present study were found to be 0.988 for the supportive work environment subscale, 0.987 for the Work Pace and EMR Stress subscale, and 0.991 for the Mini Z total score. It was concluded from the results that the scale could be used reliably.

Limitations

The limitation of this study is that the scale was applied to the physicians as a scale. However, this limitation was eliminated by adding a control question in the Mini Z 2.0 survey to prevent responders from answering randomly. It was planned that the physicians who did not provide the desired answer to the control question "please mark the answer 'strongly agree' in this question" were excluded from the study. 12 physicians who did not provide the desired answer to the control question were excluded from the study.

CONCLUSION

The data obtained in the study suggest that Turkish version of the Mini Z 2.0 is a valid and reliable tool to assess physician burnout and the workplace-related stress factors. It is anticipated that the Mini Z 2.0 adapted to the Turkish language in this study can assess physician wellbeing and burnout to develop reformative policies.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was carried out with the permission of the Ankara University Human Researches Ethics Committee (Date: 13.04.2022, Decision No: İ04-191-22).

Informed Consent

Each participant accepted informed consent for the use of their data in the research.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

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

Author Contributions

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

Acknowledgement

We would like to thank the physicians who voluntarily participated in our research.

Scientific Responsibility Statement

The authors declare that they are responsible for the article's scientific content including study design, data collection, analysis and interpretation, writing, some of the main line or all of the preparation and scientific review of the contents and approval of the final version of the article.

REFERENCES

1. Maswadi N, Khader YS, Slaih AA. Perceived stress among resident doctors in Jordanian teaching hospitals: cross-sectional study. *JMIR Public Health Surveillan*. 2019;5(4):e14238. doi:10.2196/14238
2. Riley R, Buszewicz M, Kokab F, et al. Sources of work-related psychological distress experienced by UK-wide foundation and junior doctors: a qualitative study. *BMJ Open*. 2021;11(6):e043521. doi:10.1136/bmjopen-2020-043521
3. Occupational health: stress at the workplace. Available from: https://www.who.int/news-room/questions-and-answers/item/ccupational-health-stress-at-the-workplace. (Available at: 2022 November 3).
4. Bagheri Hosseinabadi M, Etemadinezhad S, Khanjani N, et al. Evaluating the relationship between job stress and job satisfaction among female hospital nurses in Babol: an application of structural equation modeling. *Health Promot Perspect*. 2018;8(2):102. doi:10.15171/hpp.2018.13
5. Linzer M, Smith CD, Hingle S, et al. Evaluation of work satisfaction, stress, and burnout among US internal medicine physicians and trainees. *JAMA Network Open*. 2020;3(10):e2018758-e2018758. doi:10.1001/jamanetworkopen.2020.18758
6. Carayon P, Cassel CK. Factors contributing to clinician burnout and professional well-being. In: Carayon P, Cassel CK (editors). Taking action against clinician burnout: a systems approach to professional well-being. 1st ed. Washington, DC, USA: Academical Word Press; 2019. pp. 81-126.
7. de Hert S. Burnout in healthcare workers: prevalence, impact and preventative strategies. *Local Region Anesthes*. 2020;13:171-183. doi:10.2147/LRA.S240564
8. Han R, Feng L, Wei L. Research on the influence of doctor-patient relationship on physicians' citizenship behavior. In: DEStech Transactions on Social Science Education and Human Science 2017. doi:10.12783/dtssehs/mess2016/9753
9. Olson K, Sinsky C, Rinne ST, Long T, Vender R et al. Cross-sectional survey of workplace stressors associated with physician burnout measured by the Mini-Z and the Maslach Burnout Inventory. *Stress Health*. 2019;35(2):157-175. doi:10.1002/smi.2849
10. Maslach C, Jackson SE, Leiter MP. Maslach Burnout Inventory: Third edition. In: Maslach C, Jackson SE, Leiter MP (editors). Evaluating stress: a book of resources. Lanham, MD, USA: Scarecrow Education; 1997. pp. 191-218.
11. Linzer M, Poplau S, Babbott S, et al. Worklife and wellness in academic general internal medicine: results from a national survey. *J General Int Med*. 2016;31(9):1004-1010. doi:10.1007/s11606-016-3720-4
12. Nagasaki K, Shikino K, Nishimura Y, et al. Translation, cultural adaptation and validation of the Mini-Z 2.0 Survey among Japanese physicians and residents. *Int Med*. 2021;60(15):2405-2411. doi:10.2169/internalmedicine.6749-203720-4
13. Williams ES, Konrad TR, Linzer M, et al. Refining the measurement of physician job satisfaction: results from the physician worklife survey. *Medical Care*. 1999;37(11):1140-1154.
14. Williams ES, Manwell LB, Konrad TR, Linzer M. The relationship of organizational culture, stress, satisfaction, and burnout with physician-reported error and suboptimal patient care: results from the MEMO study. *Health Care Manag Rev*. 2007;32(3):203-212. doi:10.1097/01.HMR.0000281626.28363.59
15. Maslach C, Jackson SE. The measurement of experienced burnout. *J Organizat Behav*. 1981;2(2):99-113. doi:10.1002/job.4030020205
16. Ergin C. Doktor ve hemşirelerde tükenmişlik ve Maslach Tükenmişlik Ölçeğin uyarlanması. VII. Ulusal Psikoloji Kongresi Bilimsel Çalışmaları El Kitabı. 1992:143-154.
17. Alpar R. Spor, sağlık ve eğitim bilimlerinden örneklerle uygulamalı istatistik ve geçerlik-güvenirlik. Detay Yayıncılık. 2020;5:493-604.
18. Çapık C. Geçerlik ve güvenirlik çalişmalarinda doğrulayici faktör analizinin kullanimi. *Anad Hemşir Sağlık Bil Derg*. 2014;17(3):196-205.
19. Wallace JE, Lemaire JB, Ghali WA. Physician wellness: a missing quality indicator. *Lancet*. 2009;374(9702):1714-1721. doi:10.1016/S0140-6736(09)61424-0
20. Bianchi R, Schonfeld IS, Laurent E. Burnout-depression overlap: a review. *Clin Psychol Rev*. 2015;36:28-41. doi:10.1016/j.cpr.2015.01.004
21. West CP, Dyrbye LN, Shanafelt TD. Physician burnout: contributors, consequences and solutions. *J Int Med*. 2018;283(6):516-529. doi:10.1111/joim.12752
22. Shimotsu S, Poplau S, Linzer M. Validation of a brief clinician survey to reduce clinician burnout. *J General Int Med*. 2015;30(2):79-80.

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1733877
J Health Sci Med. 2025;8(5):765-769

Survey study on volunteers participating in clinical phase trials of the national vaccine developed against COVID-19*

Zafer Sezer, Ahmet İnal, Seda Caymaz Yılmaz
Department of Pharmacology, Faculty of Medicine, Erciyes University Hakan Çetinsaya Good Clinical Practise and Research Center, Kayseri, Turkiye

Cite this article as: Sezer Z, İnal A, Caymaz Yılmaz S. Survey study on volunteers participating in clinical phase trials of the national vaccine developed against COVID-19. *J Health Sci Med.* 2025;8(5):765-769.

Received: 03.07.2025 ◆ **Accepted**: 21.07.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: This study aimed to evaluate the sociodemographic characteristics, motivations for participation, and experiences of volunteers who took part in the phase 1 and phase 2 clinical trials of Turkovac, the first national COVID-19 vaccine developed in Turkiye.

Methods: A total of 230 volunteers participated in the study. Data on their demographics and perspectives on clinical research were collected using a structured questionnaire.

Results: The majority of the volunteers were male (76.1%) and residing in urban areas (87.8%). Of all participants, 91.7% were involved in the phase 2 trial and 8.3% in the phase 1 trial. Most volunteers reported that their motivation for participation was based on trust in the national vaccine. Furthermore, a significant part of the participants evaluated their clinical trial experience positively and expressed willingness to participate in future studies.

Conclusion: The findings of this study may provide valuable insights for the planning of future clinical trials and the development of effective strategies to increase volunteer participation.

Keywords: Survey, COVID-19, volunteer, clinical trials, Turkovac, national vaccine

*This study Presented as an oral presentation at the 26th National and 1st International Pharmacology Congress, November 4-6, 2021, Online, Turkiye.

INTRODUCTION

The COVID-19 pandemic emerged in December 2019 in Wuhan, China, and rapidly evolved into a global public health crisis. The World Health Organization (WHO) declared COVID-19 a pandemic on March 11, 2020; this process has profoundly affected societies, healthcare systems, and economies worldwide.[1] In the early stages of the pandemic, measures such as quarantine, social distancing, mask-wearing, and travel restrictions were implemented to control the spread of the disease, significantly impacting individuals' lifestyles and overall health. Vaccines, one of the most effective methods of protection against COVID-19, have played a key role in controlling the pandemic. Innovative vaccine technologies, particularly mRNA-based vaccines, enabled a rapid and effective response to the crisis. The World Health Organization has regarded COVID-19 vaccines as critical tools not only in ending the current pandemic but also in preparing for future outbreaks. In this context, increasing global access to vaccines and building public trust in vaccination have become central goals of global health policies.[2,3]

In this context, Turkovac, the first nationally developed COVID-19 vaccine in Turkiye, has marked a significant milestone in the management of the pandemic. Turkovac is an inactivated whole-virion SARS-CoV-2 vaccine. The vaccine was well tolerated after administration. The most common side effects were pain at the injection site and headache.[4] Phase 1 and phase 2 clinical trials of this vaccine, developed in collaboration between Erciyes University and the Turkish Health Institutes Directorate, were conducted at the Hakan Çetinsaya Center for Good Clinical Practice and Research of Erciyes University.[4] Turkovac has served as an example that highlights the importance of national vaccine development capacity not only in Turkiye but also globally.[5,6]

Although COVID-19 is no longer considered a global emergency, this study remains relevant and important in terms of understanding volunteers' perspectives on clinical trials related to the development of new vaccines. The development of national vaccines such as Turkovac plays a critical role in ensuring both national and global health security in the face of potential future outbreaks. Therefore, the aim of this study is to identify the sociodemographic characteristics of volunteers who participated in the Turkovac trials, understand their reasons for participation in clinical research, and evaluate their experiences throughout the process. The findings obtained may contribute to the planning of future clinical trials and to efforts aimed at increasing volunteer participation. Under pandemic conditions, the rapid completion of phase studies is crucial for public health. This study also crucial for demonstrating public orientation.

Corresponding Author: Ahmet İnal, ainal@erciyes.edu.tr

METHODS

Study Volunteers and Questionnaire Design

This study was conducted with individuals who volunteered to participate in the phase 1 and phase 2 clinical trials of the Turkovac vaccine at the Hakan Çetinsaya Center for Good Clinical Practice at Erciyes University. The study was approved by the Erciyes University Non-interventional Clinical Researches Ethics Committee (Date: 03.03.2021, Decision No: 2021/149). All procedures were carried out in accordance with the ethical rules and the principles of the Declaration of Helsinki. Volunteers were informed that participation was entirely voluntary, and all participants signed an informed consent form prior to the administration of the questionnaire. Data were collected through face-to-face interviews between March 10 and April 10, 2021. A total of 230 volunteers were included in the study.

The questionnaire was designed to assess the demographic characteristics of the volunteers (gender, age, education level, place of residence, marital status, and occupation), their reasons for participating in vaccine trials, and any concerns experienced during the process. The questionnaire consisted of five-point Likert-Scale items, multiple-choice questions, closed-ended, and open-ended questions. The Likert-Scale items were rated as follows: 5: Strongly agree, 4: Agree, 3: Neutral, 2: Disagree, and 1: Strongly disagree. The questionnaire was developed by reviewing similar studies in the literature and was structured using scales and question sets from previous research, tailored to suit the purpose of this study.[7-11]

Statistical Analysis

The data analyses of the collected data were performed using IBM SPSS Statistics version 22.0 (IBM Corp., Armonk, NY, USA). The normality of data distribution was assessed using the Kolmogorov-Smirnov test and box plot graphs. For the comparison of numerical variables, the independent samples t-test was used when the data were normally distributed, while the Mann-Whitney U test was applied for non-normally distributed data. The chi-square test (X^2) was used to compare categorical variables. A p-value of ≤0.05 was considered statistically significant.

RESULTS

A total of 230 volunteers participated in this study, of whom 23.9% were female (n=55) and 76.1% were male (n=175). When the age distribution was examined, the largest group was in the 35-44 age range, accounting for 42.6% of the participants. Regarding educational status, 39.1% of the volunteers were university graduates (n=90), 28.7% were high school graduates (n=66), and 8.7% held a postgraduate degree (n=20). The vast majority of the volunteers were living in urban areas (87.8%, n=202). In terms of marital status, 65.2% were married (n=150) and 34.8% were single. The distribution of sociodemographic characteristics by phase 1 and phase 2 groups is presented in Table 1.

At the time of the survey, 91.7% of the volunteers (n=211) were part of the phase 2 vaccine trial group, while 8.3% (n=19) were in the phase 1 group.

Table 1. Distribution of volunteers in phase 1 and phase 2 groups according to sociodemographic characteristics

	Total (n=230)	Phase 1 (n=19)	Phase 2 (n=211)	p-value
Gender	Female: 55 (23.9%)	1 (5.3%)	54 (25.6%)	0.050
	Male: 175 (76.1%)	18 (94.7%)	157 (74.4%)	
Education	Primary education: 39 (17%)	4 (21.1%)	35 (16.6%)	0.632
	Secondary education:15 (6.5%)	1 (5.3%)	14 (6.6%)	
	High school: 66 (28.7%)	8 (42.1%)	58 (27.5%)	
	University: 90 (39.1%)	5 (26.3%)	85 (40.3%)	
	Postgraduate: 20 (8.7%)	1 (5.2%)	19 (9.0%)	
Marital status	Married: 150 (65.2%)	8 (42.1%)	142 (67.3%)	0.027
	Single: 80 (34.8v)	11 (57.9%)	69 (32.7%)	
Occupation	Employed: 153 (66.5%)	11 (57.9%)	142 (67.3%)	0.563
	Unemployed 77 (33.5%)	8 (42.1%)	69 (32.7%)	

Data are presented as number and percentage (%). n indicates the number of volunteers in each group. A p-value of ≤0.05 was considered statistically significant

When asked, "Are you worried about participating in the vaccine trial?", volunteers in the phase 1 group responded with an average score of 4.05±1.22, indicating "Strongly not worried." Similarly, the phase 2 group reported an average score of 4.3±0.99 (p=0.921), suggesting a similarly low level of concern. In both groups, the level of anxiety was found to be low. In response to the question "Has participating in this clinical trial changed your perspective on clinical research in a positive way?", the phase 1 group gave an average score of 4.42±0.60, while the phase 2 group responded with 4.26±0.81 (p=0.406), indicating a generally positive change in perception in both groups.

When the tendency to participate in another clinical trial in the future was evaluated, volunteers in the phase 1 group showed a higher inclination with an average score of 4.32±0.88, while those in the phase 2 group reported a lower tendency with a score of 3.88±0.88. This difference was statistically significant (p=0.041), indicating that the phase 1 group had a higher motivation to participate in future studies.

Responses to the question "Would you recommend participation in a clinical drug or vaccine trial to people around you?" were recorded as 4.11±0.80 in the phase 1 group and 4.27±0.66 in the phase 2 group (p=0.312). These results indicate that both groups were inclined to recommend participation in clinical research to others.

Responses to the question "Had you previously heard of terms such as phase 1 and phase 2 used in the development of drugs or vaccines?" were recorded as 4.05±1.17 in the phase 1 group and 3.79±1.09 in the phase 2 group. This difference was not statistically significant (p=0.323). These findings are presented in Table 2.

When volunteers were asked which vaccine they would prefer if there were equal access to all COVID-19 vaccines, 85.7% (n=197) stated that they would choose the national vaccine, Turkovac. This rate was 57.9% (n=11) in the phase 1 group and 88.2% (n=186) in the phase 2 group. The difference between the groups was statistically significant (p=0.002). Among other vaccine options, the mRNA-based Biontech/Pfizer

Table 2. Volunteers' responses regarding their clinical research experience

	Phase 1 (n=19)	Phase 2 (n=211)	p-value
Are you worried about participating in the vaccine trial?	4.05±1.22	4.03±0.99	0.921
Has your perspective on clinical research changed in a positive way?	4.42±0.60	4.26±0.81	0.406
Would you participate in another clinical trial?	4.32±0.88	3.88±0.88	0.041
Would you recommend participating in clinical trials?	4.11±0.80	4.27±0.66	0.312
Had you heard of terms like "Phase 1" and "Phase 2"?	4.05±1.17	3.79±1.09	0.323

All questions were evaluated using a 5-point Likert Scale (1: Strongly disagree, 5: Strongly agree). Data are presented as mean±standard deviation. n indicates the number of volunteers in each group. A p-value of ≤0.05 was considered statistically significant

vaccine was the second most preferred, chosen by 9.6% (n=22) of the volunteers. A detailed distribution of these findings is presented in Table 3.

When volunteers were asked, "What is your reason for participating in this clinical vaccine trial?", 42.1% (n=8) of the phase 1 group and 49.8% (n=105) of the phase 2 group stated that they participated because the vaccine was nationally produced. This difference between the groups was statistically significant (p=0.012). Another frequently reported reason for participation was the desire to contribute to science and to the development of a national vaccine. This reason was cited by 42.1% (n=8) of the phase 1 group and 45.0% (n=95) of the phase 2 group (Table 4).

When volunteers were asked whether they had previously participated in any clinical trial, 80.4% (n=179) stated that they had not taken part in any prior clinical research. In contrast, 19.6% (n=51) reported previous participation in a clinical trial.

Table 5 presents comparative data between volunteers with prior clinical trial experience and those participating for the first time.

DISCUSSION

In this study, the sociodemographic characteristics, perspectives on clinical research, motivations for participation, and vaccine preferences of volunteers who participated in the phase 1 and phase 2 clinical trials of Turkovac, Turkiye's first national COVID-19 vaccine, were examined. The findings indicate that the majority of volunteers had a strong sense of trust in the Turkovac vaccine, and their motivation to participate in clinical trials was largely based on this trust. General trends of trust toward COVID-19 vaccines play a critical role in increasing vaccine acceptance rates.[11]

When examining the sociodemographic characteristics of the volunteers, it was observed that the participation rate of female volunteers was low. This finding suggests that participation in clinical research may be influenced not only by individual factors but also by social gender norms and interpersonal dynamics.[6,11] The literature suggests that women's participation rates in clinical research are generally lower than those of men, which may be attributed to a higher perceived risk among women regarding clinical trials. This situation highlights the need for developing specific strategies to ensure gender equality in research participation.[12,13] Additionally, the majority of volunteers participating in our study were between the ages of 32 and 44 and university graduates, suggesting that

Table 3. Volunteers' vaccine preferences

	Total (n=230)	Phase 1 (n=19)	Phase 2 (n=211)	p-value
Inactivated virus vaccine; National vaccine – Turkovac	197 (85.7%)	11(57.9%)	186 (88.2%)	0.002
MessengerRNA (mRNA) vaccine; Biontech/Pfizer	22 (9.6%)	6 (31.6%)	16 (7.6%)	
Other options	11 (4.7%)	2 (10.5%)	9 (4.2%)	

Data are presented as number and percentage (%). n indicates the number of volunteers in each group. A p-value of ≤0.05 was considered statistically significant

Table 4. Responses to the Question "What is your reason for participating in this clinical vaccine trial?"

	Total (n=230)	Phase 1 (n=19)	Phase 2 (n=211)	p-value
Because the vaccine is nationally produced	113 (49.1%)	8 (42.1%)	105 (49.8%)	0.012
To contribute to science and the development of a national vaccine	103 (44.8%)	8 (42.1%)	95 (45.0%)	
Other (financial reasons)	4 (1.7%)	3 (15.8%)	1 (0.5%)	

Data are presented as number and percentage (%). n indicates the number of volunteers in each group. A p-value of ≤0.05 was considered statistically significant

Table 5. Responses based on previous participation in clinical trials

	Previously participated (n=45)	First-time participants (n=185)	p-value
Has your perspective on clinical research changed positively?	4.29±0.78	4.27±0.80	0.889
Would you participate in another clinical trial?	4.22±0.76	3.84±0.90	0.010
Would you recommend participation in clinical trials?	4.27±0.75	4.25±0.66	0.911
Have you heard of terms like phase 1 and phase 2?	3.93±1.07	3.78±1.10	0.414

The questions were evaluated using a 5-point Likert Scale; data are presented as mean±standard deviation. n indicates the number of volunteers within the group. A p-value of ≤0.05 was considered statistically significant

this group is more willing to participate in clinical research. While a previous study found no age-related difference, participation in clinical trials decreased with increasing education level.[14] This suggests that pandemic conditions may have influenced individuals' willingness to participate in clinical trials, considering their age and education level. The fact that the vaccine was well tolerated and had few side effects in phase 1 may have also influenced the profile of people who preferred the vaccine in phase 2. On the other hand, unless clinical trial protocols mandate equal gender distribution among volunteers, it should be considered that researchers may prefer male participants, especially when taking into account the potential risk of pregnancy in female volunteers.

The findings of this study revealed that financial motivation, which is frequently emphasized in the literature, was not the primary reason for participation among the volunteers.[15-18] Instead, the majority of volunteers participated in the study not for financial gain, but due to their trust in the national vaccine and their desire to contribute to science. Although previous studies have suggested that volunteers are often motivated by financial incentives, this study highlights that trust in the national vaccine was the predominant factor for participation.[19] However, in clinical trials involving patient volunteers rather than healthy individuals, motivations such as the expectation of treatment or potential benefit from the investigational drug may become more prominent factors influencing participation.

The majority of volunteers stated that they would prefer Turkovac if there were equal access to all COVID-19 vaccines. This finding indicates that vaccine preference is not based solely on medical factors but is also influenced by social, cultural, and psychological elements. The literature emphasizes that individuals' attitudes toward vaccines are closely related to trust, identity, a sense of collective belonging, and support for national production. Especially during the pandemic, the spread of misinformation and anti-vaccine content on social media has been shown to significantly impact individual decision-making processes. Nevertheless, in this study, the high level of trust in the national vaccine appears to have combined with a sense of national solidarity, positively influencing vaccine acceptance. However, one of the limitations of this study is that the survey was conducted only with volunteers who participated in the Turkovac trial; thus, individuals who preferred other vaccines or who hold anti-vaccine views were not included.

Vaccine hesitancy cannot be explained solely by a lack of information; this attitude is also shaped by trust, values, and sociopolitical factors, and therefore, proposed solutions must also be multidimensional.[20] During the pandemic, uncertainties occasionally arose in society regarding the safety and efficacy of rapidly developed vaccines, which contributed to increased public distrust toward vaccination. Misinformation and anti-vaccine content, particularly those spread through social media, may have negatively influenced individual attitudes. Vaccine opposition has evolved into a global movement that is not limited to individual choices but is rooted in historical, sociopolitical, cultural, and even religious dynamics.[21] In this context, the high level of trust that the volunteers expressed toward Turkovac and their active participation in the process can be interpreted as indicators of both individual awareness and trust in nationally conducted clinical research. Indeed, a previous study also emphasized that gaining public trust plays a critical role in increasing demand for vaccination.[22] It is known that anti-vaccine campaigns are not limited to scientific arguments but are also supported by emotional and ideological content. Previous studies have reported that opposing narratives are based on various psychological factors such as distrust, fear, non-scientific beliefs, emphasis on personal autonomy, and conspiracy thinking.[23,24] In this study, the fact that the vast majority of volunteers actively participated in the process with trust in Turkovac suggests that the psychological resistance factors underlying anti-vaccine attitudes were limited in this sample and that the national vaccine gained meaningful public acceptance. In particular, the sense of trust provided by national production, the transparency of the scientific process, and clear and effective communication with volunteers may have played a role. In this regard, the development of Turkovac as a national vaccine is not only part of efforts to address the pandemic but also significant in that it represents the first locally developed vaccine to complete all stages-from laboratory to licensure-after a long period in which Turkiye lacked vaccine production capabilities. Overcoming the challenges encountered in the development of national vaccines holds great importance for sustaining public health and advancing vaccine development capacity in Turkiye.[6] In future vaccine development efforts and clinical trials, gaining public trust and enhancing volunteer motivation in areas such as scientific integrity, contribution to science, and service to humanity may help increase participation rates.[25]

Limitations

One of the main strengths of this study is that it is the first to evaluate the motivations and experiences of volunteers who participated in the clinical trials of Turkovac, Turkiye's first national COVID-19 vaccine. However, the study has several limitations. The relatively small number of participants is primarily due to the study being conducted at a single center and under ongoing pandemic conditions, which prevented it from being planned as a large-scale, multicenter study. Therefore, the findings may not fully reflect the broader societal perspective. Additionally, the phase 1 and phase 2 groups were not equally represented, as the survey was initiated while phase 1 trials were still ongoing, resulting in a lower number of phase 1 participants. Most volunteers lived in urban areas, which limits the generalizability of the results to the general population. Furthermore, the low proportion of female participants led to gender imbalance. These limitations could be addressed in future research by including a larger and more demographically diverse sample. The phase 1 study, due to its small number of volunteers and its first application to humans, affected both the gender and education levels of participants. Moving to phase 2, it was observed that women and those with higher education levels were more willing to participate.

CONCLUSION

As a result, this study reveals that volunteers' trust in a nationally produced vaccine played an important role in motivating their participation in clinical research. Moreover, since it is known that women and men use medications similarly under real-world conditions, special strategies should be developed to increase the participation of female volunteers in drug clinical trials, taking gender distribution into account.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was carried out with the permission of the Erciyes University Non-interventional Clinical Researches Ethics Committee (Date: 03.03.2021, Decision No: 2021/149).

Informed Consent

Written consent was obtained from volunteers who participated in the study.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

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

Author Contributions

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

Acknowledgements

We would like to thank all our volunteers who contributed to the national vaccine study and our team for their efforts.

REFERENCES

1. Lin Y, Hu Z, Zhao Q, Alias H, Danaee M, Wong LP. Understanding COVID-19 vaccine demand and hesitancy: a nationwide online survey in China. *PLoS Negl Trop Dis*. 2020;14(12):e0008961. doi:10.1371/journal.pntd.0008961
2. Vaccines and immunization: what is vaccination? https://www.who.int/news-room/questions-and-answers/item/vaccines-and-immunization-what-is-vaccination (Updated 01.04.2025).
3. Schmid P, Rauber D, Betsch C, Lidolt G, Denker ML. Barriers of influenza vaccination intention and behavior-a systematic review of influenza vaccine hesitancy, 2005-2016. *PLoS One*. 2017;12(1):e0170550. doi:10.1371/journal.pone.0170550
4. Ozdarendeli A, Sezer Z, Pavel STI, et al. Safety and immunogenicity of an inactivated whole virion SARS-CoV-2 vaccine, TURKOVAC, in healthy adults: interim results from randomised, double-blind, placebo-controlled phase 1 and 2 trials. *Vaccine*. 2023;41(2):380-390. doi:10.1016/j.vaccine.2022.10.093.
5. Sezer Z. COVID-19 aşı klinik araştırmaları. Klinik araştırmalar. Editörler , Özet A, Tunçok Y, Özdemir N, Yazıcı O. 1. Baskı. Ankara: Turkiye Klinikleri; 2022. p.169-74.
6. Kochhar S, Salmon DA. Planning for COVID-19 vaccines safety surveillance. *Vaccine*. 2020;38(40):6194-6198. doi:10.1016/j.vaccine.2020.07.013
7. Mazıcıoğlu MM, Göğüsten B, Erenmemişoğlu A. İlaç biyoeşdeğerlik çalışmalarında katılımcı profili. *Turk Klin J Med Sci*. 2004;24(4):328-331.
8. Erdoğdu Y, Koçoğlu F, Sevim C. COVID-19 pandemisi sürecinde anksiyete ile umutsuzluk düzeylerinin psikososyal ve demografik değişkenlere göre incelenmesi. *Klin Psikiyatr Derg*. 2020:23. doi:10.5505/kpd.2020.35403
9. Karlsson LC, Soveri A, Lewandowsky S, et al. Fearing the disease or the vaccine: the case of COVID-19. *Personal Individ Differ*. 2021;172:110590. doi:10.1016/j.paid.2020.110590
10. Lin Y, Hu Z, Zhao Q, Alias H, Danaee M, Wong LP. Understanding COVID-19 vaccine demand and hesitancy: a nationwide online survey in China. *PLoS Neglect Tropic Dis*. 2020;14(12):e0008961. doi:10.1371/journal.pntd.0008961
11. Freeman D, Loe BS, Chadwick A, et al. COVID-19 vaccine hesitancy in the UK: the Oxford coronavirus explanations, attitudes, and narratives survey (Oceans) II. *Psychol Med*. 2022;52(14):3127-3141. doi: 10.1017/S0033291720005188
12. Ding EL, Powe NR, Manson JE, Sherber NS, Braunstein JB. Sex differences in perceived risks, distrust, and willingness to participate in clinical trials: a randomized study of cardiovascular prevention trials. *Arch Int Med*. 2007;167(9):905-912. doi:10.1001/archinte.167.9.905
13. van Diemen J, Verdonk P, Chieffo A, et al. The importance of achieving sex- and gender-based equity in clinical trials: a call to action. *Eur Heart J*. 2021;42(31):2990-2994. doi:10.1093/eurheartj/ehab457
14. Lobato L, Bethony JM, Pereira FB, Grahek SL, Diemert D, Gazzinelli MF. Impact of gender on the decision to participate in a clinical trial: a cross-sectional study. *BMC Public Health*. 2014;14:1156. doi:10.1186/1471-2458-14-1156
15. Almeida L, Azevedo B, Nunes T, Vaz-da-Silva M, Soares-da-Silva P. Why healthy subjects volunteer for phase I studies and how they perceive their participation? *Eur J Clin Pharmacol*. 2007;63(11):1085-1094. doi:10.1007/s00228-007-0368-3
16. Bigorra J, Baños JE. Weight of financial reward in the decision by medical students and experienced healthy volunteers to participate in clinical trials. *Eur J Clin Pharmacol*. 1990;38(5):443-446. doi:10.1007/bf02336681
17. Hassar M, Pocelinko R, Weintraub M, Nelson D, Thomas G, Lasagna L. Free-living volunteer's motivations and attitudes toward pharmacologic studies in man. *Clin Pharmacol Ther*. 1977;21(5):515-519. doi:10.1002/cpt1977215515
18. Hermann R, Heger-Mahn D, Mahler M, et al. Adverse events and discomfort in studies on healthy subjects: the volunteer's perspective. A survey conducted by the German Association for Applied Human Pharmacology. *Eur J Clin Pharmacol*. 1997;53(3-4):207-214. doi:10.1007/s002280050364
19. Halpern SD. Financial incentives for research participation: empirical questions, available answers and the burden of further proof. *Am J Med Sci*. 2011;342(4):290-293. doi:10.1097/MAJ.0b013e3182297925
20. Dubé È, Ward JK, Verger P, MacDonald NE. Vaccine hesitancy, acceptance, and anti-vaccination: trends and future prospects for public health. *Annu Rev Public Health*. 2021;42:175-191. doi:10.1146/annurev-publhealth-090419-102240
21. Oliveira IS, Cardoso LS, Ferreira IG, et al. Anti-vaccination movements in the world and in Brazil. *Rev Soc Bras Med Trop*. 2022;55:e05922021. doi:10.1590/0037-8682-0592-2021
22. Mills MC, Salisbury D. The challenges of distributing COVID-19 vaccinations. *E Clin Med*. 2020 (8);31:100674. doi: 10.1016/j.eclinm.2020.100674
23. Barrett JS, Yang SY, Muralidharan K, et al. Considerations for addressing anti-vaccination campaigns: how did we get here and what can we do about it? *Clin Transl Sci*. 2022;15(6):1380-1386. doi:10.1111/cts.13273
24. Fasce A, Schmid P, Holford DL, Bates L, Gurevych I, Lewandowsky S. A taxonomy of anti-vaccination arguments from a systematic literature review and text modelling. *Nat Hum Behav*. 2023;7(9):1462-1480. doi:10.1038/s41562-023-01644-3
25. Harapan H, Wagner AL, Yufika A, et al. Acceptance of a COVID-19 vaccine in Southeast Asia: a cross-sectional study in Indonesia. *Front Public Health*. 2020;8:381. doi:10.3389/fpubh.2020.00381

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1707830
J Health Sci Med. 2025;8(5):770-774

Dietary macronutrient composition and perinatal outcomes according to pre-pregnancy BMI

Ayşe Çiğdem Bayrak, Recep Taha Ağaoğlu
Department of Obstetrics and Gynecology, Ankara Etlik City Hospital, Ankara, Turkiye

Cite this article as: Bayrak AÇ, Ağaoğlu RT. Dietary macronutrient composition and perinatal outcomes according to pre-pregnancy BMI. *J Health Sci Med.* 2025;8(5):770-774.

Received: 28.05.2025 ◆ **Accepted**: 23.07.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: Dietary habits and nutritional balance during pregnancy have been linked to maternal and neonatal well-being. This study aimed to compare the quantity and distribution of dietary macronutrient intake among pregnant women with varying pre-pregnancy body-mass index (BMI) categories, to evaluate their nutritional habits, and to explore the potential impact of these factors on perinatal outcomes.

Methods: This study was designed as a prospective cohort with retrospective collection of baseline data included 120 pregnant women between 24-28 weeks of gestation who underwent an oral glucose tolerance test. Participants were divided into two groups based on pre-pregnancy BMI: those with a BMI between 18.5 and 24.9 kg/m^2, classified as having normal weight, and those with a BMI between 30 and 39.9 kg/m^2, classified as having obesity. Women diagnosed with gestational diabetes, impaired glucose tolerance, or pre-existing metabolic disorders were excluded. Daily dietary intake was assessed using 72-hour food diaries, which included detailed information on the timing and content of six daily meals (three main and three snacks), the types and amounts of foods and beverages consumed, preparation methods, and the location of each meal. Perinatal outcomes, including gestational age at delivery, birth weight, Apgar scores, and neonatal intensive care unit (NICU) admission, were recorded and compared between groups.

Results: Sixty participants were included in each group. The mean daily energy intake was higher among women with obesity (2117.1 kcal vs. 2004.6 kcal), with a significantly greater proportion of energy derived from carbohydrates (51.7% vs. 44.6%; p=0.026). Weight gain during pregnancy was significantly higher in women with obesity across all trimesters. Based on participants' self-reports, physical activity levels decreased with advancing gestation in both groups, though the difference between groups was not statistically significant. No significant differences were observed in gestational age at delivery, birth weight, or Apgar scores. However, NICU admission was more frequent among women with obesity (9 cases vs. 2 cases; p=0.046).

Conclusion: Pregnant women with higher pre-pregnancy BMI demonstrated greater carbohydrate intake and increased weight gain throughout pregnancy, potentially indicating elevated metabolic risk. Balanced macronutrient intake and early monitoring of gestational weight gain may be essential components of antenatal care. Larger prospective studies are needed to validate these findings and support tailored nutritional interventions in this population.

Keywords: Dietary proteins, maternal obesity, weight gain

INTRODUCTION

Obesity is recognized as one of the most significant public health challenges of the 21st century.[1] According to the World Health Organization (WHO), by 2024, approximately 43% of the global adult population was classified as overweight [body mass index (BMI): 25-29.9 kg/m^2], and 16% as obese (BMI ≥30 kg/m^2).[2] The global prevalence of obesity, which was reported as 6.4% in 1975, has more than doubled over the past few decades, demonstrating an alarming trend.[3]

The association between obesity and increased morbidity and mortality has long been established. During pregnancy, obesity poses various health risks not only for the mother but also for the fetus. It has been associated with increased maternal morbidity as well as obstetric complications such as preterm birth, macrosomia, and a higher rate of cesarean delivery.[4,5] Obesity-related impairment of placental function may alter fetal nutrient transport, while elevated levels of insulin resistance and inflammation in pregnant women with obesity may adversely affect fetal metabolism.[6,7]

Fetal growth, oxidative metabolism, and fundamental biological processes are largely dependent on protein synthesis sustained by placental amino acid transport.[8] In pregnancies affected by maternal obesity, elevated levels of insulin, leptin, and insulin-like growth factor 1 (IGF-1) may enhance fetal amino acid transfer by activating various placental

Corresponding Author: Ayşe Çiğdem Bayrak, drcigdembayrak@gmail.com

signaling pathways.[9] Disruptions in maternal-fetal transport mechanisms may influence fetal metabolic status from the time of implantation. Moreover, insufficient maternal protein intake has been associated with fetal growth restriction (FGR) and increased rates of perinatal morbidities. Conversely, excessive protein intake may negatively affect embryonic and fetal development due to increased amino acid catabolism and competitive inhibition among transporters.[10,11]

In this context, assessing both the quantity and quality of the maternal diet has important implications for pregnancy outcomes, as the amount and source of protein intake may influence fetal growth and contribute to long-term metabolic development in childhood.[12] However, there is limited evidence in the literature directly comparing macronutrient intake and distribution among pregnant women with different pre-pregnancy BMI categories.

Therefore, this study aimed to evaluate the relationship between daily dietary intake patterns-including macronutrient distribution-maternal BMI, and overall lifestyle characteristics during pregnancy across BMI-based groups. Furthermore, we sought to explore the potential impact of these differences on key perinatal outcomes.

METHODS

Participant Selection

This was prospective cohort study with retrospective collection of baseline data, including pregnant women between 24 and 28 weeks of gestation who underwent an oral glucose tolerance test (OGTT) at a tertiary maternity hospital between January 2 and May 8, 2025. The study was approved by the Ankara Etlik City Hospital Ethics Committee (Date 14.05.2025, Decision No: AESH-BADEK1-2025-206) and conducted in accordance with the Declaration of Helsinki. All participants received verbal and written information about the study, and written informed consent was obtained prior to enrollment.

Participants were divided into two BMI-based groups to enable a clear comparison between metabolically distinct populations. Overweight women (BMI 25.0-29.9 kg/m^2) were excluded to reduce potential heterogeneity and avoid dilution of group differences in dietary and metabolic parameters. Group 1 consisted of participants with a pre-pregnancy BMI between 18.5 and 24.9 kg/m^2 (classified as having normal weight), and group 2 included those with a BMI between 30 and 39.9 kg/m^2 (classified as having obesity).

During the study visit, we measured weight using a Calibrated Digital Scale, with participants wearing light clothing and no shoes. Height was measured using a stadiometer, and BMI was calculated as weight in kilograms divided by height in meters squared (kg/m^2). First and second trimester weight gain was documented based on pregnancy records, while third-trimester weight gain was recorded at the time of hospital admission for delivery. Additionally, during the same visit, participants' physical activity levels were recorded based on self-reported changes and categorized as "increased," "decreased," or "unchanged."

We excluded women who underwent OGTT at their initial antenatal visit, those who received a 100 g OGTT following an abnormal 50 g glucose challenge test, and those diagnosed with impaired glucose tolerance or gestational diabetes based on OGTT results. Additional exclusion criteria included maternal age over 40 years, known metabolic disorders, smoking, alcohol or substance use, history of bariatric surgery, prior dietary counseling, or working night shifts. To minimize potential confounding factors affecting neonatal outcomes, we also excluded cases with multiple gestation, cervical insufficiency, structural and/or chromosomal fetal anomalies, and preterm birth before 37 weeks of gestation. Following participant selection, we recorded perinatal outcomes including gestational age at delivery, birth weight, 1- and 5-minute Apgar scores, and the need for neonatal intensive care unit (NICU) admission.

A priori power analysis using G*power 3.1.9.7 indicated that 60 participants per group were required to detect a moderate effect size (Cohen's d=0.52) with 80% power and a two-tailed alpha of 0.05. Accordingly, 120 participants were included in the study.

Assessment of Nutrient Intake

We assessed participants' daily dietary intake using a retrospective 24-hour dietary recall questionnaire administered during their morning visit to the outpatient clinic. For each of the three main meals and three snacks consumed the previous day, the time of consumption, types and amounts of food and beverages, preparation methods, place of consumption, and the participant's perceived level of hunger before eating were recorded.

Participants were also asked to complete two additional 24-hour dietary recall forms-one on a weekday and one on a weekend day-resulting in a total of 72 hours of dietary intake data. Although this 3-day recall method is widely used in nutritional studies, we acknowledge that it may not fully reflect long-term dietary habits during pregnancy, when eating patterns may vary. To minimize recall bias and potential under- or over-reporting, participants were provided with visual guidance on portion sizes using a photographic food atlas and were instructed to complete the forms on the same day as food consumption.

We analyzed the dietary data using the Nutrition Information System (BeBIS), a computer-assisted dietary assessment software developed within the scope of the national nutrition monitoring program. BeBIS enables standardized analysis of food records by referencing an extensive Turkish food composition database and calculating macro- and micronutrient content based on portion size, preparation methods, and frequency of consumption. Through this system, we obtained detailed estimates of total energy intake and the distribution of protein, fat, and carbohydrate intake for each participant.

Statistical Analysis

We performed statistical analyses using IBM SPSS Statistics for Windows, Version 23.0 (IBM Corp., Armonk, NY, USA).

We assessed the distribution of continuous variables using the Kolmogorov-Smirnov and Shapiro-Wilk tests. For normally distributed data, we compared groups using the independent samples T test and reported the results as mean±standard deviation (SD). For non-normally distributed data, we used the Mann-Whitney U test and presented the results as median (minimum-maximum).

Categorical variables were expressed as numbers (n) and percentages (%), and group comparisons were performed using the Pearson chi-square test or Fisher's exact test, as appropriate. A p-value <0.05 was considered statistically significant for all analyses.

RESULTS

We enrolled a total of 120 pregnant women in this study, with 60 participants in each group based on BMI classification. There were no statistically significant differences between the groups in terms of maternal age, gravida, parity, or number of previous abortions (p>0.05). Similarly, the distribution of educational levels was comparable between the groups (p=0.66). The mean pre-pregnancy BMI was 22.7 (range: 19.2-24.8) in the normal-weight group and 31.6 (range: 30.0-38.6) in the group of women with obesity.

Total gestational weight gain was higher among women with obesity, with an average of 17.6 kg compared to 13.6 kg in the normal-weight group. This difference was consistent across all three trimesters, with statistically significant increases particularly during the first two trimesters (p=0.009).

We found no significant differences between the groups regarding changes in physical activity throughout pregnancy (p=0.95). However, most participants in both groups reported a decline in physical activity frequency after the onset of pregnancy. Other clinical and demographic characteristics of the participants are summarized in Table 1.

Daily energy and nutrient intake data are presented in Table 2. There were no substantial differences in the number of main meals or snacks between the groups. The mean daily energy intake was higher among women with obesity (2117.1 kcal vs. 2004.6 kcal; p=0.045), and a significantly greater proportion of this energy was derived from carbohydrates (51.7% vs. 44.6%; p=0.026). In contrast, the normal-weight group obtained a larger proportion of their energy from protein (p=0.025), whereas fat-derived energy did not differ significantly between the groups (p>0.05).

There were no significant differences between the groups in gestational age at delivery, birth weight, or Apgar scores. The mean birth weight was 3170 (±410) g in the normal-weight group and 3210 (±395) g in the group of women with obesity (p>0.05). However, the rate of NICU admission was significantly higher in the group of women with obesity (9 cases vs. 2 cases; p=0.046). Among these cases, indications for NICU admission included respiratory problems (n=4), Apgar score <7 at 5 minutes (n=2), hypoglycemia (n=2), and hyperbilirubinemia (n=1). In the normal-weight group, both NICU admissions were due to respiratory problems. Other perinatal outcomes are summarized in Table 3.

Table 1. Comparison of demographic and clinical characteristics between study groups

	Normal weight women (n=60)	Women with obesity (n=60)	p-value
Maternal age*	29.8 (20-37)	30.6 (19-38)	0.422
Gravidity*	1.47 (1-3)	1.65 (1-4)	0.256
Parity*	1.32 (0-2)	1.45 (0-3)	0.401
Number of abortions*	0.15 (0-2)	0.20 (0-3)	0.573
Education level†			
Primary school	12 (20%)	9 (15%)	0.631
High school	13 (21.6%)	16 (26.6%)	0.670
University	24 (40%)	25 (41.6%)	0.462
Postgraduate	11 (18.3%)	10 (16.6%)	0.295
1st trimester hyperemesis†			
Yes	24 (40%)	11 (18.3%)	0.020
No	36 (60%)	49 (81.6%)	
Gestational weight gain (kg)*			
1st trimester	2.7 (0-4.6)	4.3 (0-6.2)	0.009
2nd trimester	4.4 (2.9-7.1)	5.9 (1.7-8.3)	0.009
3rd trimester	6.5 (3.1-8.2)	7.4 (2.6-9.9)	0.012
Change in physical activity†			
Increased	12 (20%)	14 (%23.3)	0.822
Decreased	27 (45%)	25 (%41.6)	0.857
No change	21 (35%)	21 (%35)	1.000

*Presented as mean (min–max); †Presented as n (%); Statistical significance was assessed using Mann-Whitney U test or chi-square test as appropriate

Table 2. Daily energy and nutrient intake in normal-weight women and women with obesity

	Normal weight women (n=60)	Women with obesity (n=60)	p-value
Number of main meals	3.4 (3-4)	3.5 (3-4)	0.350
Number of snacks	2.3 (1-3)	2.6 (1-3)	0.458
Total energy (kcal/day)	2004.6 (1704-2886)	2117.1 (1071-2969)	0.045
Carbohydrates (g/day)	221.4 (87-380)	253.5 (130-370)	0.030
%energy from carbohydrates	44.6 (34-59)	51.7 (32-70)	0.026
Protein (g/day)	77.1 (49-118)	70.2 (45-121)	0.049
%energy from protein	17.3 (9-32)	13.3 (10-28)	0.025
Fat (g/day)	84.3 (47-156)	77.6 (41-165)	0.120
%energy from fat	37.2 (16-51)	34.2 (17-48)	0.115
Saturated fatty acids (g/day)	26.6 (12-46)	22.6 (9-47)	0.210
Unsaturated fatty acids (g/day)	49.5 (7-51)	48.9 (6-49)	0.170

All variables are continuous and presented as median (minimum-maximum). Comparisons were performed using the Mann-Whitney U test

Table 3. Perinatal outcomes

	Normal weight women (n=60)	Women with obesity (n=60)	p-value
Gestational age at delivery*	38.5 (37-40.2)	38.4 (37-40.1)	0.681
Birthweight (kg)*	3170 (2470-4130)	3210 (2630-4310)	0.543
Apgar score at 1 minute*	8 (5-9)	8 (6-9)	0.217
Apgar score at 5 minutes*	9 (7-10)	9 (7-10)	0.434
NICU admission†	2 (3.3%)	9 (15.0%)	0.046

*Presented as mean (min-max); †Presented as n (%); Statistical significance was assessed using Mann-Whitney U test or chi-square test as appropriate

DISCUSSION

Pregnancy is a unique physiological period characterized by significant changes in maternal metabolism to support normal embryonic and fetal development. Throughout this period, maternal nutritional status plays a key role in shaping both maternal and fetal health outcomes.[13] It is theorized that intrauterine epigenetic changes may be influenced by maternal nutrition and have been associated with an increased risk of metabolic disorders in childhood-such as obesity and insulin resistance-with consequences extending into adulthood.[14,15] In this context, maternal obesity-which alters metabolic and hormonal pathways-has been associated with an increased risk of various adverse outcomes affecting both the mother and the fetus.[16]

Evaluating nutritional habits before and during pregnancy can provide clinically relevant insights for both the management of obesity and the improvement of perinatal outcomes. In this context, we compared dietary energy and macronutrient distribution between different maternal BMI groups. Our findings demonstrated that normal-weight women had higher daily protein intake and derived a greater proportion of their energy from protein and fat, suggesting a more balanced dietary pattern. In contrast, the higher-BMI group consumed a greater proportion of energy from carbohydrates. This carbohydrate-dominant intake could contribute to metabolic dysregulation observed in individuals with elevated pre-pregnancy BMI.

Contrary to popular belief, limited weight gain during pregnancy has not been definitively shown to adversely affect fetal development. While extreme restrictions-such as those observed during famine or war-are associated with low birth weight and increased perinatal mortality, studies have shown that women who gain minimal weight within physiological limits can still support adequate fetal growth without significantly increasing adverse outcomes.[17,18] In contrast, excessive gestational weight gain is associated with a higher risk of gestational diabetes, preeclampsia, macrosomia, cesarean delivery, and NICU admission.[19,20] Furthermore, excessive weight gain during pregnancy may increase the risk of metabolic disorders in offspring during childhood.[21]

Given these considerations, the institute of medicine (IOM) recommends limiting gestational weight gain in women with elevated pre-pregnancy BMI, particularly restricting weight gain to a maximum of 2 kg during the first trimester.[22] Despite these recommendations, participants with higher pre-pregnancy BMI in our study exhibited significantly greater weight gain during the first trimester compared to those with normal BMI. This difference may be partly attributable to the lower incidence of hyperemesis gravidarum among women with obesity, although this relationship warrants further investigation.

Nevertheless, we observed no significant differences between the groups in terms of gestational age at delivery, birth weight, or Apgar scores. However, NICU admission rates were significantly higher in the group with obesity, primarily due to respiratory problems, low Apgar scores, hypoglycemia, and hyperbilirubinemia. These findings are consistent with previous large-scale studies demonstrating that excessive gestational weight gain is associated with an increased risk of neonatal complications such as macrosomia, shoulder dystocia, and neonatal hypoglycemia.[23]

Our study also revealed a marked reduction in physical activity levels during pregnancy compared to pre-pregnancy in both groups. This finding is consistent with previous studies[24] and suggests that awareness and counseling regarding physical activity during pregnancy may be insufficient.

A combined evaluation of dietary composition and physical activity plays a critical role in achieving healthy gestational weight gain. Among women with obesity, balancing macronutrient intake and maintaining physical activity levels may support improved maternal and perinatal outcomes. Therefore, enhancing preconception counseling services and developing individualized follow-up strategies-particularly for controlling weight gain during the early stages of pregnancy, when energy requirements have not yet significantly increased, and for establishing appropriate nutritional targets in high-risk groups-should be prioritized. Similar education levels in both groups may have helped reduce the impact of socioeconomic disparities on dietary intake. However, as other dimensions of socioeconomic status-such as income or access to healthy food-were not assessed, this remains a potential source of residual confounding.

Limitations

Our study has several limitations that should be considered when interpreting the findings. First, the relatively small sample size may limit the generalizability of the results. Second, dietary intake was assessed using self-reported 24-hour recall forms, which are inherently subject to recall bias-particularly during pregnancy, when daily routines and eating behaviors may fluctuate. This may have led to under- or overestimation of certain food groups or portion sizes. Although we attempted to mitigate this risk by providing standardized visual aids and instructing participants to complete the forms on the same day as food consumption, some degree of recall bias remains inevitable. Finally, although the observational design makes it difficult to establish cause-effect relationships, the methodological rigor of our study enhance the clinical relevance and reliability of the findings.

CONCLUSION

Differences in dietary composition and macronutrient distribution across maternal BMI categories may provide valuable insights for individualized nutritional counseling during pregnancy. Encouraging a protein-rich and balanced dietary pattern, rather than a carbohydrate-dominant one, among women with higher pre-pregnancy BMI may be associated with improved maternal and fetal outcomes. Further prospective studies are essential to inform evidence-based nutritional counseling strategies tailored to maternal BMI.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was carried out with the permission of the Ankara Etlik City Hospital Ethics Committee (Date 14.05.2025, Decision No: AESH-BADEK1-2025-206).

Informed Consent

All patients signed and free and informed consent form.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

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

Author Contributions

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

REFERENCES

1. Villalobos JA. Obesity: the real pandemic of the 21st century. *Cir Cir (Engl Ed).* 2016;84(5):351-355. doi:10.1016/j.circen.2016.08.013
2. World Health Organization. Obesity and overweight. 2024. Available from: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight. Accessed May 25, 2025.
3. An R, Guan C, Liu J, Chen N, Clarke C. Trade openness and the obesity epidemic: a cross-national study of 175 countries during 1975-2016. *Ann Epidemiol.* 2019;37:31-36. doi:10.1016/j.annepidem.2019.07.002
4. Reed J, Case S, Rijhsinghani A. Maternal obesity: perinatal implications. *SAGE Open Med.* 2023;11:20503121231176128. doi:10.1177/20503121231176128
5. Mohammadi M, Maroufizadeh S, Omani-Samani R, Almasi-Hashiani A, Amini P. The effect of prepregnancy body-mass index on birth weight, preterm birth, cesarean section, and preeclampsia. *J Matern Fetal Neonatal Med.* 2019;32(22):3818-3823. doi:10.1080/14767058.2018.1473366
6. Howell KR, Powell TL. Effects of maternal obesity on placental function and fetal development. *Reproduction.* 2016;153(3):R97-R106. doi:10.1530/REP-16-0495
7. Schmatz M, Madan J, Marino T, Davis J. Maternal obesity: the interplay between inflammation, mother and fetus. *J Perinatol.* 2010;30(7):441-446. doi:10.1038/jp.2009.182
8. Vaughan OR, Rosario FJ, Powell TL, Jansson T. Regulation of placental amino acid transport and fetal growth. *Prog Mol Biol Transl Sci.* 2017;145:217-251. doi:10.1016/bs.pmbts.2016.12.008
9. Kelly AC, Powell TL, Jansson T. Placental function in maternal obesity. *Clin Sci (Lond).* 2020;134(8):961-984. doi:10.1042/CS20190266
10. Herring CM, Bazer FW, Johnson GA, Wu G. Impacts of maternal dietary protein intake on fetal survival, growth, and development. *Exp Biol Med (Maywood).* 2018;243(6):525-533. doi:10.1177/1535370218758275
11. Godfrey KM, Barker DJ. Fetal programming and adult health. *Public Health Nutr.* 2001;4(2B):611-624. doi:10.1079/phn2001145
12. Jahan-Mihan A, Rodriguez J, Christie C, Sadeghi M, Zerbe T. The role of maternal dietary proteins in development of metabolic syndrome in offspring. *Nutrients.* 2015;7(11):9185-9217. doi:10.3390/nu7115460
13. Most J, Dervis S, Haman F, Adamo KB, Redman LM. Energy intake requirements in pregnancy. *Nutrients.* 2019;11(8):1812. doi:10.3390/nu11081812
14. Barker DJ, Thornburg KL. The obstetric origins of health for a lifetime. *Clin Obstet Gynecol.* 2013;56(3):511-519. doi:10.1097/GRF.0b013e31829cb9ca
15. Lapehn S, Paquette AG. The placental epigenome as a molecular link between prenatal exposures and fetal health outcomes. *Curr Environ Health Rep.* 2022;9(3):490-501. doi:10.1007/s40572-022-00354-8
16. Leddy MA, Power ML, Schulkin J. The impact of maternal obesity on maternal and fetal health. *Rev Obstet Gynecol.* 2008;1(4):170-175.
17. Gilmore LA, Klempel-Donchenko M, Redman LM. Pregnancy as a window to future health: excessive gestational weight gain and obesity. *Semin Perinatol.* 2015;39(4):296-303. doi:10.1053/j.semperi.2015.05.009
18. Chen TL, Wu CH, Gau ML, Cheng SF. Prediction of the ideal gestational weight gain for reducing the risk of macrosomia/large for gestational age in women with gestational diabetes mellitus. *Midwifery.* 2025;140:104211. doi:10.1016/j.midw.2024.104211
19. Zhu Y, Zheng Q, Pan Y, et al. Association between prepregnancy BMI or gestational weight gain and pregnancy outcomes in Chinese women with gestational diabetes: a systematic review. *BMJ Open.* 2024;14(2):e075226. doi:10.1136/bmjopen-2023-075226
20. Minschart C, Lammertyn A, Van Crombrugge P, et al. Low gestational weight gain in women with gestational diabetes is safe with better metabolic profile postpartum. *J Clin Endocrinol Metab.* 2023;108(3):665-679. doi:10.1210/clinem/dgac599
21. Boney CM, Verma A, Tucker R, Vohr BR. Metabolic syndrome in childhood: association with birth weight, maternal obesity, and gestational diabetes. *Pediatrics.* 2005;115(3):e290–e296. doi:10.1542/peds.2004-1808
22. Yaktine AL, Rasmussen KM, eds. Weight Gain During Pregnancy: Reexamining the Guidelines. Washington, DC: National Academies Press; 2010.
23. Kominiarek MA, Saade G, Mele L, et al. Association between gestational weight gain and perinatal outcomes. *Obstet Gynecol.* 2018;132(4):875-881. doi:10.1097/AOG.0000000000002854
24. Borodulin KM, Evenson KR, Wen F, Herring AH, Benson AM. Physical activity patterns during pregnancy. *Med Sci Sports Exerc.* 2008;40(11):1901-1908. doi:10.1249/MSS.0b013e31817f1957

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1727206
J Health Sci Med. 2025;8(5):775-786

Clinical preferences and practice patterns of orthodontists in Turkiye regarding maxillary molar distalization

Samet Özden, Fırat Oğuz
Department of Orthodontics, Faculty of Dentistry, İnönü University, Malatya, Turkiye

Cite this article as: Özden S, Oğuz F. Clinical preferences and practice patterns of orthodontists in Turkiye regarding maxillary molar distalization. *J Health Sci Med.* 2025;8(5):775-786.

Received: 03.07.2025 ◆ **Accepted**: 24.07.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: The aim of this study was to evaluate the clinical approaches of orthodontists practicing in Turkiye regarding the distalization of maxillary first molars, as well as the factors influencing their treatment preferences.

Methods: This descriptive and cross-sectional survey study included responses from 132 orthodontists who are members of the Turkish Orthodontic Society. The questionnaire, delivered via Google Forms, comprised 37 items covering demographic data, distalization techniques, the use of clear aligners, and retention protocols. Data were analyzed using SPSS version 22.0. Relationships between variables were assessed using the Chi-square test, supported by Monte Carlo simulation where appropriate. A significance level of $p<0.05$ was considered. Ethical approval for this survey-based study was obtained from the İnönü University Scientific Researches and Publication Ethics Committee (Date: 24.09.2024, Decision No: 2024/6452).

Results: All participants reported utilizing maxillary molar distalization in their clinical practice. A total of 84.8% preferred skeletal anchorage-supported intraoral systems, and 62.9% reported placing miniscrews in the infrazygomatic crest region. According to the respondents, distalization was most frequently performed in adolescent patients (78.0%). In class II camouflage cases, 63.6% of orthodontists indicated a preference for premolar extraction. Clear aligner-based distalization was reported by 67.4% of participants, with Invisalign (Align Technology Inc., San Jose, CA, USA) being the most commonly used brand (56.8%) in this subgroup. The average duration of distalization was reported to be between 6 and 9 months by 43.9% of respondents. The most frequently encountered complication was miniscrew loosening (71.2%). In the post-treatment retention phase, 72.0% of clinicians reported using a combination of Essix and lingual retainers.

Conclusion: Maxillary molar distalization is widely employed among orthodontists in Turkiye, with notable variability in techniques and materials used based on clinician preference. These findings highlight the importance of individualized treatment planning and reflect the diversity in contemporary clinical practice.

Keywords: Distalization, survey, skeletal anchorage, orthodontic treatment, clear aligner

INTRODUCTION

According to Angle's classification, class II malocclusion is defined as a dental discrepancy in which the entire mandibular dentition is positioned more distally than normal, resulting in a misalignment between the dental arches.[1] The prevalence of class II malocclusion varies depending on ethnicity, environmental influences, and diagnostic criteria, with a global prevalence estimated at approximately 20%, and reports indicating rates as high as 40% in the 12-17 age group in Turkiye.[2,3] Due to this high prevalence, the management of class II malocclusion holds significant clinical importance, and various treatment modalities-including functional appliances, orthodontic camouflage, and orthognathic surgery-have been developed.[4]

Orthodontic camouflage aims to achieve acceptable occlusion and facial aesthetics without skeletal modification.[5] This treatment can be executed with or without tooth extraction. In non-extraction camouflage approaches, maxillary molar distalization is often employed to create the necessary space.[6]

Maxillary molar distalization is considered a viable non-extraction treatment option in mild to moderate sagittal discrepancies.[7] While traditional intraoral appliances eliminate the need for patient compliance associated with extraoral devices, they are often accompanied by undesirable anchorage loss.[8] To overcome this, temporary anchorage devices (TADs) have been introduced to provide resistance against reactive forces, minimizing anchorage loss during distalization.[5,9] These TADs can be placed in interradicular alveolar areas of the buccal or palatal regions or in extraradicular sites such as the infrazygomatic crest, enabling three-dimensional tooth movement with minimal anchorage loss.[5]

In addition to conventional methods, increasing aesthetic expectations and demand for patient comfort in recent

Corresponding Author: Samet Özden, samet.ozden@inonu.edu.tr

years have led to the widespread adoption of clear aligners as an alternative to labial brackets.[10] Clear aligners offer an effective solution, particularly in cases with mild to moderate crowding.[11,12] Beyond anterior alignment, aligners have also demonstrated the ability to achieve distalization of posterior teeth. Notably, Simon et al.[13] reported high success rates in upper molar distalization using clear aligner therapy (CAT).

Maintaining treatment outcomes over the long term is critical in both fixed appliance and aligner-based orthodontic treatments. As it is well-known that teeth tend to relapse following active treatment, retention protocols are necessary to preserve the achieved aesthetic and functional tooth positions.[14] However, there is no consensus among orthodontists regarding the optimal retention duration or method, resulting in varied clinical practices. Commonly used retention strategies include removable retainers and fixed lingual retainers.[15]

Given the wide variation in distalization methods, aligner protocols, and retention strategies observed in clinical orthodontic practice, there is a clear need for comprehensive, data-driven assessments of current treatment trends. To date, no previous study has systematically examined the clinical preferences of orthodontists regarding upper molar distalization within a national sample. This study fills a critical gap in the literature by providing the first large-scale, survey-based evaluation of real-world distalization practices among orthodontists in Turkiye. Its strength lies in its broad participant base, detailed assessment of both conventional and contemporary treatment approaches-including skeletal anchorage systems and CAT-and its analysis of how clinical decisions are shaped by demographic and institutional factors.

In this context, the aim of our study was to evaluate the clinical approaches of orthodontists in Turkiye regarding maxillary molar distalization, with a particular focus on appliance selection, skeletal anchorage preferences, clear aligner usage, and retention protocols.

METHODS

Ethics Committee Approval

This study was approved by the İnönü University Scientific Researches and Publication Ethics Committee (Date: 24.09.2024, Decision No: 2024/6452). All procedures were carried out in accordance with the ethical rules and the principles of the Declaration of Helsinki.

Design of the Study

This study is a descriptive, cross-sectional survey research based on responses collected from licensed orthodontists in Turkiye.

Participants and Sample Size

To ensure the collection of diverse and representative data, the study population was selected from among actively practicing orthodontists in Turkiye. A total of 132 orthodontists, all registered members of the Turkish Orthodontic Society, were contacted via email. The email correspondence included a detailed cover letter explaining the objectives of the study, accompanied by a link to the online survey form. The targeted sample encompassed orthodontists working across a variety of institutional settings, including Public Oral and Dental Health Centers, University Hospitals, and Private Clinics or Polyclinics.

Data Collection

Before distribution, the questionnaire was reviewed and approved by experts from the Turkish Orthodontic Society to ensure its clarity, relevance, and content validity. The questionnaire was created and distributed via Google Forms and consisted of 37 questions designed to assess participants' demographic characteristics and clinical preferences. It was structured under three main sections: demographic information (Table 1), distalization approaches (Table 2), and clear aligner-based distalization and retention protocols (Table 3).

Table 2 presents the questions aimed at evaluating the participants' clinical preferences regarding upper molar distalization. This section investigates factors such as the most commonly used distalization appliances, reasons for appliance selection, target patient groups, and the average duration of treatment. In addition, the frequency of miniscrew use, preferred insertion sites, and experiences with miniscrew failures are also addressed under this category.

Table 3 includes questions evaluating orthodontists' approaches to distalization treatment using clear aligners and the retention protocols implemented after treatment. This section explores the prevalence of clear aligner usage, preferred brands, clinical application strategies, as well as the types and duration of retention appliances.

Statistical Analysis

The collected data were analyzed using IBM SPSS Statistics for Windows, Version 22.0 (IBM Corp., Armonk, NY, USA). Descriptive statistics, including frequency, percentage, mean, median, and standard deviation, were calculated. Associations between categorical variables with two or more groups were evaluated using the Chi-square test. Although subgroup analyses were conducted, multivariate methods such as logistic regression were not employed, as the primary

Table 1. Questions regarding the demographic characteristics of the participants

Q1-What is your gender?	Q2-What is your age?	Q3-What type of institution do you work at?	Q4-What is your title?	Q5-How many years have you been practicing in the profession?
Male	20-25	Public oral and dental health center	Research assistant	0-3 years
	26-30	University hospital	Specialist dentist	3-6 years
	31-35		Lecturer	6-10 years
Female	36-40	Private clinic/polyclinic	Assistant professor	
			Associate professor	Over 10 years
	Over 40	Other	Professor	

Table 2. Treatment approaches of maxillary molar distalization	
Q6-Do you perform maxillary molar distalization in your clinical practice?	Yes
	No
Q7-Which radiographic methods do you commonly use when planning distalization?	Panoramic radiograph
	Cephalometric radiograph
	CBCT
	Digital setup or 3D modeling
Q8-In which age group do you more frequently perform maxillary molar distalization?	Children (8-12 years)
	Adolescents (13-18 years)
	Adults (18+ years)
Q9-Do you generally extract third molars in patients for whom you plan distalization?	Yes
	No
Q10-In class II camouflage cases, do you prefer premolar extraction or molar distalization?	Distalization
	Tooth extraction
Q11-What factors influence your decision between premolar extraction and molar distalization in class II camouflage cases?	Severity of molar relationship
	Amount of crowding
	Vertical growth pattern and anterior overbite
	Soft tissue profile
	Treatment duration
	Appliance cost
	All of the above
Q12-What is the primary criterion you consider first in making the decision between premolar extraction and molar distalization in class II camouflage cases?	Severity of molar relationship
	Amount of crowding
	Vertical growth pattern and anterior overbite
	Soft tissue profile
	Treatment duration
	Appliance cost
	Others
Q13-For patients where you indicate maxillary molar distalization, do you use extraoral distalization methods?	Occasionally
	Yes
	No
Q14-Which method do you more frequently use for intraoral molar distalization?	I do not use any
	Tooth- and tissue-supported distalization appliances
	Skeletal anchorage-supported distalization appliances
Q15-If you use tooth- and tissue-supported intraoral distalization appliances, which ones do you prefer?	I do not use any
	Pendulum
	Keleş Slider
	Carriere distalizer
	Frog
	Veltri
	First class
	Distal jet
	Jones jig
	ACCO
	Other
Q16-If you use skeletal anchorage-supported intraoral distalization appliances, which anatomical region do you most commonly used for anchorage?	I do not use any
	Infrazygomatic crest
	Buccal
	Maxillary tuber
	Palatinal
	Other
Q17-If you use palatal miniscrew-supported distalization appliances, which type do you prefer?	I do not use any
	Beneslider
	Modified Pendulum
	Modified Keleş Slider
	Mini screw supported frog
	Modified distal Jet
	Others
Q18-Do you routinely perform CBCT imaging before placing miniscrews in the palatal region to evaluate root positions and bone availability?	Yes
	No
	Sometimes

Table 2. Treatment approaches of maxillary molar distalization (table continues)	
Q19-In the buccal region, which type of miniscrew placement do you prefer most?	Extraradicular placement
	Interradicular placement
	I use both equally
	I do not use buccal miniscrew systems
Q20-What is the most commonly preferred screw size in extraradicular placement (infrazygomatic crest)?	2x10 mm
	2x11 mm
	2x12 mm
	2x13 mm
	2x14 mm
	I do not use infrazygomatic crest screws
	Other
Q21-Which material do you prefer for the infrazygomatic crest miniscrew?	Stainless steel miniscrew
	Titanium alloy miniscrew
	I do not use infrazygomatic crest screws
Q22-What is the biggest difficulty or complication you encounter during TAD insertion for maxillary molar distalization?	Finding sufficient bone density
	Soft tissue interference
	Device instability
	Patient discomfort and cooperation
	I do not encounter any difficulties during TAD placement
	Other
Q23-Which factor do you think most affects the success of molar distalization treatment?	Patient age
	Severity of malocclusion
	TAD placement site
	Appliance stability
	Patient cooperation
	Other
Q24-What is the average duration of treatment for maxillary molar distalization?	0-3 months
	3-6 months
	6-9 months
	9-12 months
	More than 12 months
Q25-What are the most frequently encountered complications during distalization?	Root resorption in molars
	Miniscrew loosening
	Soft tissue irritation
	Miniscrew fracture
	No complications encountered
	Other
Q26-What is your average miniscrew failure rate during distalization?	0-10%
	10-20%
	20-30%
	Over 30%
Q27-What is the most common biomechanical challenge you face during molar distalization?	Unwanted tooth movement
	Poor vertical control
	Occlusal plane discrepancies
	Insufficient distalization
	No biomechanical difficulties encountered
	Other
Q28-How would you rate patient compliance during distalization?	Excellent
	Good
	Moderate
	Poor
Q29-Which clinical protocol do you most commonly follow?	I initiate molar distalization separately at the beginning of treatment, proceed to the retention phase after achieving a class I or super class I molar relationship, and then begin fixed orthodontic treatment following retention.
	I initiate molar distalization separately at the beginning of treatment and proceed directly to fixed orthodontic treatment immediately after achieving a Class I or super Class I molar relationship.
	I initiate molar distalization separately at the beginning of treatment and proceed to fixed orthodontic treatment before achieving a class I or super class I molar relationship.
	I start fixed orthodontic treatment before distalization, and once the appropriate stage is reached, I perform total maxillary arch distalization using archwires.

CBCT: Cone-beam computed tomography, TAD: Temporary anchorage devices

Table 3. Clear aligner and retention approaches for maxillary molar distalization treatment	
Q30- Do you prefer clear aligners for distalization indications in your clinical practice?	Yes
	No
Q31- If you use clear aligners, which brand do you use most frequently?	ClearCorrect
	FAS Aligner System
	Inhouse aligner
	Invisalign
	Orthero
	I do not use clear aligners for distalization indications.
	Other
Q32- In your distalization cases treated with clear aligners, which sequential distalization strategy do you prefer?	33% sequential distalization
	50% sequential distalization
	I move all teeth simultaneously.
	I do not use clear aligners for distalization indications.
Q33- At which stage do you start using class II elastics in your distalization cases treated with clear aligners?	From the beginning of treatment
	When the first premolar starts to move distally
	When the second premolar starts to move distally
	When the canine starts to move distally
	I do not use elastics
Q34- After how many millimeters of distalization do you apply skeletal anchorage support when using clear aligners?	2 mm
	3 mm
	4 mm
	5 mm
	6 mm
	I do not use clear aligners for distalization indications.
Q35- How do you manage the retention phase after completing the distalization stage?	I start fixed treatment with class II elastic support without waiting for the retention period.
	I keep the distalization appliance in the mouth for a while.
	I remove the distalization appliance and use a different appliance for full-time or part-time retention.
	I remove the distalization appliance and apply skeletal anchorage support (mini screw).
	I do not apply any mechanics and directly begin fixed treatment.
	Other
Q36- How do you manage the retention phase after completing fixed orthodontic treatment in patients treated with distalization?	I use Essix retainer
	I use lingual retainer
	I use both Essix and lingual retainer
	I use Hawley retainer
	I use a different appliance that I have designed myself or obtained externally
Q37- How do you evaluate the relapse rate following upper molar distalization?	No relapse observed
	Low rate (0%-10%)
	Moderate rate (10%-30%)
	High rate (≥30%)

aim of the study was to describe general trends in clinical preferences. In cases where the assumptions of the Chi-square test were not met due to expected frequencies falling below 5 in certain cells of the contingency tables, the Monte Carlo simulation method was applied using 10,000 iterations. This approach provided more robust and reliable p-values, particularly in low-frequency cells. All statistical analyses were performed using SPSS Version 22.0, and a p-value of less than 0.05 was considered statistically significant.

RESULTS

A total of 132 orthodontists participated in this study. The gender distribution was relatively balanced, with 54.5% (n=72) female and 45.5% (n=60) male participants. In terms

Table 4. Frequency and percentage distributions of responses to the survey questions							
		n	%			n	%
Q1	Male	60	45.5		Frog	2	1.5
	Female	72	54.5		Distal jet	20	15.2
Q2	20-25	4	3.0	Q15	Jones jig	2	1.5
	26-30	56	42.4		ACCO	8	6.1
	31-35	41	31.1		Other	6	4.6
	36-40	13	9.8		I do not use any	4	3.1
	Over 40	18	13.6		Infrazygomatic crest	83	62.9
Q3	Dental health center	2	1.5	Q16	Buccal	21	15.9
	University hospital	94	71.2		Maxillary tuber	2	1.5
	Private clinic	36	27.3		Palatinal	22	16.7
Q4	Research assistant	61	46.2		I do not use any	44	33.3
	Specialist dentist	38	28.8		Beneslider	17	12.9
	Lecturer	2	1.5		Modified Pendulum	12	9.1
	Assistant professor	20	15.2	Q17	Modified Keleş Slider	38	28.8
	Associate professor	6	4.5		Miniscrew supported frog	2	1.5
	Professor	5	3.8		Modified distal jet	14	10.6
Q5	0-3 years	28	21.2		Others	5	4.0
	3-6 years	54	40.9		Yes	35	27.3
	6-10 years	17	12.9	Q18	No	56	43.8
	Over 10 years	33	25.0		Sometimes	37	28.9
Q6	Yes	132	100.0		Extraradicular	52	40.0
	No	0	0	Q19	Interradicular	44	33.8
Q7	Panoramic	118	89.4		I use both equally	26	20.0
	Cephalometric	54	40.9		I do not use	8	6.2
	CBCT	15	11.4		2x10 mm	7	5.3
	Digital setup	11	8.3		2x11 mm	3	2.3
Q8	Children	12	9.1	Q20	2x12 mm	91	68.9
	Adolescents	103	78.0		2x14 mm	11	8.3
	Adults	17	12.9		I do not use	18	13.6
Q9	Yes	124	93.9		Other	2	1.51
	No	8	6.1		Stainless steel	72	54.5
Q10	Distalization	48	36.4	Q21	Titanium alloy	43	32.6
	Tooth extraction	84	63.6		I do not use	17	12.9
Q11	Molar relationship	74	56.1		Bone density	63	47.7
	Crowding	64	48.5		Soft tissue	41	31.1
	Vertical growth	59	44.7	Q22	Device instability	32	24.2
	Soft tissue profile	50	37.9		Patient discomfort	29	22.0
	Treatment duration	24	18.2		I don't any difficulties	20	15.2
	Appliance cost	5	3.8		Other	1	0.8
	All of the above	41	31.1		Patient age	26	19.7
Q12	molar relationship	49	37.1		Severitymalocclusion	50	37.9
	Crowding	32	24.2		TAD placement site	38	28.8
	Vertical growth	28	21.2	Q23	Appliance stability	57	43.2
	Soft tissue profile	18	13.6		Patient cooperation	47	35.6
	Treatment duration	3	2.3		Other	3	2.3
	Others	2	1.5		0-3 months	1	0.8
Q13	Occasionally	26	19.7		3-6 months	50	37.9
	Yes	13	9.8	Q24	6-9 months	58	43.9
	No	93	70.5		9-12 months	18	13.6
Q14	I do not use any	2	1.5		More than 12 months	5	3.8
	Tooth-tissue support	18	13.6	Q25	Root resorption molars	4	3.0
	Skeletal anchorage	112	84.8		Miniscrew loosening	94	71.2

Table 4. Frequency and percentage distributions of responses to the survey questions (table continues)							
Q15	I do not use any	42	31.8	Q25	Soft tissue irritation	66	50.0
	Pendulum	45	34.1		Miniscrew fracture	3	2.3
	Keleş slider	50	37.9		No complications	7	5.3
	Carriere distalizer	8	6.1		Other	3	2.3
Q26	0-10%	53	40.2		beginning of treatment	26	19.7
	10-20%	44	33.3		first premolar starts move	16	12.1
	20-30%	28	21.2	Q33	second premolar starts move	47	35.6
	Over 30%	7	5.3		canine starts to move	6	4.5
Q27	Unwanted tooth movement	36	27.3		I do not use elastics	2	1.5
	Poor vertical control	21	15.9		From the beginning	26	19.7
	Occlusal plane discrepancies	43	32.6		2 mm	13	9.8
	Insufficient distalization	80	60.6		3 mm	55	41.7
	No biomechanical difficulties	8	6.1	Q34	4 mm	25	18.9
	Other	2	1.5		5 mm	2	1.5
Q28	Excellent	2	1.5		6 mm	1	0.8
	Good	63	47.7		I do not use	36	27.3
	Moderate	64	48.5		fixed treatment with class II elastic	21	15.9
	Poor	3	2.3		keep the distalization appliance	76	57.6
Q29	Molar distalization and retention phase	11	8.3		I use a different appliance	5	3.8
	Molar distalization and immediately orthodontic treatment	47	35.6	Q35	I apply skeletal anchorage	16	12.1
	Molar distalization and orthodontic treatment before class I relationship	10	7.6		I do not apply any mechanics	10	7.6
	Orthodontic treatment and total arch distalization	64	48.5		Other	4	3.0
Q30	Yes	89	67.4		I use Essix retainer	24	18.2
	No	43	32.6		I use lingual retainer	9	6.8
Q31	ClearCorrect	14	10.6	Q36	I use both	95	72.0
	FAS Aligner System	1	0.8		I use Hawley retainer	3	2.3
	Inhouse aligner	2	1.5		I use a different appliance	1	0.8
	Invisalign	75	56.8		No relapse observed	10	7.6
	Orthero	7	5.3		Low rate (0%-10%)	73	55.3
	I do not use	33	25.0	Q37	Moderate rate (10%-30%)	42	31.8
Q32	33% sequential distalization	48	36.4		High rate (≥30%)	7	5.3
	50% sequential distalization	47	35.6				
	Move all teeth simultaneously	35	26.5				
	I do not use clear aligners for distalization	2	1.5				

Note: This table presents the frequency (n) and percentage (%) distributions of all responses to the structured questionnaire items administered to orthodontists. Percentages are calculated based on the total number of valid responses per item. Multiple-response questions were allowed for some items, and their percentages may exceed 100%. CBCT: Cone-beam computed tomography, TAD: Temporary anchorage devices

of age, the majority (42.4%, n=56) were between 26-30 years, followed by 31-35 years (31.1%, n=41), 40 years and above (13.6%, n=18), 36-40 years (9.8%, n=13), and 20-25 years (3.0%, n=4) (Table 4).

Regarding the institution type, most participants were working in university hospitals (71.2%, n=94), followed by private clinics or polyclinics (27.3%, n=36), and public dental health centers (1.5%, n=2). The academic and professional title distribution showed that 46.2% (n=61) were research assistants, 28.8% (n=38) specialist dentists, 15.2% (n=20) assistant professors, 4.5% (n=6) associate professors, 3.8% (n=5) professors, and 1.5% (n=2) lecturers (Table 4).

According to professional experience, 40.9% (n=54) had been practicing orthodontics for 3-6 years, 25.0% (n=33) for 10 years or more, 21.2% (n=28) for 0-3 years, and 12.9% (n=17) for 6-10 years.

All respondents reported applying upper molar distalization in their clinical practice. A large majority (89.4%) indicated using panoramic radiographs in distalization planning, followed by cephalometric radiographs (40.9%), cone beam computed tomography (CBCT) (11.4%), and digital setup/3D modeling (8.3%). Distalization was most frequently performed in adolescent patients (78.0%), with lower frequencies in adults (12.9%) and children (9.1%). Most respondents (93.9%) reported routinely extracting third molars before initiating distalization (Table 4).

In class II camouflage cases, 63.6% of orthodontists preferred upper premolar extraction, while 36.4% opted for molar distalization. Factors influencing this decision included severity of molar relationship (56.1%), amount of crowding (48.5%), vertical growth pattern and overbite depth (44.7%), and soft tissue profile (37.9%). The molar relationship was the most frequently cited primary factor (37.1%) during initial assessment (Table 4).

Regarding extraoral appliances, 70.5% reported never using them, 19.7% used them occasionally, and only 9.8% used them regularly. For intraoral distalization, skeletal anchorage-supported systems were preferred by 84.8% of participants, whereas 13.6% used tooth- and tissue-supported appliances. The most commonly used conventional appliances were Keles Slider (37.9%), Pendulum (34.1%), and Distal Jet (15.2%) (Table 4).

For skeletal anchorage site selection, the most preferred location was the infrazygomatic crest (62.9%), followed by the palatal region (16.7%) and buccal region (15.9%). Among palatal anchorage systems, modified Keles Slider (28.8%), Beneslider (12.9%), and modified Pendulum (9.1%) were the most frequently used. While 43.8% of respondents did not use CBCT prior to palatal miniscrew placement, 27.3% routinely used it, and 28.9% used it selectively (Table 4).

In buccal miniscrew applications, 40.0% preferred extraradicular placement, 33.8% used interradicular placement, and 20.0% used both equally. The most common screw size in the infrazygomatic region was 2×12 mm, reported by 68.9% of participants. Regarding material preference, 54.5% used stainless steel and 32.6% used titanium alloy screws (Table 4).

The most common difficulty encountered during TAD placement was inadequate bone density (47.7%), followed by soft tissue interferences (31.1%) and device stability issues (24.2%). Factors perceived as most critical to treatment success included appliance stability (43.2%), severity of malocclusion (37.9%), and patient cooperation (35.6%). Average distalization duration was reported as 6-9 months by 43.9% and 3-6 months by 37.9% of the respondents (Table 4).

The most frequently observed complications during distalization were miniscrew loosening (71.2%) and soft tissue irritation (50.0%). A total of 40.2% reported a miniscrew failure rate between 0-10%, and 33.3% between 10-20%. The most common biomechanical challenge was insufficient distalization (60.6%), followed by occlusal plane disturbances (32.6%) and unwanted tooth movements (27.3%) (Table 4).

Regarding patient compliance, 48.5% rated it as “moderate” and 47.7% as “good.” In clinical practice, 48.5% reported initiating full arch distalization before fixed treatment, whereas 35.6% transitioned to fixed appliances after completing molar distalization (Table 4).

A total of 67.4% of clinicians reported using clear aligners for distalization cases. Among them, Invisalign was the most preferred brand (56.8%). Regarding distalization strategy, 33% staging (36.4%) and 50% staging (35.6%) were the most common. Class II elastics were typically initiated during the distal movement of the second premolars (35.6%). Respondents reported initiating skeletal anchorage support after approximately 3 mm of distalization (Table 4).

After the active distalization phase, 57.6% of clinicians kept the appliance in place temporarily for retention, while 15.9% moved directly into fixed treatment with class II elastics without a retention phase. The most commonly used post-treatment retention method was a combination of Essix retainer and lingual fixed retainer (72.0%). In terms of relapse, 55.3% reported a low relapse rate (0-10%), while 31.8% observed relapse between 10-30% (Table 4).

The study also examined the influence of demographic factors-such as age, gender, workplace type, and years of experience-on clinical decisions regarding appliance selection, skeletal anchorage use, and clear aligner applications. Statistical analyses revealed that age and gender had no significant impact on clinical preferences ($p>0.05$) (Table 5).

When clinical preferences were compared by institution type, four items showed statistically significant differences ($p<0.05$). Panoramic radiographs were more frequently used in private clinics, while cephalometric and CBCT imaging were more common in universities and public institutions (Q7). The infrazygomatic crest was more preferred in universities/public settings for TAD placement, while palatal placement was more common in private clinics (Q16). Stainless steel miniscrews were used more often in university settings (Q21), and clear aligners were significantly more common in private practice (Q30). No significant associations were found between institution type and other clinical variables ($p>0.05$), indicating a general standardization in many treatment approaches (Table 5).

Table 5. Chi-square test results: effects of gender, age, institution type and academic title on clinical preferences				
Questions	**Effect of gender (p)**	**Effect of age (p)**	**Effect of institution type (p)**	**Effect of academic title (p)**
Q-7	>0.05	>0.05	0.048*	>0.05
Q-16	>0.05	>0.05	0.017*	0.004**
Q-21	>0.05	>0.05	0.0005***	0.017*
Q-30	>0.05	>0.05	0.006**	0.021*
Q-31	>0.05	>0.05	>0.05	0.016*
Q-32	>0.05	>0.05	>0.05	0.010**
Q-33	>0.05	>0.05	>0.05	0.024*
Q-34	>0.05	>0.05	>0.05	0.009**
Other questions	>0.05	>0.05	>0.05	> 0.05
Statistically significant associations (p<0.05) are indicated. Non-significant results are presented as >0.05. *p<0.05; **p<0.01; **p<0.001				

Further analysis by academic rank revealed significant differences in some items (p<0.05). In Q16, research assistants mostly preferred the infrazygomatic crest, while faculty members more frequently chose the palatal region. In Q21, stainless steel screws were the most used in all groups, but titanium preference was higher among specialists. Regarding clear aligners, specialists reported the highest usage (Q30), Invisalign was the most commonly used brand across all groups (Q31), and 50% staging was more common among faculty and specialists, while 33% staging was preferred by research assistants (Q32). In Q33, specialists were more likely to initiate class II elastics during second premolar movement. In Q34, while most clinicians reported using skeletal anchorage after 3 mm of movement, this was more pronounced among faculty. No significant differences were found in other questions based on academic title (p>0.05) (**Table 5**).

DISCUSSION

In managing angle class II dental malocclusions, two common approaches for addressing maxillary anterior crowding and increased overjet include the distal movement of maxillary molars or the extraction of premolars. Advances in mechanotherapy and evolving treatment philosophies have significantly reduced the reliance on premolar extractions in various malocclusion types. Borderline cases, however, continue to present clinical challenges and differing opinions among practitioners. It has been suggested that approximately 25-30% of orthodontic patients may benefit from maxillary arch expansion, while up to 95% of class II cases could potentially be improved through a combination of molar rotation, distalization, and expansion.[7,16,17]

To the best of our knowledge, there is no prior survey-based study in the literature that specifically evaluates clinicians' preferences regarding the use of distalization methods in the treatment of class II malocclusions. In this context, the present study contributes to the field by exploring the distalization strategies adopted by orthodontists.

Upper molar distalization in orthodontics is commonly performed using extraoral or intraoral approaches. While headgear has shown clinical effectiveness, its use has declined due to aesthetic concerns, reliance on patient compliance, and reported complications such as soft tissue irritation and muscle strain.[18,19] These limitations have prompted a shift toward intraoral appliances, which offer continuous force application without the need for extraoral support. However, these systems often lead to anchorage loss in premolars and incisors, and when these teeth are later repositioned, further anchorage challenges and treatment delays may occur.[20]

Recent advancements in skeletal anchorage systems, particularly buccal and palatal miniscrew-supported mechanics, have enabled effective distalization in managing even severe class II malocclusions. These systems provide superior three-dimensional control, minimize unwanted side effects, and significantly reduce the risk of anchorage loss, thereby diminishing the need for extractions or extraoral devices.[5]

In this study, all respondents indicated that they incorporate upper molar distalization into their clinical practice. This finding suggests that, although not always the primary treatment choice-particularly in cases where premolar extraction is preferred-distalization remains a widely accepted and routinely utilized approach in managing specific malocclusion patterns.

Panoramic radiographs were the most frequently preferred imaging modality during distalization planning, likely due to their ease of use, low cost, and ability to provide basic diagnostic information.[21] Although cephalometric radiographs and CBCT offer more detailed skeletal and three-dimensional assessments, their relatively limited use may reflect practical constraints such as radiation concerns, availability, or institutional routines.[22] The low utilization of digital setup and 3D modeling also indicates that these technologies, while promising, have not yet become standard in everyday clinical workflows.

The majority of clinicians reported performing distalization most frequently in adolescent patients (78%), which aligns with the optimal timing for molar movement due to favorable growth potential and anchorage conditions during this period.[23]

A large proportion of respondents (93.9%) indicated that they routinely extract third molars before initiating distalization. This common practice is likely aimed at preventing eruption-related interferences and facilitating unobstructed molar

movement, as supported by previous studies recommending the removal of third molars to optimize distalization efficiency.[24,25]

Despite the widespread use of distalization, 63.6% of clinicians reported favoring premolar extraction over distalization in class II camouflage cases. This preference may reflect the greater predictability and anchorage control associated with extraction protocols, especially in patients with significant crowding or pronounced skeletal discrepancies. The decision appears to be shaped not only by treatment mechanics but also by factors such as malocclusion severity and long-term stability considerations.

Although distalization is widely utilized in clinical practice, a majority of clinicians (63.6%) reported preferring premolar extraction over distalization for class II camouflage treatment. According to the responses, this decision was primarily influenced by the severity of the molar relationship, the amount of crowding, the patient's vertical growth pattern or overbite depth, and the soft tissue profile. When asked about the 'primary criterion' considered during initial assessment, the molar relationship was again the most commonly selected factor (37.1%), highlighting its dominant role in treatment planning. These findings suggest that treatment planning is multifactorial, and clinicians weigh skeletal and dental characteristics carefully when determining whether to choose extraction or distalization.[26]

Extraoral distalization appliances were rarely used, with 70.5% of clinicians reporting that they do not incorporate them into treatment. This low preference is likely related to aesthetic concerns, limited patient compliance, and the availability of more effective intraoral alternatives.[5]

Intraoral distalization was predominantly performed using skeletal anchorage-supported systems, preferred by 84.8% of clinicians, while only 13.6% reported using tooth- and tissue-supported appliances. This finding reflects the growing reliance on miniscrew-assisted mechanics due to their superior anchorage control and reduced side effects.[17] Among traditional appliances, the Keleş Slider (37.9%) and Pendulum (34.1%) remained the most frequently used, indicating that despite the shift toward skeletal anchorage, conventional systems still hold a place in selected cases.

Clinicians in this study most frequently preferred the infrazygomatic crest (62.9%) as the site for skeletal anchorage placement, with lower rates for palatal (16.7%) and buccal (15.9%) regions. This preference may be attributed to the IZC region's favorable cortical bone density, ease of access without the need for complex appliances, and cost-effectiveness compared to palatal systems that often require custom laboratory components.[27] Additionally, its compatibility with direct force application makes it a practical choice in routine clinical settings.[28]

Among palatal miniscrew-supported systems, the most commonly preferred appliance was the modified Keleş Slider (28.8%), followed by the Beneslider (12.9%) and the modified Pendulum (9.1%). These preferences may reflect clinicians' familiarity with specific biomechanics, ease of appliance activation, and prior clinical training. The modified Keleş Slider, in particular, offers controlled molar movement with minimal reliance on patient compliance, which may explain its frequent use in palatal anchorage protocols.[29] Despite the anatomical complexity of the palatal region, only 27.3% of clinicians reported routinely using CBCT prior to miniscrew placement, while 43.8% did not use it at all. Given the risk of root damage and the need for precise identification of adequate bone volume, CBCT imaging is often considered essential in planning safe and effective miniscrew insertion.[30] The limited use observed in this study may be attributed to factors such as radiation concerns, additional cost, or lack of routine access to CBCT in certain clinical environments.

In the buccal region, clinicians showed a slight preference for extraradicular miniscrew placement (40.0%) over interradicular sites (33.8%). This may be due to the increased risk of root proximity in interradicular applications, especially when anatomical spacing is limited. Extraradicular sites may offer more consistent cortical engagement and lower risk of root contact, making them a safer option in selected cases.[31] The remaining clinicians reported using both approaches equally, likely adapting their choice based on individual anatomical considerations.

The 2×12 mm miniscrew was the most commonly preferred dimension in this study (68.9%), aligning with previous literature suggesting that this length offers optimal balance between mechanical stability and safety in extra-alveolar sites such as the infrazygomatic crest. Its sufficient length ensures effective cortical engagement while minimizing the risk of root proximity or maxillary sinus perforation. Regarding material preference, stainless steel miniscrews were selected more frequently (54.5%) than titanium (32.6%), likely due to their higher fracture resistance and cost-effectiveness in high-stress clinical applications.[32] Although titanium is known for its superior biocompatibility, its increased flexibility and higher cost[33] may limit its routine use in heavy-load mechanics like distalization.

The most frequently reported challenge during TAD placement was inadequate bone density (47.7%), a finding consistent with previous studies emphasizing the importance of cortical bone thickness for primary stability.[34] Insufficient bone support may compromise miniscrew retention, particularly in anatomically variable regions such as the infrazygomatic crest or palatal slope. Soft tissue interference (31.1%) and appliance instability (24.2%) were also noted as limiting factors. Notably, appliance stability (43.2%) was cited as the most critical determinant of treatment success, followed by malocclusion severity (37.9%) and patient cooperation (35.6%), reflecting the multifactorial demands of effective distalization mechanics.

Most clinicians reported an average distalization duration of 6-9 months (43.9%), which is consistent with previous studies reporting similar treatment timelines for molar distalization using both skeletal and conventional intraoral mechanics.[35,36] Variations in treatment duration may depend on factors such as the amount of distal movement required, appliance design, and anchorage quality.

The most commonly reported complication during distalization was miniscrew loosening (71.2%), followed by

soft tissue irritation (50.0%). These findings are consistent with previous studies identifying primary stability loss and soft tissue overgrowth as frequent issues in TAD-based mechanics.[37,38] Despite these complications, most clinicians reported relatively low miniscrew failure rates, with 40.2% estimating a loss rate between 0-10%, and 33.3% between 10-20%. The most frequent biomechanical challenge was insufficient distalization (60.6%), likely related to anatomical limitations or force application inefficiencies, while occlusal plane alterations and unwanted tooth movements were reported to a lesser extent.

Following active distalization, the majority of clinicians (57.6%) preferred maintaining the appliance intraorally for a period to ensure retention before initiating fixed therapy. This strategy may enhance post-distalization stability by allowing for periodontal and occlusal adaptation. After comprehensive treatment, the most commonly used retention protocol was a combination of Essix and lingual retainers (72.0%), likely reflecting efforts to minimize relapse risk through dual mechanical control. Despite these precautions, 31.8% of respondents reported moderate relapse rates (10-30%), suggesting that even with reinforcement, distalized molars may be susceptible to post-treatment movement.

With rising aesthetic demands and advancements in aligner technology, clear aligners have become a viable option for performing complex tooth movements, including molar distalization. In this study, 67.4% of clinicians reported using aligners for distalization, with Invisalign being the most preferred brand (56.8%). This finding aligns with recent surveys and clinical reports indicating a growing reliance on clear aligners for class II correction, particularly among practitioners seeking aesthetic, compliance-friendly alternatives.[39] Sequential distalization protocols were common, with 33% and 50% staging strategies most frequently employed-an approach shown to improve anchorage control and reduce undesired reciprocal movements.[40] Notably, 35.6% of respondents initiated class II elastic use during the distalization of the second premolars, which is consistent with current recommendations that favor delayed elastic engagement to prevent premature anchorage loss.[41] Furthermore, clinicians reported initiating skeletal anchorage-typically in the form of palatal TADs-after an average of 3 mm of distal movement, reflecting the limitations of aligners alone in achieving bodily molar translation without auxiliary support. This threshold is consistent with previous clinical and biomechanical studies suggesting that clear aligners can predictably achieve 2-3 mm of molar distalization, though primarily with distal tipping rather than bodily movement.[42] Beyond this point, the incorporation of TADs or class II elastics has been shown to significantly improve anchorage control and enhance the efficiency of posterior tooth movement.

This study also explored the potential influence of demographic and professional variables on clinical decision-making. While factors such as age and gender showed no significant association with treatment preferences, institutional setting and professional experience were found to impact specific choices-particularly in imaging modality, anchorage site selection, and clear aligner use. For example, university-based clinicians more frequently preferred infrazygomatic anchorage and steel alloy miniscrews, whereas private practitioners showed greater use of palatal TADs and aligner therapy. Additionally, variations were observed across academic titles, suggesting that training background and clinical exposure may influence appliance selection and biomechanics. These findings indicate that, although many clinical approaches appear standardized, institutional resources and practitioner experience can still shape treatment planning in molar distalization.

Limitations

While the results offer valuable insights into treatment preferences, they are based on self-reported data and may be influenced by recall bias or institutional variability. Additionally, the number of participants could have been higher to further strengthen the generalizability of the findings. Future studies with clinical outcome data and broader international samples are needed to validate these patterns and inform evidence-based protocols. Moreover, multivariate modeling approaches are recommended in future research to evaluate the independent effects of variables such as age, clinical experience, and academic title, which were not explored in the current study.

CONCLUSION

This survey-based study provides a comprehensive overview of current clinical practices among orthodontists in Turkiye regarding maxillary molar distalization. The findings indicate a strong preference for skeletal anchorage-supported intraoral appliances, particularly those utilizing infrazygomatic crest and palatal insertion sites. Although distalization is widely utilized, premolar extraction remains the more common approach in class II camouflage cases. Clear aligner systems have gained significant popularity, especially when combined with sequential staging and auxiliary anchorage. Despite these advancements, concerns such as miniscrew stability, anatomical limitations, and relapse remain critical factors influencing treatment success. These results highlight the need for individualized biomechanical planning and continued evaluation of long-term clinical outcomes in distalization therapy.

ETHICAL DECLARATIONS

Ethics Committee Approval

This study was approved by the İnönü University Scientific Researches and Publication Ethics Committee (Date: 24.09.2024, Decision No: 2024/6452).

Informed Consent

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

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

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

Author Contributions

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

REFERENCES

1. Angle EH. Classification of malocclusion. *Dent Cosmos*. 1899;41:350-357.
2. Stomatologic S. Worldwide prevalence of malocclusion in the different stages of dentition: a systematic review and meta-analysis. *Eur J Paediatr Dent*. 2020;21(2):115-122. doi:10.23804/ejpd.2020.21.02.05
3. Gelgör IE, Karaman IA, Ercan E. Prevalence of malocclusion among adolescents in central Anatolia. *Eur J Dent*. 2007;1(3):125-131. doi:10.1055/s-0039-1698327
4. Graber T, Swain B. Orthodontics Current Principles and Techniques. St. Louis: CV Mosby Co; 1985:681.
5. Oğuz F, Özden S, Cicek O. Distalization methods for maxillary molars utilizing temporary anchorage devices (TADs): a narrative review. *Appl Sci*. 2024;14(23):11333. doi:10.3390/app142311333
6. Bowman SJ, Johnston LE Jr. The esthetic impact of extraction and nonextraction treatments on Caucasian patients. *Angle Orthod*. 2000;70(1):3-10. doi:10.1043/0003-3219(2000)070<0003:TEIOEA>2.0.CO;2
7. Abdelhady NA, Tawfik MA, Hammad SM. Maxillary molar distalization in treatment of angle class II malocclusion growing patients: uncontrolled clinical trial. *Int Orthod*. 2020;18(1):96-104. doi:10.1016/j.ortho.2019.11.003
8. Grec RH, Janson G, Branco NC, Moura-Grec PG, Patel MP, Castanha Henriques JF. Intraoral distalizer effects with conventional and skeletal anchorage: a meta-analysis. *Am J Orthod Dentofacial Orthop*. 2013;143(5):602-615. doi:10.1016/j.ajodo.2012.11.024
9. Kırcalı M, Yüksel AS. Evaluation of dentoalveolar and dentofacial effects of a mini-screw-anchored pendulum appliance in maxillary molar distalization. *Turk J Orthod*. 2018;31(4):103-109. doi:10.5152/TurkJOrthod.2018.18004
10. AlSeraidi M, Hansa I, Dhaval F, Ferguson DJ, Vaid NR. The effect of vestibular, lingual, and aligner appliances on the quality of life of adult patients during the initial stages of orthodontic treatment. *Prog Orthod*. 2021;22(1):3. doi:10.1186/s40510-020-00346-0
11. Weir T. Clear aligners in orthodontic treatment. *Aust Dent J*. 2017;62(Suppl 1):58-62. doi:10.1111/adj.12480
12. Drake CT, McGorray SP, Dolce C, Nair M, Wheeler TT. Orthodontic tooth movement with clear aligners. *ISRN Dent*. 2012;2012:657973. doi:10.5402/2012/657973
13. Simon M, Keilig L, Schwarze J, Jung BA, Bourauel C. Treatment outcome and efficacy of an aligner technique-regarding incisor torque, premolar derotation and molar distalization. *BMC Oral Health*. 2014;14:68. doi:10.1186/1472-6831-14-68
14. Bearn DR. Bonded orthodontic retainers: a review. *Am J Orthod Dentofacial Orthop*. 1995;108(2):207-213. doi:10.1016/s0889-5406(95)70085-4
15. Pandis N, Vlahopoulos K, Madianos P, Eliades T. Long-term periodontal status of patients with mandibular lingual fixed retention. *Eur J Orthod*. 2007;29(5):471-476. doi:10.1093/ejo/cjm042
16. Corbett MC. Slow and continuous maxillary expansion, molar rotation, and molar distalization. *J Clin Orthod*. 1997;31(4):253-263.
17. Soheilifar S, Mohebi S, Ameli N. Maxillary molar distalization using conventional versus skeletal anchorage devices: a systematic review and meta-analysis. *Int Orthod*. 2019;17(3):415-424. doi:10.1016/j.ortho.2019.06.002
18. Bolla E, Muratore F, Carano A, Bowman SJ. Evaluation of maxillary molar distalization with the distal jet: a comparison with other contemporary methods. *Angle Orthod*. 2002;72(5):481-494. doi:10.1043/0003-3219(2002)072<0481:EOMMDW>2.0.CO;2
19. Samuels RH, Jones ML. Orthodontic facebow injuries and safety equipment. *Eur J Orthod*. 1994;16(5):385-394. doi:10.1093/ejo/16.5.385
20. Bondemark L, Karlsson I. Extraoral vs intraoral appliance for distal movement of maxillary first molars: a randomized controlled trial. *Angle Orthod*. 2005;75(5):699-706. doi:10.1043/0003-3219(2005)75[699:EVIAFD]2.0.CO;2
21. Izzetti R, Nisi M, Aringhieri G, Crocetti L, Graziani F, Nardi C. Basic knowledge and new advances in panoramic radiography imaging techniques: a narrative review on what dentists and radiologists should know. *Appl Sci*. 2021;11(17):7858. doi:10.3390/app11177858
22. De Vos W, Casselman J, Swennen G. Cone-beam computerized tomography (CBCT) imaging of the oral and maxillofacial region: a systematic review of the literature. *Int J Oral Maxillofac Surg*. 2009;38(6):609-625. doi:10.1016/j.ijom.2009.02.028
23. Ruan MJ, Chen G, Xu TM. Comparison of orthodontic tooth movement between adolescents and adults based on implant superimposition. *PLoS One*. 2018;13(5):e0197281. doi:10.1371/journal.pone.0197281
24. Aboalnaga AA, Fouda AS. Evaluation of the effect of extraction in comparison to distalization on the maxillary third molars in class II malocclusion: a retrospective study. *Clin Oral Investig*. 2024;28(3):191. doi:10.1007/s00784-024-05576-8
25. Kassem NM, Farouk K. Effect of upper third molar extraction on distalization using carriere motion appliance: a prospective clinical study. *Al-Azhar J Dent Sci*. 2024;27(3):407-416. doi:10.21608/ajdsm.2022.174768.1384
26. Alsaggaf DH, Afify AR, Zawawi KH, Alsulaimani FF. Factors influencing the orthodontic treatment plan in class II malocclusion. *Am J Orthod Dentofacial Orthop*. 2022;161(6):829-837.e1. doi:10.1016/j.ajodo.2021.01.034
27. Chang C, Huang C, Roberts WE. 3D cortical bone anatomy of the mandibular buccal shelf: a CBCT study to define sites for extra-alveolar bone screws to treat class III malocclusion. *Int J Orthod Implantol*. 2016;41(1):74-82.
28. Wu X, Liu H, Luo C, Li Y, Ding Y. Three-dimensional evaluation on the effect of maxillary dentition distalization with miniscrews implanted in the infrazygomatic crest. *Implant Dent*. 2018;27(1):22-27. doi:10.1097/ID.0000000000000706
29. Özdemir G. Kemik ve diş-doku destekli keleş slider apareylerinin etkilerinin 3 boyutlu görüntüleme yöntemiyle karşılaştırılması. Doktora Tezi, Atatürk Üniversitesi Sağlık Bilimleri Enstitüsü, 2013.
30. Jakovljevic I, Vasiljevic M, Milanovic J, et al. Importance of CBCT analysis in the preoperative planning of TAD placement in the anterior maxillary region. *Appl Sci*. 2025;15(12):6866. doi:10.3390/app15126866
31. Truong VM, Kim S, Kim J, Lee JW, Park YS. Revisiting the complications of orthodontic miniscrew. *Biomed Res Int*. 2022;2022:8720412. doi:10.1155/2022/8720412
32. Liou EJ, Pai BC, Lin JC. Do miniscrews remain stationary under orthodontic forces? *Am J Orthod Dentofacial Orthop*. 2004;126(1):42-47. doi:10.1016/j.ajodo.2003.06.018
33. Mecenas P, Espinosa DG, Cardoso PC, Normando D. Stainless steel or titanium mini-implants? *Angle Orthod*. 2020;90(4):587-597. doi:10.2319/081619-536.1
34. Miyawaki S, Koyama I, Inoue M, Mishima K, Sugahara T, Takano-Yamamoto T. Factors associated with the stability of titanium screws placed in the posterior region for orthodontic anchorage. *Am J Orthod Dentofacial Orthop*. 2003;124(4):373-378. doi:10.1016/s0889-5406(03)00565-1
35. Memè L, Bambini F, Del Vecchio G, et al. Dental distalization techniques with clear aligners: a narrative review. *Oral Implantol*. 2024;16(3 Suppl 1):461-475. doi:10.11138/oi163.1suppl461-475
36. Papadopoulos MA, Tarawneh F. The use of miniscrew implants for temporary skeletal anchorage in orthodontics: a comprehensive review. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*. 2007;103(5):e6-e15. doi:10.1016/j.tripleo.2006.11.022
37. Chen YJ, Chang HH, Huang CY, Hung HC, Lai EH, Yao CC. A retrospective analysis of the failure rate of three different orthodontic skeletal anchorage systems. *Clin Oral Implants Res*. 2007;18(6):768-775. doi:10.1111/j.1600-0501.2007.01405.x
38. Park HS. The use of microimplants in orthodontics. In: Current Therapy in Orthodontics. Elsevier; 2010:291-300.
39. Dianiskova S, Rongo R, Buono R, Franchi L, Michelotti A, D'Antò V. Treatment of mild class II malocclusion in growing patients with clear aligners versus fixed multibracket therapy: a retrospective study. *Orthod Craniofac Res*. 2022;25(1):96-102. doi:10.1111/ocr.12500

40. Shen C, Park TH, Chung CH, Li C. Molar Distalization by clear aligners with sequential distalization protocol: a systematic review and meta-analysis. *J Funct Biomater.* 2024;15(6):137. doi:10.3390/jfb15060137
41. Lombardo L, Colonna A, Carlucci A, Oliverio T, Siciliani G. Class II subdivision correction with clear aligners using intermaxilary elastics. *Prog Orthod*. 2018;19(1):32. doi:10.1186/s40510-018-0221-5
42. Ravera S, Castroflorio T, Garino F, Daher S, Cugliari G, Deregibus A. Maxillary molar distalization with aligners in adult patients: a multicenter retrospective study. *Prog Orthod*. 2016;17:12. doi:10.1186/s40510-016-0126-0

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1695337
J Health Sci Med. 2025;8(5):787-793

Bonding treatment of CAD/CAM milled denture resins repaired with visible light-cured resin

Gökçen Dinçer[1], Şebnem Özatik[2], Sina Saygılı[1], Canan Bural Alan[1]

[1]Department of Prosthodontics, Faculty of Dentistry, İstanbul University, İstanbul, Turkiye
[2]Department of Dental Prosthetics Technology, Vocational School of Health Services, İstanbul Bilgi University, İstanbul, Turkiye

Cite this article as: Dinçer G, Özatik Ş, Saygılı S, Bural Alan C. Bonding treatment of CAD/CAM milled denture resins repaired with visible light-cured resin. *J Health Sci Med.* 2025;8(5):787-793.

Received: 08.05.2025 ◆ **Accepted**: 28.07.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: This study evaluated and compared the effect of bonding agents on the flexural strength (FS) of denture base resins repaired with visible light cured (VLC) denture resin.

Methods: A total of 100 specimens (65x10x2.5 mm) were fabricated using two types of denture base materials: pre-polymerized PMMA-based blocks designed for CAD/CAM milling and conventional heat-polymerized denture base acrylic resin (control). The specimens were sectioned in the middle with 2 mm repair gap and 45° margin design. Repair surfaces were first treated with various light-cured bonding agents then repaired using VLC resin. The bonding agents either conventional or combined with acrylic primers and dual cure agents were tested. All the specimens were subjected to 3-point bending test and FS was calculated. Data were statistically analyzed using two-way analysis of variance according to the denture base material and the bonding treatments ($p<0.05$).

Results: Among repaired groups, acrylic primer + G-Premio BOND produced the highest FS within each material (A3: 17.31±4.69 MPa; B3: 9.80±2.57 MPa). Between materials, CAD/CAM exceeded conventional in groups 1-4 ($p<0.05$)-including the intact controls-whereas group 5 showed no between-material difference ($p>0.05$).

Conclusion: The use of a bonding agent specifically designed for the surface treatment of acrylic resins can be clinically beneficial when repairing denture bases with VLC resin.

Keywords: Acrylic primer, bonding agent, CAD/CAM, denture repair, flexural strength

INTRODUCTION

Computer-aided design and manufacturing (CAD/CAM) technologies have gained popularity in removable denture fabrication due to their numerous clinical and technical advantages.[1-3] Digitally fabricated dentures-defined as prostheses milled from pre-polymerized polymethyl methacrylate (PMMA) blocks using CAD/CAM systems-address several limitations associated with conventionally moulded PMMA dentures, which are typically produced through compression molding of heat-polymerized acrylic resin.[4] These digitally produced prostheses have been associated with improved patient and clinician satisfaction, primarily due to fewer required clinical appointments.[5,6] Additionally, they offer superior fit, reduced polymerization shrinkage,[7] lower microbial adhesion,[8] and the advantages of digital data storage and rapid reproducibility.[9,10]

Common complications associated with complete dentures include cracks, fractures, and debonding of artificial teeth, with fractures reported as the most frequent.[11-14] Elderly patients, who constitute the majority of denture wearers, often experience accidental denture fractures due to weakened reflexes and reduced motor control.[14,15] Additionally, poor denture design[16] and insufficient mechanical properties of denture base materials contribute significantly to denture failures.[17,18]

The re-fabrication of digital prosthetic restorations is optimal in the presence of any complications, the financial implications associated with computer systems and the requisite materials represent a considerable economic drawback. Consequently, the repair of these systems, which have emerged as the prevailing treatment modality, is paramount in selecting suitable repair materials and surface modifications.[19] Repair materials must be widely used, easily accessible, and cost-effective for both dental laboratories and clinics.[13,20]

Clinically effective denture repair is highly dependent on the bond strength between repair materials and the denture base, as well as appropriate surface modifications.[21] Common repair materials include autopolymerizing, visible light-cured (VLC), and heat-cured acrylic resins.[22-24] Autopolymerizing and VLC

Corresponding Author: Gökçen Dinçer, gokcenates@istanbul.edu.tr

resins are particularly favored in clinical practice due to their simplicity and minimal equipment requirements.[25-27] Testing VLC resin based on urethane dimethacrylate (UDMA) within standardized protocols may offer additional clinical benefits.[26,28]

Chemical and/or mechanical surface treatments are employed during denture repair to enhance surface characteristics and bond strength.[29,30] Mechanical treatments, such as sandblasting or abrasion, improve micromechanical retention by increasing surface area.[21] Chemical approaches include acid etching, methyl methacrylate (MMA) application, or organic solvents.[31,32] Limited studies have evaluated the fracture strength of CAD/CAM milled and conventionally fabricated denture bases repaired with VLC.[33,34] Although previous studies have indicated that VLC resins may present inadequate flexural strength,[24,33] their clinical advantages warrant further investigation. The incorporation of bonding agents as a chemical surface treatment may improve their bonding potential and mechanical performance in denture repair.

This study aims to evaluate the flexural strength (FS) of CAD/CAM milled and conventionally fabricated denture base materials repaired with UDMA-based VLC resin, with and without the application of chemical bonding agents. The null hypothesis of the present study is that the use of UDMA-based VLC repair material will not differ in FS at fracture between CAD/CAM milled and conventionally fabricated denture base materials when chemically treated with bonding agents.

METHODS

Ethics

This study is entirely in vitro and does not involve human or animal participants. Therefore, ethics committee approval is not required for this research. All procedures were carried out in accordance with the ethical rules and the principles of the Declaration of Helsinki.

Preparation of Test Samples

For the fabrication of denture base samples, a commercially available, pre-polymerized PMMA-based puck specifically produced for CAD/CAM techniques (Merz Dental GmbH, Lütjenburg, Germany) with standard dimensions (98 mm in diameter and 25 mm thickness) (group A) as experimental group and PMMA denture base material (Paladent 20, Heraeus Kulzer GmbH & Co. KG, Hanau, Germany) as control group (group B) were included. A VLC cured UDMA bases material (Eclipse Prosthetic Resin, Dentsply Int., New York, NY, USA) with a paste consistency were used as repair material.

The pre-polymerized PMMA-based CAD/CAM blocks were milled into standardized specimens measuring 65×10×2.5 mm using a universal lathe device (Trens SN50C/1000, Slovakia). Initially, cylindrical blocks were trimmed into rectangular forms, followed by horizontal and vertical cutting using 2 mm cutting burs.

For the control group (group B) acrylic resin was prepared at a powder/liquid ratio of 23.4 g/10 ml, according to the manufacturer's recommendations. The mixture homogenized at room temperature (23±2°C) for 60 seconds and then allowed to rest for 15 minutes. Afterward, the mixture was poured into plaster molds. For the polymerization process, the metal flasks were first placed in a thermostatically controlled water bath (Kavo Elektrotechnisches Werk GmbH, Biberach, Germany) at room temperature and then heated to 74°C according to the manufacturer's instructions. After keeping at 74°C for 30 min, the temperature was raised to 100°C and held for an additional 30 minutes. The flasks were then allowed to cool to room temperature in the water bath. Once the flasking process was completed, the acrylic samples were removed, and excess material was cleaned using a hand tool and a tungsten carbide bur.

Preparation of Repair Surfaces

To simulate the clinical repair process, intact samples were initially placed into plaster molds. Each sample and its corresponding mold were numbered and recorded, after which the samples were removed from the molds.

To simulate a denture fracture, the samples were divided into two equal parts using a tungsten carbide bur (Rapidy Microbur, Bredent GmbH, Senden, Germany) at a speed of 2,000 rpm. To set the repair gap at 2 mm with a 45° angle, guide marks were drawn on the sample surfaces, with a distance of 2 mm from the top and 7 mm from the bottom. All repair surfaces of the samples were milled with a tungsten carbide bur (Frank Dental, Gmund am Tegernsee, Germany) at a speed of 1,000 rpm and then smoothed under running tap water using two different grades of sandpaper (200 and 400 grit, Waterproof silicon carbide paper, English Abrasives Ltd., London, UK). The final dimensions of the samples were checked using a digital caliper (Absolute Digimatic Caliper, Mitutoyo, Kawasaki, Japan). Once the samples were adjusted to the desired dimensions, each pair of samples was placed into the corresponding mold cavities.

For VLC resin repair, samples in molds were preheated (55°C, 2 min) in an oven (Eclipse Conditioning Oven, Dentsply Sirona Int., Ontario, Canada) to facilitate resin adaptation.

Surface Treatments

After the heating, the plaster molds were removed and samples were divided into five subgroups for surface treatments (**Table 1, 2**):

Table 1. Information on the trade name, manufacturer, abbreviation and polymerization type of the denture base materials used in this study

	Manufacturer	Group	Polymerization type
CAD/CAM M-pm disc	Merz Dental GmbH, Lütjenburg, Germany	A	prepolymerized puck
Paladent 20	Heraeus Kulzer GmbH, Hanau, Germany	B	Heat-activated polymerization powder and liquid

CAD/CAM: Computer-aided design and manufacturing

Table 2. The groups in the study and the surface treatments applied

	Group A	Group B
Control group	A1	B1
G-Premio BOND	A2	B2
GC acrylic primer + G-Premio BOND	A3	B3
GC acrylic primer + G-Premio BOND DCA	A4	B4
G-Premio BOND DCA	A5	B5

Group 1 (control group-group A1-B1): No repair or surface treatment applied.

Group 2 (G-Premio BOND-group A2-B2): G-Premio BOND (GC, Tokyo, Japan) was applied to the repair surfaces with a clean, dry brush as shown in Figure 1.

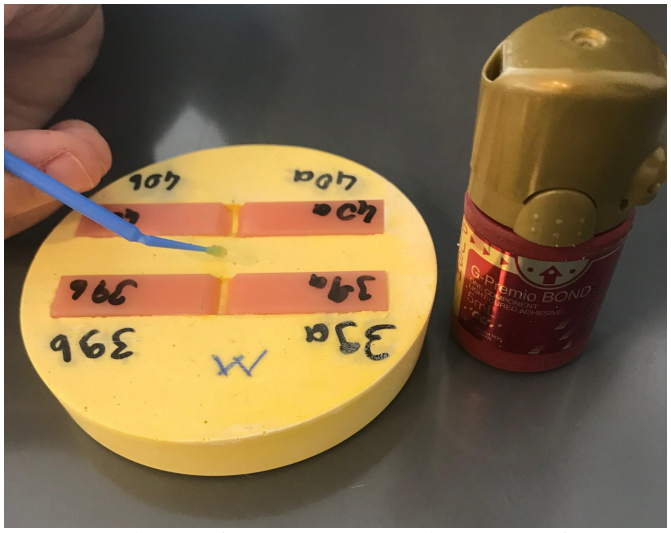

Figure 1. Application of G-Premio BOND to the repair interface using a clean microbrush

Group 3 (acrylic primer+G-Premio BOND-group A3-B3): GC acrylic primer (GC, Tokyo, Japan) was applied first, followed by air drying for 30 seconds, then G-Premio BOND was applied with a new brush as shown in Figure 1, 2.

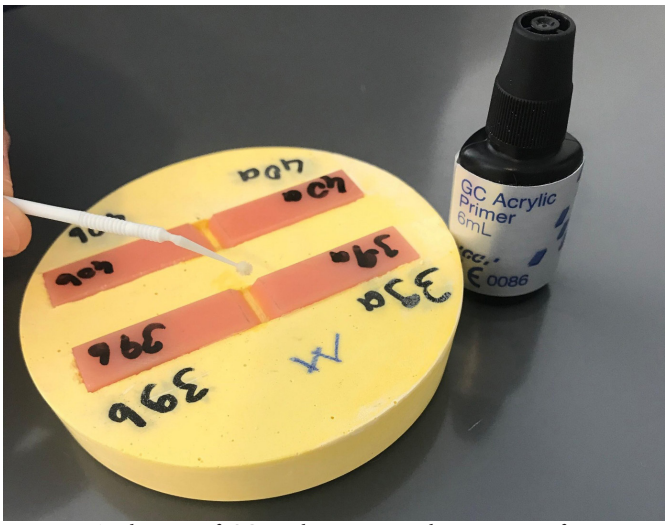

Figure 2. Application of GC acrylic primer to the repair interface using a clean microbrush

GC: Gradia composite

Group 4 (acrylic primer+G-Premio BOND DCA-group A4-B4): GC acrylic primer was applied and air-dried, then G-Premio BOND DCA (GC, Tokyo, Japan) was applied using a separate brush as shown in Figure 2, 3.

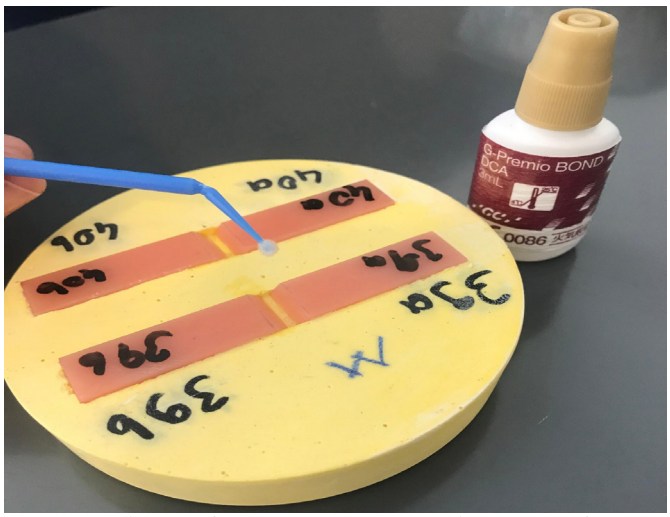

Figure 3. Application of G-Premio BOND DCA to the repair interface using a clean microbrush

Group 5 (G-Premio BOND DCA group A5-B5): G-Premio BOND DCA was applied directly with a clean brush as shown in Figure 3.

Separate clean brushes were used for each bonding agent to prevent cross-contamination. In all treatment groups, bonding agents were uniformly spread into a thin layer using air spray and polymerized (20s) with a light-curing device (Smartlite Max, Model 644050, Dentsply, USA, intensity: 1000 mW/cm^2) as shown in Figure 4.

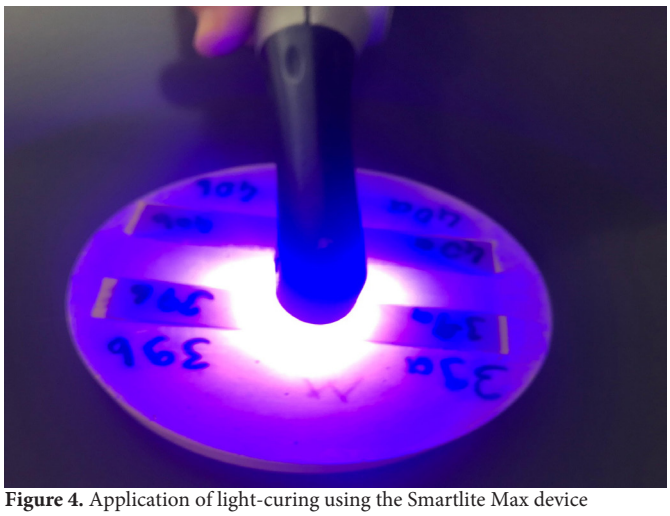

Figure 4. Application of light-curing using the Smartlite Max device

Subsequently, 2 mm of the VLC repair material in a paste form was applied and condensed firmly into the repair gap using finger pressure. To prevent oxygen inhibition during polymerization, an air barrier coating (Eclipse Air Barrier Coating, Dentsply Sirona Inc, New York, USA) was applied over the repair resin with a brush. Polymerization was performed (10 min) using the Eclipse Junior light-curing unit (Dentsply Sirona, Germany), according to the manufacturer's instructions.[34]

Samples were cooled to room temperature (23±2°C), rinsed with distilled water to remove the coating, and carefully removed from the plaster molds using a fine-tipped spatula. All samples repaired were finished using a hard bur (Frank Dental GmbH, Gmund, Germany) at 1000 rpm and then smoothed under running water using 200 and 400 grit sandpapers (Waterproof silicon carbide paper, English Abrasives Ltd., London, United Kingdom). The final sample dimensions were verified at three separate points using a digital caliper (±0.01 mm precision), and the average of these measurements was used to confirm compliance with dimensional criteria. Material specifications, including the chemical composition and curing recommendations of the UDMA-based VLC resin, were obtained from the manufacturer's technical datasheet (Dentsply Sirona, Eclipse Prosthetic Resin, MSDS).[35]

The samples prepared for the 3-point bending test were grouped according to their respective group, and all samples were stored in distilled water at 37°C for 48 hours after the complete repair process, prior to mechanical testing.

The sample size was determined based on a previous study that conducted power analysis using the G*power software (version 3.1.9.7) with an effect size (d=0.861) and standard deviation of 6.[33] According to that analysis, a minimum of five specimens per group was sufficient to detect statistically significant differences at a power of 80% and an alpha level of 0.05. In the present study, 10 specimens were included in each subgroup. A total of 100 specimens were tested in this study.

Flexural Strength Test

The flexural strength was assessed using a 3-point bending test on a Universal Testing Machine (EZ Test Series, Shimadzu, Japan). The span length between the metal supports was set at 50 mm, and the crosshead speed was maintained at 5 mm/min. A compressive force was applied perpendicularly to the midpoint of each specimen until fracture occurred. The maximum load (N), deflection at fracture (mm), and corresponding flexural data were automatically recorded via the connected software. Flexural strength (FS) in megapascals (MPa) was calculated using the following formula, as previously described:[33] $FS=3FL/(2bd^2)$ where F is the maximum load at fracture (N), L is the support span (mm), b is the specimen width (mm), and d is the specimen thickness (mm).

Statistical Analysis

Data analyses were performed using the SPSS 22 software (IBM Corp., Armonk, NY, USA). Descriptive statistics were reported mean±standard deviation (SD). A two-way analysis of variance (ANOVA) was conducted to investigate differences in flexural strength (FS) among different acrylic materials and surface treatment groups. The model included main effects for material (CAD/CAM milled vs. conventional), surface treatment (group 1 to group 5), and the interaction term (material×surface treatment). When significant interaction effects were observed, post-hoc comparisons were performed using simple effects analysis with Bonferroni correction. A p-value <0.05 was considered statistically significant for all analyses.

RESULTS

The statistical evaluation of the flexural strength values of the experimental groups was performed using arithmetic mean values and two-way ANOVA, and the results are presented in Table 3, 4, respectively. There was a statistically significant difference after the surface treatments, except for group A5-B5 (p<0.001) (Table 4).

Table 3. Description statistics of flexural strength

	Group A			Group B		
Group	n	Mean±SD	(min-max)	n	Mean±SD	(min-max)
1	10	81.73±8.21	(68.28-91.88)	10	75.75±3.86	(70-80.94)
2	10	11.13±2.03	(8.13-14.06)	10	5.47±2.43	(1.56-10.47)
3	10	17.31±4.69	(10.94-25.78)	10	9.8±2.57	(7.19-15)
4	10	11.56±2.95	(5.94-15.78)	10	3.25±1.73	(1.09-5.63)
5	10	1.11±0.27	(0.78-1.72)	10	3.39±1.50	(1.41-6.09)

Values are shown in MPa

Table 4. Comparison of fracture forces by groups and materials

	Material				p		
	Group A		Group B				
Group	n	Mean±SD	n	Mean±SD	Material	Group	Material* group
1	10	$81.73±8.21^{a,A}$	10	$75.75±3.86^{a,B}$			
2	10	$11.13±2.03^{c,A}$	10	$5.47±2.43^{bc,B}$			
3	10	$17.31±4.69^{b,A}$	10	$9.8±2.57^{b,B}$	<0.001	<0.001	<0.001
4	10	$11.56±2.95^{c,A}$	10	$3.25±1.73^{c,B}$			
5	10	$1.11±0.27^{d,A}$	10	$3.39±1.50^{c,A}$			

A, B: Different uppercase letters in the same row indicate statistically significant differences between materials within each surface treatment (p<0.05). a, b, c, d: Different lowercase letters in the same column indicate statistically significant differences between surface treatments within each material group (p<0.05). (Post-hoc comparisons performed with Bonferroni correction)

In group A, the highest fracture strength was observed in the intact samples (group A1: 81.73±8.21 MPa, SD), followed by groups where the bonding agent was applied with the acrylic primer (group A3: 17.31±4.69 MPa, SD); group A4: 11.56±2.95 MPa,SD). Moderate FS was recorded in the G prime bonding group (group A2: 11.13±2.03 MPa, SD). In group B, the highest FS was also observed in the intact specimens (group B1: 75.75±3.86 MPa, SD), a statistically significant difference was noted when the acrylic primer was applied in combination with the G-Premio BOND agent [group B3: 9.8±0.81 MPa (SD)], followed by the G-Premio BOND group (group B2: 2.57±2.43 MPa, SD), and the acrylic primer with DCA Bond (group B4: 3.25±1.73 MPa, SD). The lowest fracture strength in both groups was found in the group where only the dual cure activator agent was applied [group A5: 1.11±0.27 MPa (SD); group B5: 3.39±1.50 MPa (SD)].

The pairwise comparisons were performed using Bonferroni post-hoc test, and significant differences were denoted with superscript letters in Table 4. A statistically significant interaction was observed between material and surface treatment group (p<0.001).

DISCUSSION

The null hypothesis of the present study was rejected, as significant differences were found between the FS of CAD/CAM milled and conventionally produced denture base materials after surface modifications with different chemical agents, except for one group when using VLC polymerized UDMA repair material.

Denture base materials produced by CAD/CAM systems exhibit a range of FS values. Previous studies have demonstrated that these materials provide significantly greater FS compared to conventionally heat-polymerized denture base resins.[32,33] The results obtained from the present study, which show higher FS in intact CAD/CAM specimens than in conventionally polymerized ones, are in agreement with earlier reports.[24,33] This enhancement can be attributed to the optimized material properties, the employment of pre-polymerized blocks fabricated under elevated pressure,[27,28] reduced polymerization shrinkage,[4] and a minimal residual monomer presence.[35]

To date, there is limited in vivo evidence concerning the mechanical failure of digitally fabricated complete dentures. While intraoral fractures in CAD/CAM dentures are infrequent, extraoral fractures remain a possibility.[37] With the expanding clinical use of CAD/CAM denture base materials, further in vivo investigations on intraoral and extraoral failure modes are needed to comprehensively assess their long-term mechanical performance.

The primary objective in denture repair is to re-establish the mechanical strength and ensure adequate bonding between the base and repair material. Surface geometry plays a vital role in this process. Literature supports that a 45° beveled joint with rounded edges increases bonding area and modifies stress distribution from tensile to shear forces, which enhances repair durability.[28,37] In the present study, the selection of a 45° angled repair surface design was made to promote effective preparation and to assure an improved distribution of adverse stresses.

The gap between the fractured surfaces is another crucial parameter. Research suggests ideal repair gaps ranging from 1.5 to 3 mm, although gaps as large as 10 mm have been tested.[37] In our study, a 2 mm repair gap was chosen due to ease of application and aesthetic concerns, as narrower gaps can create application difficulties due to the thickness of the bur, and variations in the repair gap could affect the results. For future studies, a new study design could be proposed using different repair gaps.

Despite their mechanical inferiority to autopolymerizing and heat-polymerizing resins, VLC resins are still used in clinical applications. Their advantages include reduced residual monomer content and superior color stability.[21,29,31] However, previous studies such as Lewinstein et al.,[38] which reported no significant differences in bond strength between these materials, were conducted under different surface preparation and polymerization conditions, limiting direct comparisons. Additionally, VLCs are commonly hand-mixed and applied without pressure, increasing the risk of internal voids and defects.[32] Consequently, their mechanical performance may be compromised. Nonetheless, they may be suitable in specific clinical situations prioritizing esthetics and reduced irritation over mechanical strength. In a study examining the repair process of milled denture base materials using VLC repair resin,[28] the authors suggested that, in addition to investigating surface treatments for milled PMMA, there is a need to develop a bonding agent when VLC material is preferred. In this study, different commercially available bonding agents were used as repair resins for milled and conventionally produced PMMA, and their effect on flexural strenght was investigated.

In this study, several commercially available bonding agents were tested. In the CAD/CAM group, the highest FS (17.31±4.69 MPa) was achieved when bonding agent was used together with acrylic primer (group A3). This indicates that primer application enhances surface energy and facilitates bonding agent adhesion. In contrast, the bonding agent alone (group A5: 1.11 ± 1.09 MPa) resulted in lower FS. Therefore, the combined use of primer and bonding agent is advised for repairs of CAD/CAM milled PMMA bases. These findings support the notion that the bonding agent alone may be insufficient due to inadequate interaction with the CAD/CAM substrate, which has low surface energy and high crosslinking density. The primer enhances wettability and promotes better diffusion and micromechanical interlocking.

Similarly, in conventional specimens, the combination of bonding agent and primer (group B3: 9.8±2.57 MPa) yielded superior FS. The lowest FS was observed in the group treated solely with a dual-cure activator (group B5: 3.39±1.50 MPa). These results confirm that using bonding agents alone may be insufficient, highlighting the importance of surface pre-treatment.

Clinically, although CAD/CAM denture repairs may incur higher costs, combining primers with bonding agents can enhance repair strength and patient satisfaction. Appropriate selection of materials and protocols can improve prosthesis longevity and treatment outcomes.

Nevertheless, the claim that reproduction is superior to repair should be made with caution. While re-fabricating a denture using stored digital data can provide excellent mechanical results, it is not always feasible due to clinical, economic, or logistic constraints. In many cases, repair remains a valid and timely solution.

Limitations

Limitations of this study include its in vitro setting, the use of only one VLC resin, and the absence of long-term clinical or aging simulations. Although power analysis was performed and the sample size was above the minimum threshold, future research should involve larger and more diverse samples to enhance generalizability.

CONCLUSION

This study has demonstrated the impact of different bonding agents and surface treatments on the repair of PMMA denture base materials, showing that the combination of acrylic primer and bonding agents provides the highest flexural strenght. These findings highlight the importance of selecting

appropriate repair materials and surface treatments in clinical practice and offer guidance for achieving more durable and long-lasting denture repairs.

ETHICAL DECLARATIONS

Ethics Committee Approval

This study is entirely in vitro and does not involve human or animal participants. Therefore, ethics committee approval is not required for this research.

Informed Consent

Since the study was conducted without the participation of any living being, no written consent form was obtained.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

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

Author Contributions

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

REFERENCES

1. Carlsson GE, Omar R. The future of complete dentures in oral rehabilitation. A critical review. *J Oral Rehabil.* 2010;37(2):143-156. doi:10.1111/j.1365-2842.2009.02039.x
2. Yilmaz B, Azak AN, Alp G, Ekşi H. Use of CAD-CAM technology for the fabrication of complete dentures: an alternative technique. *J Prosthet Dent.* 2017;118(2):140-143. doi:10.1016/j.prosdent.2016.10.016
3. Kalberer N, Mehl A, Schimmel M, Müller F, Srinivasan M. CAD-CAM milled versus rapidly prototyped (3D-printed) complete dentures: an in vitro evaluation of trueness. *J Prosthet Dent.* 2019;121(4):637-643. doi:10.1016/j.prosdent.2018.08.002
4. Goodacre BJ, Goodacre CJ, Baba NZ, Kattadiyil MT. Comparison of denture base adaptation between CAD-CAM and conventional fabrication techniques. *J Prosthet Dent.* 2016;116(2):249-256. doi:10.1016/j.prosdent.2016.02.017
5. Fouda A, Tonogai J, McDermott P, Wang D, Dong CS. A systematic review on patient perceptions and clinician-reported outcomes when comparing digital and analog workflows for complete dentures. *J Prosthodont.* 2024. doi:10.1111/jopr.13723
6. Ohara K, Isshiki Y, Hoshi N, et al. Patient satisfaction with conventional dentures vs. digital dentures fabricated using 3D-printing: a randomized crossover trial. *J Prosthodont Res.* 2022;66(4):623-629. doi:10.2186/jpr.JPR_D_21_00149
7. Arslan M, Murat S, Alp G, Zaimoglu A. Evaluation of flexural strength and surface properties of prepolymerized CAD/CAM PMMA-based polymers used for digital 3D complete dentures. *Int J Comput Dent.* 2018;21(1):31-40.
8. Al-Fouzan AF, Al-Mejrad LA, Albarrag AM. Adherence of *Candida* to complete denture surfaces in vitro: a comparison of conventional and CAD/CAM complete dentures. *J Adv Prosthodont.* 2017;9(5):402-408. doi:10.4047/jap.2017.9.5.402
9. Infante L, Yilmaz B, McGlumphy E, Finger I. Fabricating complete dentures with CAD/CAM technology. *J Prosthet Dent.* 2014;111(5):351-355. doi:10.1016/j.prosdent.2013.10.014
10. Steinmassl O, Dumfahrt H, Grunert I, Steinmassl PA. CAD/CAM produces dentures with improved fit. *Clin Oral Investig.* 2018;22(8):2829-2835. doi:10.1007/s00784-018-2369-2
11. Dimitrova M, Vlahova A, Hristov I, Kazakova R. Bonding efficiency between artificial teeth and denture base in CAD/CAM and conventional complete removable dentures. *Materials (Basel).* 2024;17(13):3138. doi:10.3390/ma17133138
12. Wiegand A, Stucki L, Hoffmann R, Attin T, Stawarczyk B. Repairability of CAD/CAM high-density PMMA-and composite-based polymers. *Clin Oral Investig.* 2015;19(8):2007-2013. doi:10.1007/s00784-015-1424-8
13. Jeong KW, Kim SH. Influence of surface treatments and repair materials on the shear bond strength of CAD/CAM provisional restorations. *J Adv Prosthodont.* 2019;11(2):95-104. doi:10.4047/jap.2019.11.2.95
14. Kumari R, Bala S. Assessment of cases of complete denture fracture. *J Pharm Bioallied Sci.* 2021;13(Suppl 2):S1558-S1560. doi:10.4103/jpbs.JPBS_726_20
15. Bosanceanu DN, Beldiman A, Baciu RE, Bolat M, Bosanceanu DG, Forna NC. Complete dentures fractures-causes and incidence. *Ro J Oral Rehabil.* 2017;9(1):54-59.
16. Ates M, Cilingir A, Sulun T, Sunbuloglu E, Bozdag E. The effect of occlusal contact localization on the stress distribution in complete maxillary denture. *J Oral Rehabil.* 2006;33(7):509-513. doi:10.1111/j.1365-2842.2005.01578.x
17. Ali IL, Yunus N, Abu-Hassan MI. Hardness, flexural strength, and flexural modulus comparisons of three differently cured denture base systems. *J Prosthodont.* 2008;17(7):545-549. doi:10.1111/j.1532-849X.2008.00357.x
18. Huggett R, Bates JF, Packham DE. The effect of the curing cycle upon the molecular weight and properties of denture base materials. *Dent Mater.* 1987;3(2):107-112. doi:10.1016/s0109-5641(87)80023-9
19. Murakami N, Wakabayashi N, Matsushima R, Kishida A, Igarashi Y. Effect of high-pressure polymerization on mechanical properties of PMMA denture base resin. *J Mech Behav Biomed Mater.* 2013;20:98-104. doi:10.1016/j.jmbbm.2012.12.011
20. Venkat R, Gopichander N, Vasantakumar M. Comprehensive analysis of repair/reinforcement materials for polymethyl methacrylate denture bases: mechanical and dimensional stability characteristics. *J Indian Prosthodont Soc.* 2013;13(4):439-449. doi:10.1007/s13191-013-0303-4
21. Sahin Z, Ozer NE, Akan T, Kılıcarslan MA, Karaagaclıoglu L. The effect of various surface treatments on the repair bond strength of denture bases produced by digital and conventional methods. *Odontology.* 2024;112(3):782-797. doi:10.1007/s10266-023-00808-1
22. Tuğut F, Koyu T. Effect of repair and thermal cycling on the flexural strength of denture base materials fabricated from different methods. *Cumhuriyet Dent J.* 2023;26(2):150-156. doi:10.7126/cumudj.1201130
23. Stipho HD, Talic YF. Repair of denture base resins with visible light-polymerized reline material: effect on tensile and shear bond strengths. *J Prosthet Dent.* 2001;86(2):143-148. doi:10.1067/mpr.2001.117055
24. AlQahtani M, Haralur SB. Influence of different repair acrylic resin and thermocycling on the flexural strength of denture base resin. *Medicina (Kaunas).* 2020;56(2):50. doi:10.3390/medicina56020050
25. Bural C, Aktaş E, Deniz G, Ünlüçerçi Y, Kizilcan N, Bayraktar G. Effect of postpolymerization heat-treatments on degree of conversion, leaching residual MMA and in vitro cytotoxicity of autopolymerizing acrylic repair resin. *Dent Mater.* 2011;27(11):1135-1143. doi:10.1016/j.dental.2011.08.007
26. Cilingir A, Bilhan H, Geckili O, Sulun T, Bozdag E, Sunbuloglu E. In vitro comparison of two different materials for the repair of urethane dimethacrylate denture bases. *J Adv Prosthodont.* 2013;5(4):396-401. doi:10.4047/jap.2013.5.4.396
27. Mahajan H, Chandu GS, Mishra SK. An in vitro study of the effect of design of repair surface on the transverse strength of repaired acrylic resin using autopolymerizing resin. *Niger J Clin Pract.* 2014;17(1):38-42. doi:10.4103/1119-3077.122835
28. Vasthare A, Shetty S, Kamalakanth SK, Shetty M, Parveen K, Shetty R. Effect of different edge profile, surface treatment, and glass fiber reinforcement on the transverse strength of denture base resin repaired with autopolymerizing acrylic resin: an in vitro study. *J Interdiscip Dent.* 2017;7(1):31-37. doi:10.4103/jid.jid_76_16
29. Al-Dwairi ZN, Tahboub KY, Baba NZ, Goodacre CJ, Ozcan M. A comparison of the surface properties of CAD/CAM and conventional polymethylmethacrylate (PMMA). *J Prosthodont.* 2019;28(4):452-457. doi:10.1111/jopr.13035

30. Minami H, Suzuki S, Minesaki Y, Kurashige H, Tanaka T. In vitro evaluation of the influence of repairing condition of denture base resin on the bonding of autopolymerizing resins. *J Prosthet Dent.* 2004;91(2):164-170. doi:10.1016/j.prosdent.2003.11.016
31. Özcan M, Alander P, Vallittu PK, Huysmans MC, Kalk W. Effect of three surface conditioning methods to improve bond strength of particulate filler resin composites. *J Mater Sci Mater Med.* 2005;16(1):21-27. doi:10.1007/s10856-005-6442-4
32. Erbulak Z, Ergun G. The effects of different surface treatments applied to milled PMMA denture base material on repair bond strength. *Odontology.* 2023;111(4):953-970. doi:10.1007/s10266-023-00794-4
33. Özatik Ş, Bural Alan C. Flexural strength of repaired denture base materials manufactured for the CAD-CAM technique. *J Oral Sci.* 2024;66(2):120-124. doi:10.2334/josnusd.23-0220
34. Dentsply Sirona Prosthetics. Eclipse® Prosthetic Resin Materials (Baseplate, Set-Up and Contour Resin) Safety Data Sheet. Rev. 5. York, PA: Dentsply Sirona; 2019. Available from: https://www.dentsplysirona.com
35. Menon RK, Xin YH, Wei BCT, et al. CADCAM versus conventional denture bases: network meta-analysis of in vitro studies comparing accuracy and surface properties. *Int Dent J.* 2025;75(3):2062-2070. doi:10.1016/j.identj.2024.12.032
36. Ölçer Us O, Yüzbaşıoğlu E, Özdemir G, Albayrak B. Clinical outcomes and complications of CAD-CAM fabricated complete dentures: an update and review. *J Exp Clin Med.* 2021;38(S2):92-97. doi:10.52142/omujecm.38.si.dent.3
37. Beyli MS, von Fraunhofer JA. Repair of fractured acrylic resin. *J Prosthet Dent.* 1980;44(5):497-503. doi:10.1016/0022-3913(80)90067-0
38. Lewinstein I, Zeltser C, Mayer CM, Tal Y. Transverse bond strength of repaired acrylic resin strips and temperature rise of dentures relined with VLC reline resin. *J Prosthet Dent.* 1995;74(4):392-399. doi:10.1016/s0022-3913(05)80377-3

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1709742

J Health Sci Med. 2025;8(5):794-799

Does first trimester prenatal screening impact on maternal anxiety? A prospective cohort study

Ömer Gökhan Eyisoy[1], Kemal Sandal[2]

[1]Department of Obstetrics, Division of Perinatology, Yozgat City Hospital, Yozgat, Turkiye

[2]Department of Obstetrics and Gynecology, Göztepe Prof. Dr. Süleyman Yalçın City Hospital, İstanbul, Turkiye

Cite this article as: Eyisoy ÖG, Sandal K. Does first trimester prenatal screening impact on maternal anxiety? A prospective cohort study. *J Health Sci Med.* 2025;8(5):794-799.

Received: 03.06.2025 ◆ **Accepted**: 29.07.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: The objective of the present study was to evaluate the anxiety levels of pregnant women who were admitted for combined first trimester screening test, and to determine the effect of screening results on maternal anxiety.

Methods: The present study was carried out prospectively between March and June 2025, and comprised 148 pregnant women. The Spielberger State-Trait Anxiety Inventory was utilised in order to evaluate the participants' state (STAI-I) and trait (STAI-II) anxiety levels. The STAI-Ia and STAI-II were administered to participants at their first admission, and first-trimester screening was performed. Following the counselling of first trimester screening results the State Anxiety Index (STAI-Ib) was re-administered to the participants. Changes in participants' anxiety levels were evaluated.

Results: The mean pre-screening state anxiety scores (STAI-Ia) of the patients were higher than their trait anxiety scores (STAI-II) (39.80±6.08 and 36.68±4.70 respectively; p<0.001). A total of 12 patients (13.9%) exhibited positive results in their first trimester screening tests. Patients with positive test results had higher state anxiety scores after first-trimester screening and counselling (STAI-Ib), compared to pre-screening scores (STAI-Ia) (40.00±7.26 vs. 43.25±7.38, p=0.003). 75% of the patients with positive test results demonstrated a probable clinical anxiety score. In patients with negative test result, state anxiety scores reduced compared to pre-sceening scores (39.79±6.00 vs. 36.93±5.27, p<0.001).

Conclusion: Admission for first trimester screening has been shown to induce state anxiety in the patients. Social and psychological support should be provided to prevent the negative effects of maternal anxiety, especially in patients with positive screening results.

Keywords: Maternal anxiety, first trimester screening, down syndrome, fetal anomaly

INTRODUCTION

Pregnancy is a physiological period of life that brings about many biological, hormonal and psychological changes in the pregnant woman. In the course of adapting to such changes, there is an elevated risk of emotional and psychological disorders, including anxiety, depression, stress and sadness. Recent research has revealed that emotional disorders during pregnancy are more prevalent than previously estimated. Socioeconomic status, unplanned pregnancy, advanced maternal age, and high-risk pregnancy have been identified as factors that can influence the psychological well-being of pregnant women.[1]

The possibility of a structural or genetic malformation in the fetus is a cause of concern for the mother, and is often associated with maternal anxiety.[2] The advent of medical technology has made prenatal screening and diagnosis of a wide range of structural and genetic abnormalities possible. Down syndrome (trisomy 21) is the most common chromosomal abnormality compatible with life and can be detected prenatally.[3] A combined nuchal translucency (NT) measurement and biochemical screening test for Down syndrome performed between 11-14 weeks of gestation has been shown to detect up to 87% of fetuses with Down syndrome and is also recommended by national pregnancy follow-up programmes.[4,5] However, screening for fetal anomalies can cause unnecessary concern or anxiety in women because it raises the possibility of a problem with the fetus.[6] It has been reported that maternal anxiety during the prenatal period is affected by obstetric complications, and that children of pregnant women with anxiety are at increased risk of adverse outcomes in terms of socioemotional development.[7,8]

The objective of the present study was to investigate the impact of combined first trimester screening test on maternal anxiety.

METHODS

This prospective cohort study was conducted at Yozgat City Hospital between March and June 2025. Ethical approval for the study was obtained from the Yozgat Bozok University Rectorate Non-interventional Clinical

Corresponding Author: Ömer Gökhan Eyisoy, dr.gokhaneyisoy@gmail.com

Researches Ethics Committee (Date: 05.03.2025, Decision No: 2025-GOKAEK-255_05.03.2025-370). The study was conducted in accordance with the tenets of the Declaration of Helsinki. Prior to participation, all subjects were thoroughly informed about the study and its associated procedures, and subsequently provided written informed consent.

During the course of the study, a total of 224 patients underwent first-trimester Down syndrome screening and basic first-trimester fetal anatomy scan. All examinations were performed between 11 and 14 weeks of gestation by a perinatologist or obstetrician qualified in nuchal translucency measurement. Examinations were performed according to the international guidelines.[9] Patients over 35 years of age, patients referred with suspicion of fetal anomaly, patients who could not speak Turkish, multiple pregnancies, patients with active psychiatric disease or history of psychiatric disease, patients with detected fetal anomaly or non-living fetus during examination and patients who refused to participate in the study were excluded from the study. A total of 148 patients who met the study's inclusion criteria were included in the study.

A standardised scale, the Spielberger State-Trait Anxiety Inventory (STAI), was employed in order to evaluate the anxiety levels of the patients.[10] The version of the scale that had been adapted for use in Turkey by Öner in 1985 was utilized.[11] The scale comprises a total of 40 questions; 20 questions are aimed at assessing state anxiety (STAI-I), which indicates how one feels at a particular moment, and 20 questions are aimed at assessing trait anxiety (STAI-II), which indicates how one generally feels. Participants are invited to respond to each question using a 4-point Likert Scale, allocating 4 points to the situation that best reflects them and 1 point to the situation that least reflects them. The STAI is comprised of two distinct categories of statements. Direct statements are employed to express negative emotions (e.g. I am tense, I feel strained), whilst inverted statements are used to express positive emotions (e.g. I feel calm, I feel secure). Firstly, the total weights of the direct and inverted statements are calculated separately, then the sum of the weights score of the inverted statements is subtracted from the sum of the weights score of the direct statements. A constant and predetermined value is added to this number. For the state anxiety scale, this value is 50; for the trait anxiety scale, it is 35. The final value represents the individual's anxiety score. Scores on this scale range from 20, indicating minimal anxiety, to 80, reflecting severe anxiety. A score greater than 40 is regarded as an indicator of possible clinical anxiety levels in a pregnant women.[12]

The flowchart illustrating the study's methodology is presented in **Figure 1**. The clinical and sociodemographic characteristics of the patients admitted for combined first trimester screening test were obtained upon initial presentation. Patients who met the study's inclusion criteria were thoroughly informed about the study and provided with written consent. The STAI-I and STAI-II scales were administered to the participants by the study assistant prior to the combined first trimester screening test. In order to ascertain the participants' pre-screening anxiety levels, the STAI-I scores were utilised (STAI-Ia). Furthermore, in order to determine the baseline anxiety levels of the participants, the STAI-II scores were utilised. A standardised first trimester ultrasonographic examination, incorporating NT measurement, was subsequently conducted. To complete combined first trimester screening test, blood samples were obtained from patients exhibiting ultrasonographically normal fetal anatomy. Once the screening test had been completed within one week, patients with a risk value above the 1/270 threshold for Down syndrome were grouped as positive test result, while patients with a risk value below the threshold were grouped as negative test result. The combined first trimester screening test was not performed in patients with abnormal findings on sonographic examination and the medical counselling was provided to the patient. All patients were counselled on their screening test results. The counselling interview encompassed a comprehensive information of the principles, reliability, accuracy, and the potential for false positive and negative results associated with combined first trimester screening test. The interview also addressed cell-free fetal DNA testing and invasive diagnostic procedures such as chorion villus sampling and amniocentesis, and the associated risks. During counselling, all patients with a positive test result were offered invasive diagnostic testing. Following the counselling about the results, the state anxiety test was re-administered (post-screening; STAI-Ib). The STAI-II was not re-administered to the participants, as it is considered to represent a more stable state of anxiety, and the change in trait anxiety scores would not be significant. Individuals manifesting clinical anxiety were referred to the clinical psychologists.

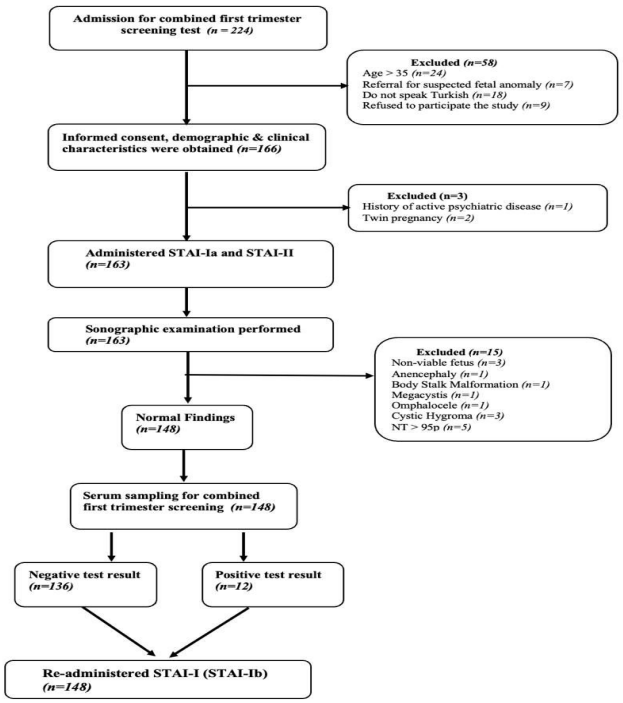

Figure 1. Follow-up chart of the patients

STAI: State-Trait Anxiety Inventory

An a priori power analysis was performed using G*power version 3.1.9.7 to estimate the required sample size. In the study conducted by Akalın et al.[13] on fetal congenital heart

disease during pregnancy, state anxiety scores were found to be 46.28±7.18 before fetal echocardiography and 43.48±7.97 after fetal echocardiography. The effect size, as determined by this study, was calculated to be 0.3681385. When the relevant effect size was employed (linear bivariate regression one group size of slope), it was determined that a minimum of 85 pregnant patients were required for the study to be adequate with 95% power and 5% alpha error. Consequently, the sample size of 148 participants is more than sufficient to test the hypotheses of the study.

Continuous data are presented as mean±standard deviation or median (range) according to the data distrubution, while categorical data are presented as frequency and percentages. In order to compare the clinical and sociodemographic characteristics of the groups Mann-Whitney U test, chi-square test or Fischer exact test (due to low expected cell counts) was employed. Spearman's rank correlation was used to evaluate the association between quantitative variables. A paired samples T test or Wilcoxon test (in instances where the assumptions of the paired T test were not met) was conducted to compare anxiety levels before and after combined first-trimester screening test. The Statistical Package for the Social Sciences, version 22 for Windows (IBM Corporation, Armonk, NY, USA), was utilised for all statistical analyses. Statistical significance was interpreted as $p<0.05$.

RESULTS

In total, 148 patients who met the specified inclusion criteria were enrolled in the study during its course. Figure 1 presents the flow chart outlining the selection of patients included and excluded from the study. In the present study, 12 patients were identified as positive test result due to their combined screening test results for Down syndrome exceeding the 1/270 threshold. In 136 patients, no major sonographic abnormalities were detected and the combined screening test results were below the threshold value, so the screening was considered negative.

The clinical and sociodemographic characteristics of the patients are presented in Table 1. The positive test result and negative test result groups did not differ with regard to age, gestational age, gravidity, parity, history of miscarriage, rate of consanguineous couples, rate of having a healthy child, rate of having a history of a child with anomaly, rate of maternal comorbidity, educational level, working status and monthly income level ($p>0.05$).

The mean pre-screening state anxiety scores of the patients were found to exceed their trait anxiety scores (STAI-Ia, STAI-II; 39.80±6.08 and 36.68±4.70 respectively, $p<0.001$). This finding indicates that patients referred for combined first-trimester screening test exhibit a higher state anxiety level in comparison to their trait anxiety. Spearman's correlation analysis revealed no correlation between patients' pregnancy associated plasma protein-A (PAPP-A) and human chorionic gonadotropin (hCG) levels and their pre- and post-screening state anxiety scores ($p>0.05$, n=148). A weak positive correlation was identified between the patients' trait anxiety scores (STAI-II) and hCG levels (Figure 2; $r=0.235$, $p=0.004$, n=148); however, no correlation was observed with PAPP-A levels ($p>0.05$, n=148).

Table 1. Clinical and sociodemographic characteristics of the patients

		Negative test result (n=136)	Positive test result (n=12)	p-value
Age (years)		27.5 (18-35)	27.5 (21-35)	0.757
Gestational age (weeks)		12.0 (11.0-14.0)	12.0 (11.0-14.0)	0.749
Gravida	1 ≥2	42 (30.9) 94 (69.1)	4(33.3) 8 (66.7)	0.545
Parity	No ≥1	51 (37.5) 85 (62.5)	5 (41.7) 7 (58.3)	0.501
History of miscarriage	No Yes	104 (76.5) 32 (23.5)	10 (83.3) 2 (16.7)	0.450
Consanguineous marriage	No Yes	117 (86.0) 19 (14.0)	12 (100.0) 0 (0.0)	0.179
Having healty child	No Yes	58 (42.6) 78 (57.4)	5 (41.7) 7 (58.3)	0.948
Anomalous child*	No Yes	125 (91.9) 11 (8.1)	12 (100.0) 0 (0.0)	0.381
Maternal comorbidity	No Yes	115 (84.6) 21 (15.4)	10 (83.3) 2 (16.7)	0.587
Education	Primary High school or University	52 (38.2) 84 (61.8)	4 (33.3) 8 (66.7)	0.499
Working status	Housewife Employed	76 (55.9) 60 (44.1)	9 (75.0) 3 (25.0)	0.164
Monthly income level	Low Middle or high	60 (44.1) 76 (55.9)	6 (50.0) 6 (50.0)	0.461

Data was presented as median (min-max), number and percentage (%), *History of pregnancy termination due to an anomalous fetus or having a child with an anomaly

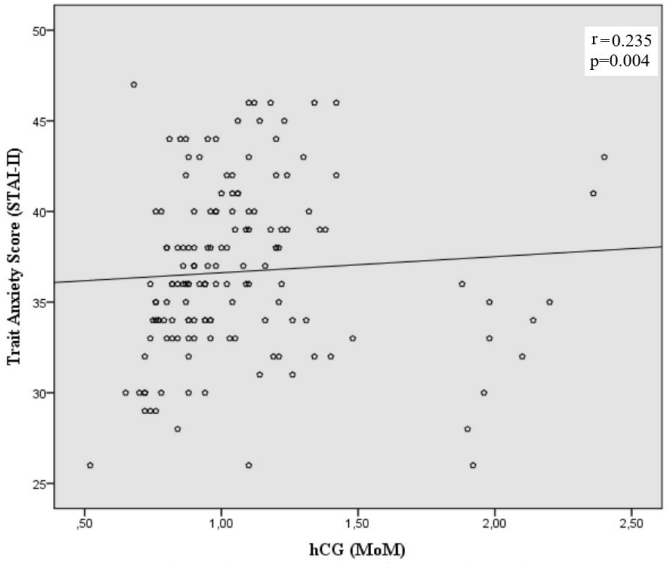

Figure 2. Scatterplot and regression lines for correlation between STAI-II scores and hCG levels (MoM) of pregnant women (n=148) (r: Correlation coefficient)

STAI: State-Trait Anxiety Inventory, hCG: Human chorionic gonadotropin

The state anxiety scores of the patients are presented in Table 2, for both pre-screening (STAI-Ia) and post-screening (STAI-IIb). In patients who has positive test results, state anxiety scores were observed to increase following first-trimester screening test and counselling when compared to the scores obtained pre-screening (40.00±7.26 vs. 43.25±7.38, p=0.003). Anxiety scores increased in all patients with positive test results. Moreover, in 9 of the 12 patients (75%), the anxiety scores exceeded the probable clinical anxiety threshold of 40 points. In patients with negative test results, state anxiety scores reduced compared to pre-sceening scores (39.79±6.00 vs. 36.93±5.27, p<0.001). Although state anxiety scores decreased in patients with a history of miscarriage (39.76±6.71 vs. 38.41±6.58, p=0.115), consanguineous marriage (39.42±6.18 vs. 38.11±6.75, p=0.122) and maternal comorbidity (37.13±5.37 vs. 36.96±4.16, p=*0.766*) after first trimester screening and counselling compared with to the pre-screening scores, these findings were not statistically significant. Patients with a history of pregnancy termination due to an anomalous fetus or having a child with an anomaly had lower state anxiety scores after the first trimester screening test and counselling (41.36±6.08 vs. 35.91±6.12, p=0.01).

DISCUSSION

The three principal findings of the study are as follows: Firstly, the pre-screening state anxiety scores of the patients were found to exceed their trait anxiety scores. Secondly, state anxiety scores were found to decrease in patients with a negative test results, and increase in patients with a positive test results. Thirdly, state anxiety scores of patients with a history of miscarriage, consanguineous marriage and maternal comorbidities remained unchanged following combined first trimester screening test.

In a large meta-analysis, it was reported that the anxiety levels of patients with early pregnancy were similar to the general population and that offering prenatal screening tests to patients had positive effects on maternal anxiety with a low - moderate level of evidence. Furthermore, no difference was identified between the anxiety scores of patients who accepted or refused to have a Down syndrome screening test.[14] All participants in the present study comprised patients who had applied for combined first-trimester screening test for Down syndrome and their pre-screening state anxiety scores were higher than their trait anxiety scores. This finding can be interpreted as reactive concerns exhibited by patients due to their application for screening rather than being intrinsic. The present study also found a weak relationship between hCG levels and trait anxiety scores. However, it would be erroneous to conclude that elevated hCG levels are associated with increased maternal anxiety, as this interpretation would imply higher anxiety levels during the weeks of pregnancy when elevated hCG values are observed. Moreover, it has been reported that the levels of anxiety experienced by pregnant

Table 2. Prescreening (STAI-Ia) and postscreening (STAI-Ib) state anxiety scores of the patients

		n (%)	STAI-Ia (mean±SD)	STAI-Ib (mean±SD)	p-value
Gravida	1	46 (31.1)	40.35±6.01	37.33±6.33	<0.001
	≥2	102 (68.9)	39.56±6.13	37.50±5.44	<0.001
Parity	No	56 (37.8)	40.39±6.19	37.66±6.48	<0.001
	≥1	92 (62.2)	39.45±6.03	37.32±5.22	<0.001
History of miscarriage	No	114 (77.0)	39.82±5.91	37.16±5.42	<0.001
	Yes	34 (23.0)	39.76±6.71	38.41±6.58	0.115
Consanguineous marriage	No	129 (87.2)	39.86±6.09	37.35±5.56	<0.001
	Yes	19 (12.8)	39.42±6.18	38.11±6.75	0.122[α]
Having healty child	No	63 (42.6)	40.22±6.23	37.76±6.51	<0.001
	Yes	85 (57.4)	39.49±5.99	37.21±5.06	<0.001
Anomalous child*	No	137 (92.6)	39.68±6.09	37.57±5.68	<0.001
	Yes	11 (7.4)	41.36±6.08	35.91±6.12	0.01[α]
Maternal comorbidity	No	125 (84.5)	40.30±6.10	37.54±5.96	<0.001
	Yes	23 (15.5)	37.13±5.37	36.96±4.16	0.766[α]
Screening result	Negative	136 (91.9)	39.79±6.00	36.93±5.27	<0.001
	Positive	12 (8.1)	40.00±7.26	43.25±7.38	0.003[α]

Data was presented as mean±SD, number and percentage (%). α Wilcoxon test, *History of pregnancy termination due to an anomalous fetus or having a child with an anomaly, STAI: State-Trait Anxiety Inventory, SD: Standard deviation

individuals are similar to those observed in the general population.[15] In the study conducted by Mousavi et al.,[16] no relationship was found between anxiety scores and hCG levels in the second trimester.

The present study observed an increase in the state anxiety scores of the patients with positive test results. Many historical studies have shown that state anxiety scores of patients with positive test results for Down syndrome increased or have higher state anxiety scores than patients with negative test results.[15,17,18] Increased state anxiety levels in patients with positive test results can be attributed to the possibility of having a child with an anomaly and the emotional, social and economic burden this will bring. The present study also demonstrated that anxiety scores increased in all patients with a positive test result, and in 75 percent of patients the level of anxiety reached a probable clinical level. Therefore, patients with positive test results should be provided with psychological support in addition to medical counseling. In a study conducted by Richmond et al.,[19] it was demonstrated that anxiety levels of the patients with positive test results in the combined first trimester screening test was decreased following the non-invasive prenatal test (NIPT) being reported as low risk. These findings demonstrate the reliability of patients to the NIPT screening test, which exhibits a higher performance for the detection of Down syndrome when evaluated in terms of perinatology. Due to its lower false positive rate, it can be considered an alternative to the combined first trimester screening test, helping to avoid unnecessary maternal anxiety.

Consistent with the existing literature, this study also demonstrates that the post-screening state anxiety scores of patients with negative test results decrease in comparison to the pre-screening scores. In the meta-analysis which evaluated the effect of screening tests on maternal anxiety, it was reported that anxiety scores decreased in patients with negative test results and reached their lowest value in the postpartum period and also suggested that no evidence to support an assumption of residual anxiety in women with negative test results.[14] Bardi et al.[20] reported that 99% of patients expressed a desire to learn as soon as possible whether there was a structural anomaly in the fetus, and that patients' state anxiety scores decreased after normal fetal anatomy was confirmed. The findings of this study also demonstrated a significant decrease in the state anxiety scores of patients with a history of having an anomalous child or a history of pregnancy termination due to anomalous fetus after combined first trimester screening test and basic fetal anatomical scan. This decline may be attributable to the relief by the initial favorable results following a negative experience.

The study revealed no significant decrease in maternal anxiety scores following first trimester screening in pregnancies characterised by a history of miscarriage, consanguineous marriage, and maternal comorbidity. A history of miscarriage has been evidenced to cause negative psychological effects on patients, but the duration of post-miscarriage anxiety remains uncertain.[21] Despite the fact that screening test results are negative, patients' fear of reliving a negative experience, such as a miscarriage, may be the reason why their anxiety scores did not decrease. The increased risk of genetic diseases other than Down syndrome in consanguineous marriages and the increased risk of adverse pregnancy outcomes in those with maternal comorbidities may be the reason why patients' state anxiety scores did not show a significant change after the first trimester screening. As indicated by the findings of the present study, the investigation conducted by Akalın et al.[13] into the impact of fetal echocardiography on maternal anxiety revealed that the effect of fetal echocardiography on maternal anxiety scores was found to be insignificant in patients with a history of miscarriage and in patients with maternal comorbidities.

Limitations

The present study is not without its limitations. Firstly, the relatively small number of patients with positive test results is a notable factor. Secondly, state anxiety scores were assessed after the screening test results but could not be assessed after the diagnostic test, mid-trimester fetal anatomical scan or in postpartum period. The underlying factors contributing to this issue included patient non-compliance with follow-up appointments at the same center, the limitations of genetic diagnostic facilities within the laboratory, and patients' preference for delivering in different cities or institutions. It is recommended to confirm the findings with prospective studies with a larger patient population and longer follow-up period.

CONCLUSION

Consequently, patients' requests for screening tests have been shown to induce state anxiety in the patient. State anxiety levels decreases in patients with negative test results and increases in patients with positive test results. In addition to prenatal screening, diagnosis and follow-up, the importance of providing social and psychological support should be emphasised, particularly for patients with positive screening test results, to prevent the detrimental effects of maternal anxiety on pregnancy outcomes.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was carried out with the permission of the Yozgat Bozok University Rectorate Non-interventional Clinical Researches Ethics Committee (Date: 05.03.2025, Decision No: 2025-GOKAEK-255_05.03.2025-370).

Informed Consent

All patients signed and free and informed consent form.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

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

Author Contributions

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

Acknowledgments

The authors would like to express their gratitude to all the obstetricians at Yozgat City Hospital for their invaluable contributions in recruiting participants for the study.

Disclosure

The authors report no conflicts of interest in this work.

REFERENCES

1. Dryer R, Brunton R. Pregnancy-related anxiety. Theory, Research, and Practice. 1st ed. Routledge 2022.
2. Abdelmoula NB. The association between pregnancy-specific anxiety and fetal malformations: a cohort study. *Eur Psychiatry.* 2023;66(S1): S502-S502. doi:10.1192/j.eurpsy.2023.1068
3. Oyelese Y, Schioppo D, O'Brien B. Prenatal screening and diagnosis: time for a paradigm shift. *Am J Perinatol.* 2025;42(4):538-545. doi:10.1055/a-2312-8824
4. Alldred SK, Takwoingi Y, Guo B, et al. First trimester ultrasound tests alone or in combination with first trimester serum tests for Down's syndrome screening. *Cochrane Database Syst Rev.* 1996;2017(3): CD012600. doi:10.1002/14651858.CD012600
5. Halk Sağlığı Genel Müdürlüğü. Doğum öncesi bakım yönetim rehberi. Retrieved October; 2024.
6. Georgsson OS, Saltvedt S, Waldenström U, Grunewald C, Olin-Lauritzen S. Pregnant women's responses to information about an increased risk of carrying a baby with Down syndrome. *Birth.* 2006; 33(1):64-73. doi:10.1111/j.0730-7659.2006.00075.x
7. Fischbein RL, Nicholas L, Kingsbury DM, Falletta LM, Baughman KR, VanGeest J. State anxiety in pregnancies affected by obstetric complications: a systematic review. *J Affect Disord.* 2019;257:214-240. doi:10.1016/j.jad.2019.07.007
8. Madigan S, Oatley H, Racine N, et al. A meta-analysis of maternal prenatal depression and anxiety on child socioemotional development. *J Am Acad Child Adolesc Psychiatry.* 2018;57(9):645-657.e8. doi:10.1016/j.jaac.2018.06.012
9. Volpe N, Sen C, Turan S, et al. First trimester examination of fetal anatomy: clinical practice guideline by the World Association of Perinatal Medicine (WAPM) and the Perinatal Medicine Foundation (PMF). *J Perinat Med.* 2022;50(7):863-877. doi:10.1515/jpm-2022-0125
10. Spielberger CD, Gonzalez-Reigosa F, Martinez-Urrutia A, Natalicio LF, Natalicio DS. The State-Trait Anxiety Inventory. *Interam J Psychol.* 1971;5:3-4.
11. Öner N, LeCompte WA. Durumluk-sürekli kaygı envanteri el kitabı. Boğaziçi Üniversitesi Yayınları; 1985.
12. Grant KA, McMahon C, Austin MP. Maternal anxiety during the transition to parenthood: a prospective study. *J Affect Disord.* 2008; 108(1-2):101-111. doi:10.1016/j.jad.2007.10.002
13. Akalın M, Yalçın M, Demirci O, et al. Positive effects of fetal echocardiography on maternal anxiety: a prospective study in a tertiary center in Turkey. *J Psychosom Obstet Gynaecol.* 2022;43(4):585-592. doi:10.1080/0167482X.2022.2124911
14. Lou S, Mikkelsen L, Hvidman L, Petersen OB, Nielsen CP. Does screening for Down's syndrome cause anxiety in pregnant women? A systematic review. *Acta Obstet Gynecol Scand.* 2015;94(1):15-27. doi:10.1111/aogs.12482
15. Müller M, Bleker O, Bonsel G, Bilardo C. Nuchal translucency screening and anxiety levels in pregnancy and puerperium. *Ultrasound Obstet Gynecol.* 2006;27(4):357-361. doi:10.1002/uog.2761
16. Mousavi S, Hantoushzadeh S, Sheikh M, Shariat M. Maternal anxiety and the second-trimester prenatal screening: a prospective cohort study. *Fetal Diagn Ther.* 2015;38(4):269-275. doi:10.1159/000381640
17. Green JM, Hewison J, Bekker HL, Bryant LD, Cuckle HS. Psychosocial aspects of genetic screening of pregnant women and newborns: a systematic review. *Health Technol Assess.* 2004;8(33):iii-109. doi:10.3310/hta8330
18. Georgsson Ohman S, Grunewald C, Waldenström U. Perception of risk in relation to ultrasound screening for Down's syndrome during pregnancy. *Midwifery.* 2009;25(3):264-276. doi:10.1016/j.midw.2007.04.007
19. Richmond Z, Fleischer R, Chopra M, et al. The impact of non-invasive prenatal testing on anxiety in women considered at high or low risk for aneuploidy after combined first trimester screening. *Prenat Diagn.* 2017; 37(10):975-982. doi:10.1002/pd.5110
20. Bardi F, Bakker M, Kenkhuis MJ, et al. Psychological outcomes, knowledge and preferences of pregnant women on first-trimester screening for fetal structural abnormalities: a prospective cohort study. *PloS One.* 2021;16(1):e0245938. doi:10.1371/journal.pone.0245938
21. Campillo ISL, Meaney S, McNamara K, O'Donoghue K. Psychological and support interventions to reduce levels of stress, anxiety or depression on women's subsequent pregnancy with a history of miscarriage: an empty systematic review. *BMJ Open.* 2017;7(9):e017802. doi:10.1136/bmjopen-2017-017802

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1729141
J Health Sci Med. 2025;8(5):800-804

Social media usage among orthopaedic patients: what do they value most in orthopaedic surgeons' accounts?

Hayati Kart[1], Görkem Çat[2]

[1]Department of Orthopedics and Traumatology, Faculty of Medicine, İstanbul Medeniyet University, İstanbul, Turkiye
[2]Department of Orthopedic and Traumatology, Faculty of Medicine, Marmara University, İstanbul, Turkiye

Cite this article as: Kart H, Çat G. Social media usage among orthopaedic patients: what do they value most in orthopaedic surgeons' accounts? *J Health Sci Med.* 2025;8(5):800-804.

Received: 28.06.2025 ◆ **Accepted**: 29.07.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: Individuals use social media for various health-related purposes, such as gathering information about their illnesses, contacting healthcare professionals, learning about other patients' experiences with doctors, and making informed decisions when selecting doctors. The aims of this study are (I) to evaluate orthopaedic patients' utilisation of social media and (II) to ascertain the factors that are of greatest importance to patients in relation to orthopaedic surgeons' social media accounts.

Methods: A cross-sectional study was conducted with a total of 1018 patients in the orthopaedic outpatient department between November and December 2023. Patients filled out a questionnaire consisting of 15 questions about patients' personal information (gender, age, and education level), social media preferences, social media experiences, and the content of orthopaedic surgeons' social media accounts.

Results: 80.4% of the orthopedic patients surveyed were actively using at least one social media platform. Within the group of social media users, Instagram stood out as the most preferred platform, with a usage rate of 82.4%. Patients reported that 45.8% believe it is essential for doctors to maintain a presence on social media Patients were most interested (65.1%) in educational videos on orthopaedic surgeons' social media accounts.

Conclusion: The vast majority of orthopaedic patients are active users of social media, and a significant proportion of orthopaedic patients would like their doctors to have a social media presence. Orthopaedic surgeons should adapt their social media strategies to align with their patient population in order to effectively reach and engage with them.

Keywords: Social media, orthopaedic patients, orthopaedic surgeons, social media usage

INTRODUCTION

As of October 2023, there are 5.3 billion internet users worldwide, with 4.95 billion individuals, constituting 61.4% of the global population, actively using social media.[1] By the year 2027, it is anticipated that the number of individuals utilizing social media will reach 6 billion.[2] Mobile phones facilitate connectivity and allow mobile phone users to access social media anytime, anywhere.[3] It is clear that social media has become an integral part of our daily routines and its impact extends across diverse sectors and plays a crucial role in the realm of medicine.[4]

A considerable number of orthopaedic patients actively participate in social media platforms.[5,6] The patients use social media for various health-related purposes, such as gathering information about their illnesses, contacting healthcare professionals, learning about other patients' experiences with doctors, and making informed decisions when selecting doctors. Just like their patients, orthopaedic surgeons have not been indifferent to the recent changes in the social media world. The use of social media by orthopaedic surgeons has also reached significant levels.[7,8] A recent study of 208 orthopaedic surgeons in Germany reported that all participating orthopaedic surgeons used social media.[9] Doctors generally use social media platforms for purposes such as communicating with their patients, raising health awareness, sharing information on preventive health issues and enhancing their reputation. Irrespective of its intended purpose, it has become a common communication tool for both patients and orthopaedic surgeons.

Social media is often a useful research tool in the search for health-related information for patients.[10] Although it is widely used to obtain information, there are still doubts about its use as a source of accurate information due to the lack of restrictions and controls.[11] In order to address these concerns, the doctors' account is crucial for the patients. The use of social media has also changed the nature of the doctor-patient relationship. Social media accounts enable patients to reach out to their doctors in a more accessible manner than traditional methods.[12] The social media accounts allow doctors to exert a positive influence on their patients and to foster better relationships with them, thereby increasing the

Corresponding Author: Hayati Kart, hayatikart@gmail.com

number of patients who follow them. Thus, patients receive accurate health information from a doctor, not an unqualified person. However, there is a lack of studies in the literature that examine orthopaedic surgeons' social media accounts. The objective of this study is (I) to evaluate orthopaedic patients' utilisation of social media and (II) to ascertain the factors that are of greatest importance to patients in relation to orthopaedic surgeons' social media accounts.

METHODS

After obtaining ethical approval the Marmara University Faculty of Medicine Clinical Researches Ethics Committee (Date: 06.10.2023, Decision No: 09.2023.1261) our study was conducted with the participation of 1,018 orthopaedic patients at the Orthopedics and Traumatology Outpatient Clinic of Marmara University Hospital in November and December 2023. All procedures were carried out in accordance with the ethical rules and the principles of the Declaration of Helsinki. The study was designed as a cross-sectional study.[13]

Patients in our country with general health insurance utilize the system known as MHRS (Centralized Hospital Appointment System) to schedule appointments for examinations at state hospitals through phone or internet. Our university hospital is also a state hospital. The study included voluntary patients aged 18 and above who had made an appointment through the MHRS system and attended the orthopedic outpatient clinic. Non-voluntary and illiterate patients were excluded from the scope of the study.

The study was conducted face to face. Patients filled out a questionnaire ensuring the confidentiality of their identity information. The survey consisted of 15 questions regarding patients' personal information (gender, age, and education level), social media preferences, social media experiences, and the content of orthopaedic surgeons' social media accounts.

The study data was evaluated using the SPSS 26 (Statistical Package for the Social Sciences) program for statistical analyses. Descriptive statistical methods, such as mean, standard deviation, median, minimum and maximum values for quantitative variables, and frequency and percentage for qualitative variables, were employed. The normal distribution of the data was assessed using the Shapiro-Wilks test and Box Plot graphs. For quantitative variables with a normal distribution, we used the student T test to compare two groups, the One-way ANOVA test for comparisons involving three or more groups, and the Bonferroni test to determine the group responsible for the difference. For qualitative data comparisons, we used the Chi-square test. We evaluated the results at a significance level of $p<0.05$ with a 95% confidence interval.

RESULTS

The study was conducted with a total of 1018 participants, of whom 563 (55.3%) were women and 455 (44.79%) were men. The age of the participants ranged from 18 to 69 years, with a mean age of 39.10±12.81 years. The educational backgrounds of the patients are given in Figure 1. The study revealed that 80.4% of orthopaedic patients surveyed reported using at least one social media site. Patients preferring to use social media are shown in Figure 2. The responses of 818 patients utilizing social media to evaluate their social media experiences are presented in Table 1, and what patients consider to be more important regarding the content of doctors' social media accounts is detailed in Table 2.

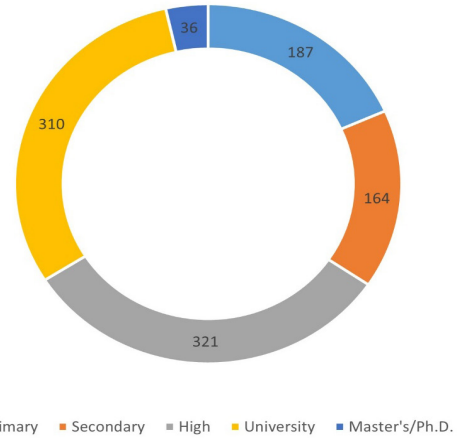

Figure 1. The educational backgrounds of the patients

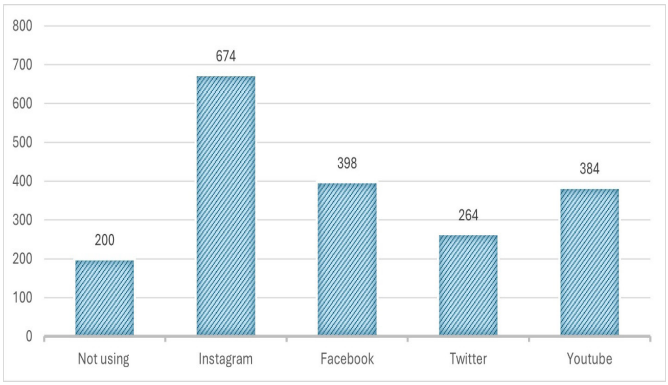

Figure 2. Patients' social media use

*Participants could give more than one answer to this question

Table 1. The social media experiences of the patients

		n/(%)
Have you ever asked a question to the doctor on social media?	Yes	183 (22.4)
	No	635 (77.6)
Do you search for your doctor on social media or the internet before coming to the hospital?	Yes	205 (25.1)
	No	613 (74.9)
Is it important for you that your doctor has a social media account?	Yes	375 (45.8)
	No	443 (54.2)
Do you pay attention to the doctor's profile picture?	Yes	283 (34.6)
	No	535 (65.4)
Is the doctor's number of followers important to you?	Yes	292 (35.7)
	No	526 (64.3)

The age averages of social media users were statistically significantly lower than those of non-users ($p=0.001$; $p<0.01$). Specifically, within the social media user groups, only individuals using Facebook showed a statistically significantly higher age compared to non-users, while the age averages of users on other social media platforms were lower than those

Table 2. The importance of content on doctors' social media accounts		
		n (%)
How important are the comments under doctor posts to you?	Not important	118 (14.4)
	Slightly	63 (7.7)
	Medium	188 (23)
	Important	268 (32.8)
	Very important	181 (22.1)
How important are posts about the doctor's private life to you?	Not important	466 (57.0)
	Slightly	89 (10.9)
	Medium	132 (16.1)
	Important	98 (12.0)
	Very important	33 (4.0)
How much do the comparative x-rays of patients before and after treatment affect you?	Not important	77 (9.4)
	Slightly	59 (7.2)
	Medium	153 (18.7)
	Important	314 (38.4)
	Very important	215 (26.3)
How much do the comparative photos/videos of patients before and after treatment affect you?	Not important	114 (13.9)
	Slightly	74 (9.0)
	Medium	161 (19.7)
	Important	294 (35.9)
	Very important	175 (21.4)
How important are the doctor's informational video shares about diseases for you?	Not important	67 (8.2)
	Slightly	56 (6.8)
	Medium	163 (19.9)
	Important	309 (37.8)
	Very important	223 (27.3)
Do you do research on the internet about which university your doctor graduated from?	Not important	164 (20.0)
	Slightly	66 (8.1)
	Medium	179 (21.9)
	Important	241 (29.5)
	Very important	168 (20.5)
How important are other patients' comments about your doctor for you?	Not important	81 (9.9)
	Slightly	49 (6.0)
	Medium	172 (21)
	Important	285 (34.8)
	Very important	231 (28.2)
How important is social media for you when choosing your doctor?	Not important	232 (28.4)
	Slightly	90 (11.0)
	Medium	223 (27.3)
	Important	171 (20.9)
	Very important	102 (12.5)

of non-users (p=0.031) (p<0.01). The ages of participants who previously asked questions to doctors through social media were statistically significantly lower than those who did not ask questions (p=0.001; p<0.01). The ages of participants who searched for their doctors on social media or the internet before going to the hospital were statistically significantly lower (p=0.002; p<0.01).

The evaluation based on educational backgrounds revealed that the usage of Instagram, Twitter, and YouTube was statistically significantly higher among university and doctorate graduates (p=0.001; p<0.01). In contrast, Facebook usage was statistically significantly lower among university and doctorate graduates compared to other educational groups (p=0.001; p<0.01). A statistically significant difference was found between the responses to the question 'Have you ever asked a doctor a question on social media?'. University graduates are more likely to answer 'yes' than primary, secondary and high school graduates. (p=0.001; p<0.01) A statistically significant difference was found between the answers to the question 'Have you searched for your doctor on social media or the internet before coming to the hospital?'. The rate of university and doctorate graduates answering yes is higher than primary, secondary and high school graduates (p=0.001; p<0.01). "Is it important for you that your doctor has a social media account?" There was a statistically significant difference in the responses to the question.". The rate of answering "yes" of university graduates and doctoral graduates is higher than those who have secondary and high school education. (p=0.001; p<0.01).

DISCUSSION

With the age of technology, the use of social media has become mainstream among orthopaedic patients. The majority of orthopaedic patients actively use social media. The results of the evaluation conducted among social media users indicated that Instagram is the platform with the highest level of user preference. Orthopaedic patients express a desire to see orthopaedic surgeons on social media. Orthopaedic patients highlight the importance of educational videos about ailments on orthopaedic surgeons' accounts.

The use of social media continues to increase day by day. In a study conducted by Curry et al.[5] in 2014, the internet usage rate among orthopedic patients was reported to be 51%. Pazarcı et al.[14] observed that the prevalence of social media use among orthopaedic patients was 39% in 2015. Aydin et al.[15] reported a social media usage rate of 83% among surgical patients, including orthopaedic patients, in their studies. The results of this study showed us that 80.4% of orthopaedic patients use social media. Thorne et al.[16] reported the social media usage rate of orthopedic surgeons as 74.4% in their study in 2020. In a study conducted by Youssef et al.[9] in 2023, evaluating the social media usage of 208 orthopedic and trauma surgeons, it was reported that all participants used social media. Evaluating our results with the literature, it is now clear that social media has become a widely used communication tool for both orthopaedic patients and orthopaedic surgeons. A critical issue for patients is easy access to their doctors. In their study, Duymus et al.[6] found that 46.7% of orthopaedic patients believed that orthopaedic doctors should communicate with patients through the internet. In this study, 45.8% of patients responded yes to the question "Is it important for your doctor to have a social media account?". Social media makes it easy for patients to reach and communicate with their doctor. The fact that patients can get in touch with their doctors more easily makes them feel more confident. Orthopaedic patients

therefore find it important for their doctor to have a social media account.

In today's world, there are many different social media platforms, with Facebook, Instagram, Twitter and YouTube being the most prominent and popular.[17] Aydın et al.[15] found that Instagram was the most popular platform among patients, used by 62%. Freiberger et al.[18] reported Facebook as the most preferred platform among patients (84%), with a notably higher usage among older adults. In our study, which involved a relatively young cohort, Instagram was the most frequently used social media platform (82.4%). Pazarcı et al.[14] reported in their studies that internet use increases with increasing educational level. Gencer et al.[11] reported an association between the increase in the level of education and internet usage, as well as social media habits. Our findings also indicate that an increase in educational attainment is correlated with social media use. Among social media users, individuals with lower educational levels tend to prefer Facebook, whereas those with higher educational levels favor other platforms. Demographic factors such as age and education significantly influence patients' social media usage and preferences. With this understanding, orthopaedic surgeons serving specific demographics can more effectively reach patients by selecting social media platforms that align with their patients' age and educational backgrounds. For instance, arthroplasty operations are usually performed for the elderly population, an orthopedic surgeon who performs arthroplasty operations can reach more patients through facebook. Informational videos about arthroplasty and potential complications posted by orthopaedic surgeons on Facebook can be useful for older patients. Similarly, an orthopedic surgeon specializing in sports surgery may prefer Instagram to reach a large number of young athletes. Training information provided via Instagram can help prevent possible injuries to athletes.

The profile of a social media account is a crucial element in the expression of who you are and what you have to offer in the digital world. Users create their own profiles, which usually contain their profile pictures, biographies (personal information), posts, and comments about these posts. People's first impressions are shaped by these profiles. Klietz et al.[19] conducted a study by creating a plastic surgery account to assess the specific interests of plastic surgery patients on social media. The study found that patients paid considerable attention to personal posts, aesthetic surgery topics, and medical conditions, while scientific posts received relatively low engagement. Freiberger et al.[18] reported that orthopaedic patients demonstrated the highest level of interest in educational videos and medical information, with engagement rates reaching 62%. In this study, we investigated what patients are most interested in on orthopaedic surgeons' social media accounts. The patients expressed that the educational videos on the social media accounts of orthopedic surgeons were most important for them (65.1%). Many people use social media to access information, but one of the biggest drawbacks of social media is the lack of restrictions and controls on the information available.[11] For this reason patients believe they can access accurate information about their conditions from doctors' social media accounts. This is also true for orthopaedic patients, who expect their doctors to share informative videos about their diseases on their social media accounts. In contrast to the plastic surgery patients, orthopedic patients indicated that posts related to the doctor's private life were not significant for them (67.9%).[19] Furthermore, patients have indicated that the profile picture and number of followers of their doctor are not significant factors in their decision-making process. The results of this study demonstrate that orthopaedic patients prefer to view their doctors as reliable sources of accurate information rather than as phenomena on social media. In Turkiye, the Ministry of Health introduced a regulation in 2023 mandating that health-related information shared on social media must be provided exclusively by legally authorized healthcare professionals. The regulation explicitly prohibits any content-even from licensed practitioners-that describes medical treatments or procedures lacking scientific validation or established clinical evidence. It also bans all forms of explicit and implicit advertising in the healthcare sector. These measures aim to curb misleading health claims and prevent unethical commercial exploitation within the field.

The advent of social media has introduced a novel dimension to the relationship between doctor and patient. Nevertheless, one constant persists: the desire of patients to place trust in their doctors. The posts made by doctors on their social media accounts, the results of the patients they have previously treated, and comments made by other patients can collectively increase patients' trust in doctors. 64.7% of the patients participating in our study reported that pre- and postoperative X-Rays shared by doctors on their social media accounts are important for them. Such social media posts of the doctors give patients an idea of the doctors' abilities and experiences in this field. 63% of the patients considered comments made by other patients about their doctors as important or very important. Positive comments from a previous patient about their doctor are likely to increase new patients' confidence in their doctors. In the view of Gencer at al.,[11] sharing information about doctors' CVs is crucial for for patients' trust. In this study, 50% of the patients stated that the educational background of their doctors is significant to them. Therefore, sharing personal information in the bios of orthopaedic doctors' social media accounts, such as graduation details, specialties, and areas of expertise, can also increase patients' trust in their doctorsIn other words, when used appropriately, social media can serve as an effective tool for fostering trust-based relationships between patients and physicians.

Limitations

Our study has several limitations. A major limitation is that it is a single-centre, cross-sectional study and it may not be sufficient to understand the general population. Another limitation is that our study was conducted in a university hospital, which is a government hospital. If the same questions were repeated in a private hospital, the results might be different because, as we know from our study, socio-demographic characteristics influence social media use.

A multicentre study that includes both private and public hospitals may provide more accurate results.

There are no comprehensive studies on the evaluation of orthopaedic surgeons' social media accounts from a patient perspective. The present study has been carried out in order to fill this crucial gap. Our study of a very large-scale group of patients has revealed the social media preferences of orthopaedic patients and their expectations of their doctors. With this information, orthopaedic surgeons will be able to better engage with their patients and meet their expectations by using social media more effectively.

CONCLUSION

As a result, our study demonstrates that social media is widely used by orthopaedic patients. Many patients value their doctors having a social media presence. This makes communication more accessible. It also helps establish trust. Furthermore, it allows patients to reach accurate medical information more easily. Preference for different social media platforms varies with demographic factors such as age and education level. This suggests that orthopaedic surgeons should adapt their social media strategies to align with their patient population in order to effectively reach and engage with them. For surgeons primarily treating geriatric patients, Facebook facilitates easier patient outreach, whereas those serving younger patient groups tend to be more effective by utilizing Instagram and other social media platforms. Although social media offers a valuable platform for patient education and physician-patient interaction, it is essential that physicians use it responsibly and ensure the accuracy and credibility of the information they share.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was carried out with the permission of the Marmara University Faculty of Medicine Clinical Researches Ethics Committee (Date: 06.10.2023, Decision No: 09.2023.12615).

Informed Consent

All patients signed and free and informed consent form.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

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

Author Contributions

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

REFERENCES

1. Internet and social media users in the world 2023 | Statista. Accessed December 4, 2023. https://www.statista.com/statistics/617136/digital-population-worldwide/
2. Number of worldwide social network users 2027 | Statista. Accessed December 4, 2023. https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/
3. Deng Z, Liu S. Understanding consumer health information-seeking behavior from the perspective of the risk perception attitude framework and social support in mobile social media websites. *Int J Med Inform*. 2017;105:98-109. doi:10.1016/J.IJMEDINF.2017.05.014
4. Malyavko A, Kim Y, Harmon TG, et al. Utility of social media for recruitment by orthopaedic surgery residency programs. *JBJS Open Access*. 2021;6(3):e21.00076. doi:10.2106/JBJS.OA.21.00076
5. Curry E, Li X, Nguyen J, Matzkin E. Prevalence of Internet and social media usage in orthopedic surgery. *Orthop Rev*. 2014;6(3):5483. doi:10.4081/or.2014.5483
6. 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. doi:10.5312/wjo.v8.i2.178
7. Duymuş TM, Karadeniz H, Şükür E, Atiç R, Zehir S, Azboy İ. Social media and Internet usage of orthopaedic surgeons. *J Clin Orthop Trauma*. 2017;8(1):25. doi:10.1016/J.JCOT.2016.10.007
8. Justinia T, Alyami A, Al-Qahtani S, et al. Social media and the orthopaedic surgeon: a mixed methods study. *Acta Inform Medica*. 2019;27(1):23-28. doi:10.5455/aim.2019.27.23-28
9. Youssef Y, Scherer J, Niemann M, Ansorg J, Back DA, Gehlen T. Social media use among orthopedic and trauma surgeons in Germany: cross-sectional survey study. *JMIR Form Res*. 2023;7(1):e45665. doi:10.2196/45665
10. Montemurro P, Porcnik A, Hedén P, Otte M. The influence of social media and easily accessible online information on the aesthetic plastic surgery practice: literature review and our own experience. *Aesthetic Plast Surg*. 2015;39(2):270-277. doi:10.1007/S00266-015-0454-3
11. Gencer B, Doğan Ö, Çulcu A, et al. Internet and social media preferences of orthopaedic patients vary according to factors such as age and education levels. *Health Info Libr J*. 2024;41(1):84-97. doi:10.1111/hir.12503
12. Gross CE, Scott D, Samora JB, Khan M, Kang DG, Frank RM. Physician-rating websites and social media usage: a global survey of academic orthopaedic surgeons: AOA critical issues. *J Bone Jt Surg Am Vol*. 2022;104(2):1-6. doi:10.2106/JBJS.20.01893
13. Sharma A, Tran Minh Duc N, Luu Lam Thang T, et al. A consensus-based checklist for reporting of survey studies (CROSS). *J Gen Intern Med*. 2021;36(10):3179-3187. doi:10.1007/s11606-021-06737-1
14. Pazarcı Ö, Gölge UH, Çamurcu Y, Saygılı H, Kılınç S, Keskinbıçkı MV, Korkmaz M. Internet use by orthopedic and traumatology patients in Turkey. *Cumhur Tıp Derg*. 2015;37(4):269-275. doi:10.7197/CMJ.V37I4.5000124235
15. Aydin MA, Gokcen HB. Factors considered by patients in provider selection for elective specialty surgery: a questionnaire-based survey. *Int J Surg*. 2019;69:43-48. doi:10.1016/j.ijsu.2019.07.007
16. Thorne D, Burton C, Tulloch S. Australian orthopaedic surgeons and social media: the future of education and communication? *AMJ*. 2020;13(7):239-246. doi:10.35841/1836-1935.13.7.239-246
17. De Martino I, D'Apolito R, McLawhorn AS, Fehring KA, Sculco PK, Gasparini G. Social media for patients: benefits and drawbacks. *Curr Rev Musculoskelet Med*. 2017;10(1):141-145. doi:10.1007/s12178-017-9394-7
18. 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 Sport Med Rehabil*. 2023;5(3):e649-e656. doi:10.1016/j.asmr.2023.03.004
19. Klietz ML, Kaiser HW, MacHens HG, Aitzetmüller MM. Social media marketing: what do prospective patients want to see? *Aesthetic Surg J*. 2020;40(5):577-583. doi:10.1093/asj/sjz204

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1740940
J Health Sci Med. 2025;8(5):805-810

The impact of monocyte to HDL ratio and Prognostic Nutritional Index on survival in stage III colorectal cancer patients

Mehmet Salim Demir[1], Merve Yumak[2]
[1]Department of Medical Oncology, Van Training and Research Hospital, Van, Turkiye
[2]Department of Gastroenterological Surgery, Van Training and Research Hospital, Van, Turkiye

Cite this article as: Demir MS, Yumak M. The impact of monocyte to HDL ratio and Prognostic Nutritional Index on survival in stage III colorectal cancer patients. *J Health Sci Med.* 2025;8(5):805-810.

Received: 12.07.2025 ◆ **Accepted**: 03.08.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: This study explores how systemic inflammatory and nutritional indicators, specifically the monocyte-to-high-density lipoprotein ratio (MHR) and Prognostic Nutritional Index (PNI), influence clinical outcomes in stage III colorectal cancer (CRC) patients.

Methods: A retrospective review of 109 individuals was conducted. ROC curve analysis was employed to determine the optimal cut-off values of MHR and PNI for predicting mortality. Survival outcomes, including overall survival (OS) and disease-free survival (DFS), were evaluated using Kaplan-Meier estimates and compared with log-rank tests. Cox regression was utilized to pinpoint factors independently associated with DFS.

Results: The findings revealed significantly lower OS among patients not undergoing adjuvant chemotherapy (45.6 vs. 82.2 months; p=0.002). Additionally, diminished MHR (<0.37) and PNI (<46.8) levels were linked to poorer OS (p=0.041 and p=0.003, respectively). While low PNI was also associated with reduced DFS (p=0.021), MHR did not significantly impact DFS (p=0.42). Both MHR (AUC: 0.643) and PNI (AUC: 0.657) demonstrated moderate predictive capabilities for mortality. Importantly, perineural invasion surfaced as an independent negative prognostic factor for DFS (HR: 2.36; p=0.038).

Conclusion: In conclusion, pre-treatment MHR and PNI values serve as accessible and low-cost indicators that may assist in prognostic stratification in stage III CRC management.

Keywords: Monocyte-to-HDL ratio, Prognostic Nutritional Index, colorectal cancer, stage III, survival, prognostic markers

INTRODUCTION

Colorectal cancer (CRC) ranks among the most prevalent malignancies globally, with high incidence and mortality rates across both developed and developing nations. According to GLOBOCAN 2024 estimates, it is the third most commonly diagnosed cancer and the second leading cause of cancer-related death worldwide.[1] In Turkiye, CRC represents 12.7% of all malignancies, with age-standardized incidence rates of 23.6/100,000 in males and 16.2/100,000 in females, reflecting unique regional patterns that warrant population-specific investigations.[2-4]

Stage III CRC defined by regional lymph node involvement without distant metastasis per AJCC 8th edition criteria[5] occupies a critical therapeutic window. While surgical resection offers potential curability, occult micrometastases drive recurrence rates exceeding 30% despite adjuvant chemotherapy.[6] Current 5-year survival rates plateau around 60-65%, underscoring limitations of conventional TNM staging and highlighting the urgent need for refined prognostic tools.[7] Heterogeneity in treatment response remains poorly explained by histopathology alone, as tumors with identical staging may exhibit divergent biological behaviors influenced by host inflammatory responses and nutritional status.[8]

Systemic inflammation constitutes the seventh hallmark of cancer, fostering a tumor-permissive microenvironment through multiple pathways: angiogenesis induction, DNA damage acceleration, and immune evasion.[8] Circulating immune cells serve as quantifiable sentinels of this process, with neutrophil-to-lymphocyte ratio (NLR) and platelet-to-lymphocyte ratio (PLR) extensively validated as prognostic indicators.[9] More recently, the monocyte-to-HDL cholesterol ratio (MHR) has emerged as a superior biomarker by integrating pro-tumorigenic and cardioprotective mechanisms. Monocytes promote metastasis via matrix metalloproteinase secretion and immunosuppressive cytokine production (e.g., IL-10, TGF-β),[10] while HDL cholesterol exerts anti inflammatory effects through endothelial protection and oxidized lipid clearance.[11] Elevated MHR thus signifies disrupted homeostasis favoring tumor progression, with meta-analyses confirming its prognostic value across gastrointestinal malignancies.[12]

Corresponding Author: Mehmet Salim Demir, ahmetcagri22@gmail.com

Nutritional compromise frequently parallels cancer associated inflammation, creating a vicious cycle that accelerates cachexia and immunosuppression. The Prognostic Nutritional Index (PNI) calculated from serum albumin and lymphocyte counts quantifies this dual insult. Albumin synthesis suppression reflects hepatic inflammatory signaling (IL-6 mediated), while lymphopenia indicates adaptive immune impairment.[13,14] Multiple studies demonstrate PNI's predictive power, with thresholds <45 reducing 5-year survival by 30-40% in gastric and CRCs.[15,16] Notably, PNI's prognostic independence from body mass index makes it particularly valuable in obese populations where traditional nutritional assessments fail.[14]

Given this biological rationale, both MHR and PNI offer promising insight into the tumor microenvironment and host-tumor interactions. Despite their growing use in clinical research, few studies have examined their combined prognostic value specifically in patients with stage III colorectal adenocarcinoma, where treatment decisions hinge on accurate survival predictions. Therefore, identifying whether these simple biomarkers are capable of predicting survival and recurrence may have significant implications for clinical practice.

This study aims to investigate the prognostic significance of preoperative MHR and PNI levels in patients with stage III colorectal adenocarcinoma. By examining their association with overall survival (OS) and disease-free survival (DFS), we seek to determine whether these cost-effective indices can supplement conventional prognostic tools and contribute to more personalized treatment approaches.

METHODS

This study was approved by the Van Training and Research Hospital Non-interventional Clinical Researches Ethics Committee (Date: 04.07.2025, Decision No: GOKAEK/2025-05-11) and conducted in accordance with the Declaration of Helsinki. Informed consent was waived due to the retrospective nature.

This retrospective cohort study included 109 patients who were diagnosed with stage III colorectal adenocarcinoma and underwent surgical resection between 2015 and 2022 at a single tertiary care center. Inclusion criteria required histopathologically confirmed stage III disease. Given the sample size (n=109), multivariate analyses were limited to ≤5 covariates to avoid overfitting. Patients with chronic inflammatory diseases, active infections, hematologic disorders, second malignancies, or missing data were excluded.

Comprehensive clinical, pathological, and demographic data were retrieved from electronic medical records. Collected parameters included patient age, sex, tumor location, surgical urgency (elective vs. emergency), histological grade, T stage, lymphovascular and perineural invasion status, number of harvested lymph nodes, number of metastatic lymph nodes, and details of adjuvant chemotherapy regimen administered (Table 1).

Table 1. Patients demographic characteristics

	n	%
Gender		
Female	45	41.3
Male	64	58.7
T stage		
T2	12	11
T3	65	59.6
T4	32	29.4
Histologic grade		
Grade 1	37	33.9
Grade 2	60	55.1
Grade 3	12	11
Lymphovascular invasion	60	55.0
Perineural invasion	50	45.9
Tumor localisation		
Rectum	27	24.8
Right colon	40	36.7
Left colon	42	38.5
Emergency operation status	22	20.2
Adjuvant chemotherapy	95	87.2
Adjuvant chemotherapy (n=95)		
Capox	40	39.6
Folfox	41	40.6
Capecitabine	14	13.8
Relapse	43	39.4
Mortality	30	27.5
PNI		
Low	31	28.4
High	78	71.6
MHR		
Low	46	42.2
High	63	57.8
	Median±SD	**Median (min-max)**
Age	62.2±13.4	63 (40-84)
Harvested lymph nodes	19.6±8.3	18 (4-46)
Positive lymph nodes	5.61±3.9	5 (1-20)
PNI	48.9±3.9	49.4 (40-56.9)
MHR	0.44±0.17	0.4 (0.01-0.89)
Positive lymph nodes rates	0.29±0.17	0.26 (0.04-0.83)

PNI: Prognostic Nutritional Index, MHR: Monocyte-to-HDL cholesterol ratio, HDL: High-density lipoprotein, SD: Standard deviation, Min: Minimum, Max: Maximum

Blood samples were collected after a 12-hour fasting period. In patients undergoing elective surgery, samples were obtained within 7 days prior to the operation. For those undergoing emergency procedures, preoperative laboratory tests were drawn within 24 hours of hospital admission, as part of the routine workup. Serum HDL cholesterol and absolute monocyte counts were measured using automated spectrophotometry within 2 hours of collection.

Albumin levels were quantified via bromocresol green method. The PNI was calculated using the formula: 10×serum albumin (g/dl)+0.005×total lymphocyte count (per mm^3).[14] The MHR was derived by dividing absolute monocyte count by HDL cholesterol level.[12] All laboratory values used in the calculations were obtained from preoperative blood samples.

Statistical Analysis

Data analyses were performed using IBM SPSS Statistics for Windows, Version 25.0 (IBM Corp., Armonk, NY). Categorical variables were expressed as frequencies and percentages, while continuous variables were presented as medians±standard deviations or medians with ranges, depending on distribution. The Kolmogorov-Smirnov test was used to assess normality. Group comparisons were conducted using Chi-square or Fisher's exact tests for categorical variables.

Receiver operating characteristic (ROC) curve analysis was utilized to identify optimal cut-off values for MHR and PNI based on their ability to predict OS. Kaplan-Meier survival curves were generated to assess OS and disease-free survival (DFS), with statistical significance evaluated using the log-rank test. DFS was defined as the time from surgery to the first documented tumor recurrence or radiologic evidence of relapse. Variables that reached significance in univariate analysis were further examined in a multivariate Cox proportional hazards model to determine independent predictors of DFS. A two-tailed p-value <0.05 was considered statistically significant throughout the analysis. Multivariate Cox regression model included the following variables that were significant in univariate analysis: histologic grade, perineural invasion, emergency operation status, age group, and PNI level. Due to the limited number of death events (n=30), OS was not analyzed in multivariate fashion to avoid statistical overfitting.

RESULTS

A total of 109 patients with stage III colorectal adenocarcinoma were included in the study. The median age of the cohort was 62.2±13.4 years (medain: 63), and the gender distribution comprised 64 males (58.7%) and 45 females (41.3%).

In terms of tumor T staging, 11% of patients were classified as T2, 59.6% as T3, and 29.4% as T4. Regarding histologic differentiation, 33.9% of tumors were grade 1, 55.1% were grade 2, and 11% were grade 3. Lymphovascular invasion (LVI) was identified in 60 patients (55%), while perineural invasion was present in 50 patients (45.9%).

Tumor localization was as follows: rectum in 24.8% of cases, right colon in 36.7%, and left colon in 38.5%. Surgeries were performed electively in 87 patients (79.8%), whereas 22 (20.2%) underwent emergency procedures. Adjuvant chemotherapy was administered to 95 patients (87.2%), with the following regimens recorded: CAPOX (40 patients), FOLFOX (41 patients), and oral capecitabine (14 patients) (Table 1).

During the follow-up period, 43 patients (39.4%) experienced disease recurrence, and 30 (27.5%) died. The median number of harvested lymph nodes was 19.6±8.3, with a median of 18. The average number of positive nodes was 5.61±3.9, yielding a median lymph node ratio of 0.26.

Laboratory-based indicators revealed an average MHR of 0.44±0.17 (median: 0.40), with 42.2% of patients having values <0.37. The mean PNI was calculated at 48.9±3.9 (median: 49.4), and 28.4% of the cohort had PNI values below 46.8 (Table 2). ROC analysis, performed using OS as the endpoint, identified the cut-off points for MHR and PNI as <0.37 and <46.8, respectively.

Table 2. Examination of ROC curve test for the ability of MHR and PNI values to predict mortality

	MHR	PNI
Cut-Off	<0.37	<46.8
AUC (95% CI)	0.643 (0.545-0.732)	0.657 (0.560-0.745)
Sensitivity (95% CI)	60 (40.6-77.3)	53.33 (34.3-71.7)
Specificity (95% CI)	63.29 (51.7-73.9)	75.95 (65-84.9)
PPV (95% CI)	38.3 (29.1-48.4)	45.7 (33.5-58.5)
NPV (95% CI)	80.6 (72.3-86.9)	81.1 (74.1-86.5)
p	0.017*	0.009**

*p<0.05, ROC curve test, ROC: Receiver operating characteristic, MHR: Monocyte-to-HDL cholesterol ratio, HDL: High-density lipoprotein, PNI: Prognostic Nutritional Index, AUC: Area under curve, CI: Confidence interval, PPV: Positive predictive value, NPV: Negative predictive value

The AUC values were 0.643 (95% CI: 0.545-0.732; p=0.017) for MHR and 0.657 (95% CI: 0.560-0.745; p=0.009) for PNI, suggesting moderate prognostic utility (Figure 1).

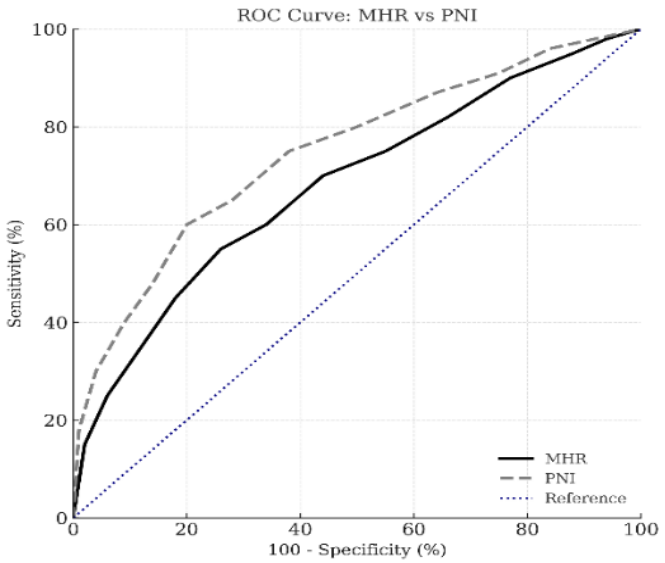

Figure 1. Examination of the mortaity prediction ability of MHR and PNI values using the ROC curve test

*The diagnostic performance of MHR and PNI values in predicting mortality among the included patients was evaluated using the ROC curve test. According to the analysis results, the cut-off values for MHR and PNI were determined to be <0.37 and <46.8, respectively. MHR: Monocyte-to-HDL cholesterol ratio, HDL: High-density lipoprotein, PNI: Prognostic Nutritional Index, ROC: Receiver operating characteristic

Survival analysis demonstrated significantly reduced OS among patients who did not receive adjuvant chemotherapy (45.6 months vs. 82.2 months; p=0.002) ant they had higher Charlson Comorbidity Index scores (median: 6.2 vs. 3.1; p=0.01) (Figure 2A). Patients with PNI <46.8 had markedly lower OS (65.3 months vs. 85.1 months; p=0.003) (Figure 2B). Likewise, individuals with MHR <0.37 had inferior OS

compared to those with higher values (71.7 vs. 84.4 months; p=0.041) (Figure 2C). No significant OS differences were observed for gender (p=0.842), tumor location (p=0.466), LVI (p=0.323), or perineural invasion (p=0.285) (Table 3).

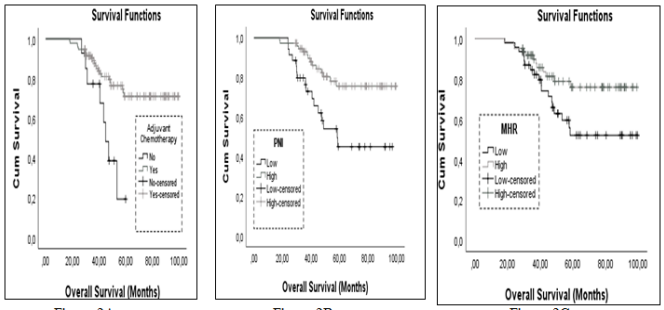

Figure 2A Figure 2B Figure 2C

Figure 2. Kaplan-Meier curves for OS stratified by adjuvant chemotherapy status, PNI, and MHR

OS: Overall survival, PNI: Prognostic Nutritional Index, MHR: Monocyte-to-HDL cholesterol ratio, HDL: High-density lipoprotein

Table 3. Differences between the median survival times of patients and variables

	Median	SE	95% CI Min	95% CI Max	p
OS (months)	78.9	3.1	72.8	85.1	
Gender					
Female	77.9	4.5	69.0	86.8	0.842
Male	78.6	4.1	70.5	86.6	
Adjuvant chemotherapy					
No	45.6	3.1	39.5	51.8	0.002**
Yes	82.2	3.2	76.0	88.5	
MHR (<0.37)					
Low	71.7	4.9	62.0	81.4	0.041*
High	84.4	3.8	77.1	91.8	
PNI (<46.8)					
Low	65.3	5.6	54.4	76.2	0.003**
High	85.1	3.4	78.4	91.8	

*p<0.05, **p<0.01, Kaplan meier, Log rank test, SE: Standard error, CI: Confidence interval, Min: Minimum, Max: Maximum, OS: Overall survival, MHR: Monocyte-to-HDL cholesterol ratio, HDL: High-density lipoprotein, PNI: Prognostic Nutritional Index

The median disease-free survival (DFS) for the cohort was 16.8±1.0 months (Table 4). Univariate analysis indicated significantly worse DFS in patients with grade 3 tumors (p=0.008), presence of perineural invasion (p=0.043) (Figure 3A), emergency surgery (p=0.013) (Figure 3B), and low nutritional PNI values (p=0.021) (Figure 3C). No significant DFS differences was observed for MHR (Figure 3D). Age >65 years (p=0.040) (Figure 3E).

Multivariate Cox regression analysis revealed that only perineural invasion remained an independent predictor of shortened DFS (HR: 2.361; p=0.038), while other variables did not retain statistical significance (Table 5). Multivariate analysis for OS was not performed due to the limited number of death events (n=30), which may reduce statistical power and increase the risk of overfitting. The median follow-up time for surviving patients was 42 months (range: 18-96 months), calculated using the reverse Kaplan-Meier method.

Table 4. Differences between the variables and the median disease-free survival time of the patients

	Median	SE	95% CI Min	95% CI Max	p
DFS (months)	16.8	1.0	14.8	18.8	
Gender					
Female	16.9	1.7	13.7	20.2	0.703
Male	16.6	1.2	14.2	19.0	
Hystologic grade					
Grade 1	17.6	1.3	15.0	20.2	0.008**
Grade 2	17.5	1.8	14.0	21.0	
Grade 3	10.7	1.7	7.4	13.9	
Perineural invasion					
No	17.7	1.5	14.8	20.6	0.043*
Yes	15.5	1.2	13.1	17.8	
Emergency operation status					
No	17.9	1.2	15.6	20.2	0.013*
Yes	13.0	1.7	9.8	16.3	
PNI (<46.8)					
Low	11.4	1.2	9.2	16.7	0.021*
High	20.5	1.4	17.7	23.3	
Age					
65 and under	17.8	1.3	15.2	20.4	0.040*
65 upper	14.8	1.4	12.1	17.5	

*p<0.05, **p<0.01, Kaplan meier, Log rank test, SE: Standard error, CI: Confidence interval, Min: Minimum, Max: Maximum, DFS: Disease-free survival, PNI: Prognostic Nutritional Index

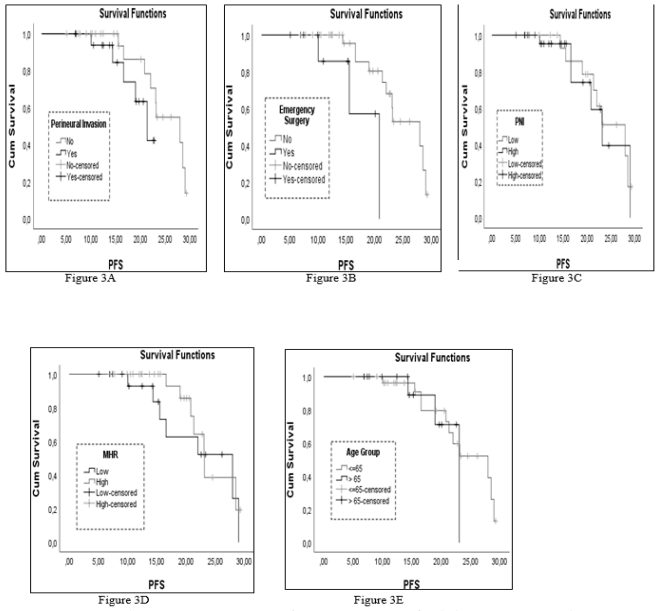

Figure 3A Figure 3B Figure 3C Figure 3D Figure 3E

Figure 3. Kaplan-Meier curves for DFS stratified by perineural invasion, surgery urgency, PNI, MHR, and age group

DFS: Disease-free survival, PNI: Prognostic Nutritional Index, MHR: Monocyte-to-HDL cholesterol ratio, HDL: High-density lipoprotein

Table 5. Examining the factors affecting the median disease-free survival time

		Hazard ratio	95% CI		
			Min	Max	
Hystologic grade					
Grade 1					0.134
Grade 2	-1.297	0.273	0.076	0.980	0.046*
Grade 3	-1.066	0.344	0.091	1.307	0.117
Perineural invasion	0.859	2.361	1.050	5.308	0.038*
Emergency operation	0.659	1.933	0.734	5.094	0.182
PNI (<46.8)					
Low	0.742	2.100	0.954	4.642	0.067
High					
Age					
65 and under					
65 upper	0.688	1.989	0.829	4.772	0.124

*p<0.05, **p<0.01, Multivariate Cox regresyon test, **Perineural invasion emerged as the only independent predictor of worse DFS (HR: 2.361; p=0.038), surpassing other variables in prognostic significance. DFS: Disease-free survival, CI: Confidence interval, Min: Minimum, Max: Maximum, PNI: Prognostic Nutritional Index

DISCUSSION

This study provides evidence supporting the prognostic utility of the MHR and PNI in patients with stage III CRC. Our findings demonstrate that lower MHR and PNI values are significantly associated with poorer OS and DFS, reinforcing the role of systemic inflammation and nutritional status in cancer progression.[14-16,19]

The considerable survival benefit observed with adjuvant chemotherapy reaffirms current clinical guidelines.[6] The discrepancy in OS between treated and untreated patients (82.2 vs. 45.6 months; p=0.002) is consistent with the IDEA collaboration and molecular subtype studies demonstrating survival benefit particularly in CMS2 and CMS4 subtypes.[6,17] The observed OS difference in the non-chemotherapy group may be confounded by higher comorbidity burden.

The prognostic impact of PNI aligns with previous reports. A Japanese prospective study reported 5-year OS below 58% for patients with PNI <45, compared to 78% for those with higher scores.[16] In our cohort, a similar pattern was observed-patients with PNI <46.8 had a median OS of 65.3 months, while those with higher values reached 85.1 months (p=0.003), underscoring the importance of nutritional resilience in CRC prognosis. Given that PNI predicted both OS and DFS, its role as a modifiable risk factor through nutritional interventions deserves attention. Early nutritional support, particularly in patients with borderline PNI values, may improve treatment tolerance and reduce recurrence.

Likewise, our data support the relevance of MHR as a marker of tumor-promoting inflammation. Monocytes promote tumor progression via angiogenesis and immune suppression,[10] while HDL exerts anti-inflammatory effects by inhibiting monocyte adhesion.[11] This mechanistic balance explains MHR's prognostic value. In our analysis, patients with MHR <0.37 experienced inferior OS (71.7 months vs. 84.4 months; p=0.041), corroborating findings from Korean retrospective cohorts and recent meta-analyses.[19]

Interestingly, MHR was significantly associated with OS but not DFS. This discrepancy may reflect the long-term systemic impact of chronic inflammation on survival, rather than short-term recurrence risk. Monocyte-driven mechanisms, such as immune exhaustion and vascular remodeling, may accelerate mortality in the absence of direct tumor progression.

Pathological perineural invasion, a classic histopathologic feature, was identified as an independent predictor of DFS (HR: 2.36; p=0.038), consistent with previous findings by Ishii et al.[17] This supports the hypothesis that perineural invasion reflects more aggressive tumor biology with greater potential for recurrence. Additionally, reduced DFS in patients undergoing emergency surgery highlights the need for early detection and optimized surgical protocols.[12]

The high rate of recurrence and death in our cohort may partially stem from the symptomatic presentation of most cases. As evidenced in European registry studies, patients diagnosed through screening had 5-year OS rates exceeding 83%, compared to only 57.5% among symptomatic patients,[22] suggesting that delayed diagnosis contributes to poorer outcomes.

In conclusion, low PNI and high MHR reflect malnutrition and systemic inflammation, directly impairing OS/DFS. These accessible biomarkers (requiring routine blood tests) offer clinical utility for risk stratification. The 45.6-month OS without adjuvant chemotherapy mirrors IDEA trial data,[6] emphasizing non-compliance as a modifiable risk factor. As the sole independent DFS predictor in multivariate analysis, presence of perineural invasion signifies aggressive tumor biology and neural spread-aligning with Ishii et al.'s[17] findings. Reduced DFS in older patients and emergent surgeries highlights needs for geriatric oncology protocols and early interventions.[12]

Our findings suggest that lower MHR and PNI values may identify high-risk patients who could benefit from more intensive monitoring and tailored therapeutic strategies. Given their accessibility and cost-effectiveness, MHR and PNI could be incorporated into standard risk stratification algorithms alongside conventional clinical and pathological parameters. Future prospective studies with larger, multicenter cohorts are warranted to confirm these observations and further validate their prognostic value. Our study is among the few to explore MHR and PNI in a Turkish cohort. Regional dietary habits, access to screening programs, and inflammatory profiles may differ from Western populations, thereby supporting the value of population-specific analyses.

Limitations

This study has several limitations. First, the relatively small sample size (n=109) may limit the statistical power of subgroup analyses and increase the risk of errors. Although multivariate analysis was restricted to ≤5 covariates to avoid overfitting, larger prospective cohorts are needed to validate our findings. In addition, the cut-off values identified via ROC analysis were not validated in an independent external

cohort, which limits their generalizability and may introduce overfitting bias.

CONCLUSION

While our findings are compelling, several limitations warrant consideration. The retrospective design introduces potential selection bias, and the lack of molecular data limits insights into tumor biology. Nonetheless, the inclusion of practical, low-cost markers such as MHR and PNI enhances the study's clinical applicability. Future prospective and multicenter studies are needed to validate these findings and further refine prognostic stratification models. In summary, this study highlights the clinical relevance of preoperative MHR and PNI as prognostic indicators in patients with stage III CRC. Both markers, derived from routine blood tests, were significantly associated with survival outcomes, offering insight into the interplay between systemic inflammation, nutritional status, and tumor progression.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was carried out with the permission of the Van Training and Research Hospital Non-interventional Clinical Researches Ethics Committee (Date: 04.07.2025, Decision No: GOKAEK/2025-05-11).

Informed Consent

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

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

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

Author Contributions

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

REFERENCES

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2024: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2024;74(3):229-263. doi:10.3322/caac.21834
2. Türkiye Cumhuriyeti Sağlık Bakanlığı. Türkiye kanser istatistikleri 2023. 2024. [Accessed via: saglik.gov.tr]
3. Öztürk A, Yalçın Ş, Demirci S, et al. Clinicopathological features of colorectal cancer in Turkey: a multicenter study. *Turk J Gastroenterol*. 2023;34(6):618-626. doi:10.5152/tjg.2023.22198
4. Demirci NS, Aydin H, Korkmaz A, et al. Systemic inflammation markers in Turkish colorectal cancer patients. *Eur J Cancer Care*. 2024;33(1):e14185. doi:10.1111/ecc.14185
5. Amin MB, Edge SB, Greene FL, et al. AJCC cancer staging manual. 8th ed. Springer; 2017. doi:10.1007/978-3-319-40618-3
6. Grothey A, Sobrero AF, Shields AF, et al. Duration of adjuvant chemotherapy for stage III colon cancer. *N Engl J Med*. 2018;378(13):1177-1188. doi:10.1056/NEJMoa1713709
7. Argilés G, Tabernero J, Taieb J, et al. ESMO Consensus guidelines for management of early colon cancer. *Ann Oncol*. 2024;35(1):10-32. doi:10.1016/j.annonc.2023.10.001
8. Mantovani A, Allavena P, Marchesi F, et al. Cancer-related inflammation: the seventh hallmark of cancer. *J Exp Med*. 2023;220(6):e20230614. doi:10.1084/jem.20230614
9. Templeton AJ, McNamara MG, Šeruga B, et al. Prognostic role of neutrophil-to-lymphocyte ratio in solid tumors: a systematic review and meta-analysis. *J Natl Cancer Inst*. 2014;106(6):dju124. doi:10.1093/jnci/dju124
10. Qian BZ, Pollard JW. Macrophage diversity in cancer. *Cancer Cell*. 2024;42(2):159-178. doi:10.1016/j.ccell.2024.01.007
11. Zaman R, Watanabe J, Hoang D, et al. HDL-mediated lipid trafficking in cancer pathogenesis. *Trends Endocrinol Metab*. 2023;34(12):875-888. doi:10.1016/j.tem.2023.09.003
12. Zhang L, Wang Y, Chen Z, et al. Prognostic value of monocyte-to-HDL ratio in cancers: a meta-analysis. *Front Oncol*. 2023;13:1127896. doi:10.3389/fonc.2023.1127896
13. Murphy AJ, Westerterp M, Tall AR. HDL in the tumor microenvironment: a multifunctional mediator. *Cancer Immunol Res*. 2024;12(1):3-12. doi:10.1158/2326-6066.CIR-23-0715
14. Sun K, Chen S, Xu J, et al. Prognostic Nutritional Index predicts survival and correlates with systemic inflammation in cancer patients. *Clin Nutr*. 2023;42(8):1350-1358. doi:10.1016/j.clnu.2023.06.020
15. Chen Y, Zhao X, Xue Y, et al. Monocyte to HDL ratio predicts survival in stage III colorectal cancer. *J Cancer Res Clin Oncol*. 2023;149(11):8765-8773. doi:10.1007/s00432-023-04820-0
16. Onodera T, Goseki N, Kosaki G. Prognostic Nutritional Index in gastrointestinal surgery. *Nutr Cancer*. 2024;76(1):1-9. doi:10.1080/01635581.2023.2265078
17. Ishii Y, Hasegawa H, Hirata N, et al. Perineural invasion as a predictor of recurrence in colorectal cancer. *Ann Surg Oncol*. 2022;29(12):7894-7902. doi:10.1245/s10434-022-12027-9
18. Guinney J, Dienstmann R, Wang X, et al. The consensus molecular subtypes of colorectal cancer. *Nat Med*. 2015;21(11):1350-1356. doi:10.1038/nm.3967
19. Lee GH, Park Y, Kim HY, et al. Emergency surgery for colorectal cancer: survival differences by surgical procedure. *J Gastrointest Surg*. 2023;27(5):1023-1032. doi:10.1007/s11605-023-05629-3
20. Wang Y, Zhao Q, Peng C, et al. Prognostic Nutritional Index and survival in gastrointestinal cancers: a meta-analysis. *JAMA Netw Open*. 2023;6(4):e238124. doi:10.1001/jamanetworkopen.2023.8124
21. Innocenti F, Pugh SL, Gonen M, et al. Precision oncology in colorectal cancer. *J Clin Oncol*. 2023;41(12):2187-2202. doi:10.1200/JCO.22.02524

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1734344
J Health Sci Med. 2025;8(5):811-817

Efficacy and safety of intravaginal gas ozone therapy in the treatment of bacterial vaginosis

Özgür Ağlamış[1], Seval Yılmaz Ergani[1], Eda Adeviye Şahin[2], Hanifi Şahin[3], Turan Şahin[4]

[1]Department of Perinatology, Ankara Etlik City Hospital, Ankara, Turkiye
[2]Departmant of Obstetrics and Gynecology, Dr. Eda Şahin Private Clinic, İstanbul, Turkiye
[3]Departmant of Obstetrics and Gynecology, İstanbul Health and Technology University, İstanbul, Turkiye
[4]Departmant of Obstetrics and Gynecology, Sancaktepe Şehit Prof. Dr. İlhan Varank Trainig Research Hospital, İstanbul, Turkiye

Cite this article as: Ağlamış Ö, Yılmaz Ergani S, Şahin EA, Şahin H, Şahin T. Efficacy and safety of intravaginal gas ozone therapy in the treatment of bacterial vaginosis. *J Health Sci Med.* 2025;8(5):811-817.

Received: 04.07.2025 ◆ **Accepted**: 04.08.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: The aim of this study is to share our experiences and results with our intravaginal gas ozone technique in the treatment of bacterial vaginosis (BV) by retrospective data analysis.

Methods: Salutem plus brand medical ozone device, 60 gamma 200 cc gas ozone was administered intravaginally through 50 cc syringes with the help of a nasogastric catheter to 102 patients diagnosed with BV using Amsel criteria. Patients whose hemogram parameters were checked before and after the procedure were called for 1st, 3rd, 6th, 9th and 12th month clinical controls in terms of recurrence, and the examination findings and digital data of these patients were analyzed retrospectively in terms of treatment, complications and recurrence.

Results: No life-threatening complications were encountered. Minor adverse effects were observed and resolved spontaneously in a short time without the need for any treatment. Recurrence rates were observed to be lower than in classical regimen treatments. Additionally, no more than 2 recurrences were observed during 1-year follow-up.

Conclusion: Our study is the first example of the literature in terms of evaluation of treatment, recurrence and complications of intravaginal gas ozone therapy in the treatment of bacterial vaginosis. These finding suggest that intravaginal gas ozone therapy, which is considered a safe form of treatment, will become more widespread in the future and the number and frequency of its application will increase even more. Further investigation with a larger number of patients should be conducted to confirm our data.

Keywords: Bacterial vaginosis, intravaginal, ozone therapy, ozone gas

INTRODUCTION

Bacterial vaginosis (BV) is a condition that usually affects women of reproductive age and can be symptomatic or asymptomatic. Approximately 50% of women are symptomatic. Foul-smelling vaginal discharge, itching and increased vaginal pH are observed.[1] Gardner first described BV in 1955.[2] BV is characterized by an overgrowth of opportunistic bacteria and a decrease in Lactobacilli levels, rather than an infection condition.[3] BV is diagnosed using Amsel criteria. Diagnosis is made when the vaginal pH value is above 4.5, when 10% potassium hydroxide solution is added to the vaginal discharge, there is an amine smell, when clue cells are found in wet preparations,homogeneous, non-viscous,milky-white discharge adherent to the vaginal wall.[4] Metronidazole and clindamycin are used in the treatment of BV. Antibiotics can be used orally or intravaginally. These recommended regimens have similar effectiveness.[5]

According to the guidelines, treatment is recommended only for symptomatic women because there is not enough evidence to support the treatment of asymptomatic women.[6] A high rate of recurrent infection, that is, recurrence, is observed in bacterial vaginosis treatments.[7] Ozone, trioxygen (O3), is a highly reactive inorganic molecule.[8] In animal experiments conducted on mice, it was observed that ozone had a positive effect on the treatment of Asherman syndrome, had a significant effect during pregnancy, and revealed a remodeling effect in tissues.[9] Ozone gas is also a powerful microorganism killer. It has antibacterial, antiviral, antifungal and antiparasitic properties.[8] Gas ozone can quickly and effectively oxidize the cell wall and cytoplasmic membrane of bacteria. It has a degrading effect on the cell membrane, thus preventing biofilm formation and reducing resistance to treatment.[10] It has recently been noticed that bacterial infections have developed resistance by forming biofilms, and these are almost impossible to treat with conventional methods.[11] Ozone gas is known to quickly destroy biofilms.[12] The activity of ozone against resistant biofilm is especially important in resistant cases where antibiotics have failed.[13] Ozone induces

Corresponding Author: Özgür Ağlamış, ozgurztb@gmail.com

moderate oxidative stress. Moderate oxidative stress activates nuclear factor-erythroid 2-related factor 2 (Nrf2), a nuclear transcriptional factor. Ozone treatment can activate nuclear factor-erythroid 2-related factor 2 through moderate oxidative stress and suppress NF-κB and inflammatory responses. Mild immune responses are triggered through other nuclear transcriptional factors, such as nuclear factor of activated T cells and activated protein-1 (AP-1). Consequently, both free antioxidants and antioxidant enzymes not only protect cells from oxidation and inflammation but also can reverse chronic oxidative stress.[13]

In our study, the aim of this study was to retrospectively analyze the treatment of bacterial vaginosis, which is one of the most common causes of vaginal discharge and has a high recurrence rate even if treated, with intravaginal gas ozone and its long-term recurrence rates.

METHODS

After the approval by the Institutional Review Board of Dr. Özgür Ağlamış Private Clinic, İstanbul, Turkiye (Date: 29.06.2024, Decision No: 12), the records were reviewed the medical records of all patients who underwent surgery. Informed consent was obtained from each patient at admission for research use of her clinical data. The study protocol was approved by the Sancaktepe Şehit Prof. Dr. İlhan Varank Trainig Research Hospital Ethics Committee İstanbul - Turkiye (Date: 28.06.2024, Decision No: 2024/189).

The study was conducted in accordance with the standards of Good Clinical Practice (ICH-E6) and the principles of the Declaration of Helsinki. This retrospective study was conducted at our private clinic of between January 2022 and October 2024. The data were obtained from hospital database system after ethics approval.

This study was designed as a retrospective cross-sectional analysis utilizing pre-existing clinical data from patients diagnosed with bacterial vaginosis using Amsel criteria and treated with intravaginal gas ozone therapy at our clinic between January 2022 and October 2024. No clinical trial, randomization, or investigational drug administration was performed. All procedures were part of routine clinical care, and the analysis focused solely on treatment outcomes, recurrence rates, and safety findings based on digital medical records obtained during standard follow-up visits. Consequently, the study does not qualify as an interventional clinical trial but rather as an observational review of existing data. During this period, 8 patients were excluded from the study because they received intravaginal suppository, oral antibiotic treatment, and fractional carbon dioxide laser treatment at external centers. The remaining 102 patients were evaluated in terms of success, complications and recurrence and were included in the study. Flowchart of patient inclusion was showed at flow **Figure 1**.

In digital data analysis, it was observed that each patient's hemogram was checked before and within one week after the treatment, and the patients were evaluated in terms of hemolysis and pancytopenia. It was learned that all patients included in the study were monogamous and that condoms were recommended after treatment in all their relationships

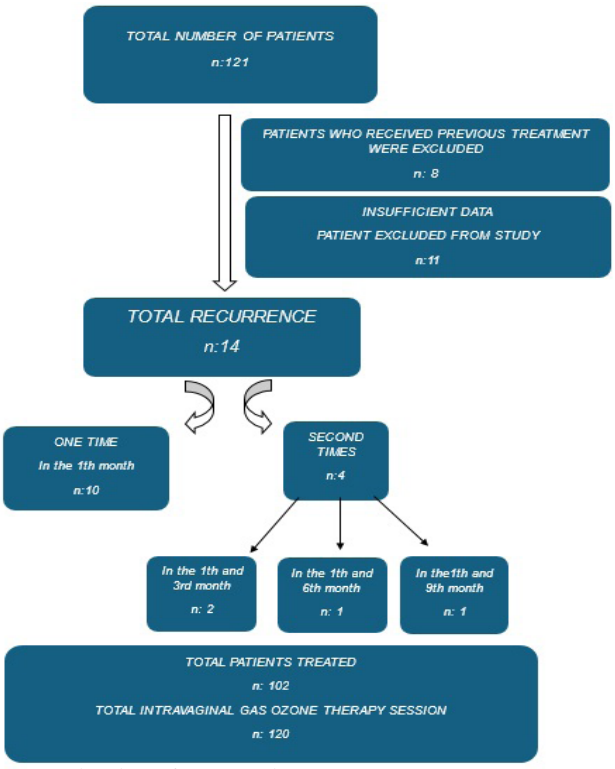

Figure 1. Flowchart of patient inclusion

during the one-year period, and condom use was confirmed verbally at the controls.

It was understood that all vaginal examination findings of the patients, who underwent clinical controls at the 1st month, 3rd month, 6th month, 9th month and 12th month after the treatment, were recorded. Patients were re-evaluated with Amsel criteria at all these visits. In this way, an attempt was made to achieve standardization. Vaginal discharge samples were collected from all patients using sterile swabs before and after ozone therapy. The samples were transported to the microbiology laboratory within 2 hours and cultured on selective media, including Columbia blood agar and *Gardnerella*-selective agar.

Bacterial identification and quantification were performed using standard gram staining and aerobic/anaerobic culture techniques. The presence of *Gardnerella vaginalis (G. vaginalis), Lactobacillus* spp., and other anaerobes was confirmed using MALDI-TOF mass spectrometry.

In case of recurrence, the files of the patients who were found to have received another session of intravaginal ozone therapy with the same gamma dose and amount were examined through the digital recording system. Complete cure, complications and recurrences were proportioned according to the total number of patients.

Intravaginal Gas Ozone Technique

The patients were placed on the lithotomy table. The inside of the vagina was washed twice with 10 cc of SF to prevent vaginal discharge from interfering with the procedure. Infection-related discharge in the vagina was cleaned with sterile sponge. A size 8 nasogastric tube was advanced intravaginally, touching the posterior fossa under the cervix.

Then the catheter was withdrawn 0.5 cm. Sponges were placed at the entrance of the vagina for pressure purposes to prevent gas escape during the procedure. Salutem plus brand medical ozone device was set to 60 gamma ozone dose. 60 gamma ozone gas was drawn into the 50 cc injector. The ozone-filled syringe was attached to the tip of the nasogastric tube and ozone gas was quickly injected into the vagina. This process was repeated four times, with a total of 200 cc of intravaginal gas ozone. After the procedure, the sponges at the entrance to the vagina were waited for five minutes before being removed. Afterwards, the procedure was terminated by removing the nasogastric tube. Patients were observed for one hour in the clinic environment for possible side effects. Early complications or adverse effects were recorded digitally. Figure 1 and 2 show the preparation for intravaginal ozone gas administration and the delivery of ozone gas.

Figure 1. Preparation for intravaginal ozone gas administration

Figure 2. When intravaginal ozone gas is administered

Statistical Analysis

All statistical analyses were performed using SPSS version XX (IBM Corp., Armonk, NY, USA). Continuous variables (e.g., age, BMI, hemogram values) were expressed as mean±standard deviation and compared using paired t-tests to evaluate pre- and post-treatment changes.

RESULTS

A total of 121 patients were included in the study. The mean age was 34.8±4.2 years, and the mean body-mass index (BMI) was 26.5±4.6 kg/m^2. 30.4% (n=31) of the patients were smokers. 35.3% (n=36) of the patients had a history of BV. 15.7% (n=16) of the patients had previously received BV treatment. Demographic and clinic characteristic data of the patients are shown in Table 1. In the retrospective evaluation made for one year after treatment, a total of 18 recurrences (18.36%) were observed in 14 patients. These recurrences, which were observed in 14 patients in total, were observed to recur more than once in only 3.9% (n=4) of the patients, and only once in the other 9.8% of the patients (n=10). Recurrence was observed in all of these four patients in the first month. Recurrence was observed in two of them in the 3rd month, in one in the 6th month, and in one in the 9th month. In these four patients with more than one BV attack, no more than two recurrences were observed. The maximum number of recurrences was two. It was noticed that there was no third recurrence in any patient. It was observed that there was no BV attack in any patient at 12 months. Of the ten patients with only one recurrence, six (60%) were observed to have recurrence in the 1st month, two (20%) in the 3rd month, and 2 (20%) in the 6th month. It was observed that none of the patients who had a BV attack had a recurrence in the 9th and 12th months after retreatment. Following vaginal ozone therapy, a significant reduction in *G. vaginalis* and anaerobic flora was observed, alongside a notable increase in protective *Lactobacillus* colonization. These findings support the microbiological efficacy of ozone therapy in restoring a healthy vaginal microbiota. Data of the patients are shown in Table 2.

Table 1. Demographic and clinic characteristics of patients

Parameters	Patients (n=102)
Age (years), mean±SD	34±5.7
Gravida, median (min-max)	1 (0-4)
Parity, median (min-max)	1 (0-3)
Abortion, median (min-max)	0 (0-2)
Number of live-born children, median (min-max)	1 (0-3)
BMI (kg/m^2), mean±SD	26.5±4.6
Smoking n (%)	31 (30.4)
History of BV n (%)	36 (35.3)
Previous BV treatment n (%)	16 (15.7)
Recurrence	
1 time, n (%)	10 (9.8)
2 times, n (%)	4 (3.9)

BMI: Body-mass index, BV: Bacterial vaginosis, SD: Standard deviation

Table 2. Microbiological evaluation		
	Pre-treatment (%)	Post-treatment (%)
Gardnerella vaginalis presence	78	19
Lactobacillus dominance	22	63
High anaerobic bacterial load	67	18

Ozone therapy led to significant improvement in clinical symptoms of bacterial vaginosis. Most patients reported complete or near-complete resolution of complaints, especially in terms of discharge and odor. Data of the patients are shown in Table 3.

Table 3. Bacterial vaginosis symptoms improvement		
	Pre-treatment (%)	**Post-treatment (%)**
Vaginal discharge	95	9
Fishy odor	82	12
Itching	70	14
Burning sensation	45	11

Categorical variables (e.g., recurrence rates, presence of *G. vaginalis*, *Lactobacillus* colonization, symptom resolution) were analyzed using chi-square tests. Microbiological assessments were based on standard gram staining and aerobic/anaerobic culture techniques performed both before and after ozone therapy; bacterial identification was confirmed by MALDI-TOF mass spectrometry. A p-value of <0.05 was considered statistically significant.

Chi-square tests demonstrated a statistically significant reduction in recurrence rates and Gardnerella presence after ozone therapy ($p<0.01$). Similarly, *Lactobacillus* colonization increased significantly post-treatment ($p<0.01$). Paired t-test analysis of clinical symptoms, including discharge, odor, and itching, also revealed significant improvement ($p<0.01$). These findings confirm that the observed therapeutic benefits are unlikely to be due to chance and support the efficacy of intravaginal ozone therapy in BV management. Data of the patients are shown in Table 4.

A total of 120 sessions of intravaginal gas ozone therapy were performed on 102 patients. No situation causing serious mortality or morbidity was encountered in any patient. No statistically significant change in blood parameters was encountered during hemogram controls. Only minor adverse events were encountered. Mild vaginal irritation 5%, Pelvic discomfort 3%, burning or itching sensation 4%, no reportes side effects 88%. Minor side effects, including mild vaginal irritation (5%), transient pelvic discomfort (3%), and a burning or itching sensation (4%), were reported in 12% of patients. All symptoms were transient, resolved spontaneously within 24-48 hours, and did not require pharmacologic intervention. Patients were observed for one hour after the procedure to ensure immediate tolerance and were given communication instructions for delayed symptom reporting. Follow-up visits were scheduled at 1, 3, 6, 9, and 12 months, and no persistent or serious side effects were noted during this time. Pre- and post-treatment hemogram monitoring confirmed the absence of systemic complications such as hemolysis or pancytopenia.

Recurrence was defined as the re-emergence of bacterial vaginosis meeting at least three of four Amsel criteria (homogeneous discharge, vaginal pH>4.5, positive whiff test, and presence of clue cells) in conjunction with clinical symptoms such as malodorous discharge or vaginal discomfort. When recurrence was suspected clinically, additional confirmation was obtained through gram staining and culture for *G. vaginalis* and *Lactobacillus* species. This combined diagnostic strategy allowed for a standardized and objective evaluation of recurrence during the 12-month follow-up period.

DISCUSSION

Although the success rates in the treatment of bacterial vaginosis are high after medical treatments, high rates of recurrence are also observed.[7] In a study, recurrence rates of up to 58% were observed even one year after treatment.[14] In another study, it was stated that the recurrence rate was close to 50%.[15] The lower recurrence rates in our study and the fact that all recurrences did not extend over a period longer than one year can be considered an advantage of our study. In the study conducted by Cook et al.,[16] the recurrence rate in the first three months after treatment with metronidazole was stated to be 30-40%. This rate is above the total recurrence rate in our study. In addition, while we counted each replaced with episodes of BV one by one in our study, Cook et al. considered at least three or more attacks in the last year as recurrence. Therefore, it can be said that our recurrence rate is lower when compared to our study. Similar to our study, Marshall et al.[17] also stated that one attack is sufficient to qualify as recurrent BV.

In the treatment of BV, it is generally recommended to use 500 mg oral metronidazole twice a day for seven days or use 0.75% metronidazole gel with an intravaginal applicator every night for five days. Another treatment method is to apply 2% clindamycin vaginal cream with an intravaginal applicator for 7 nights.[18] When we compare these standard treatment protocols with our study, the fact that the treatment can be completed in only 15 minutes under clinical conditions can be considered an advantage over other treatments in terms of time. In addition, compared to the problems experienced

Table 4. Chi-square & paired T test			
Parameter	**Test**	**p-value**	**Interpretation**
Recurrence rate (pre vs post ozone therapy)	Chi-square	<0.01	Significant reduction in recurrence after ozone therapy
Gardnerella presence (pre vs post)	Chi-square	<0.01	Significant decrease in *Gardnerella* presence post-therapy
Lactobacillus colonization (pre vs post)	Chi-square	<0.01	Significant increase in *Lactobacillus* colonization post-therapy
Symptom improvement (odor, discharge, itching)	Paired T test	<0.01	Marked improvement in clinical symptoms post-treatment

by patients in adapting to oral medication use or intravaginal applicator use for 7 days, the application of this treatment by a healthcare professional in a short time can be considered as easier for the patient.

According to Sobel et al.,[19] cases of bacterial vaginosis that recur after treatment with metronidazole or clindamycin should be treated more aggressively, especially if the patient has monthly relapses. Accordingly, in recurrent replaced with episodes of BV, intravaginal application of 0.75% metronidazole gel twice a week for 4-6 months is recommended. In this method, the length of the treatment period and the difficulty of using an intravaginal applicator can be seen as a disadvantage compared to our intravaginal gas ozone application, which takes only 15 minutes in case of recurrence.

According to Coundray et al.,[1] since current treatment methods are quite ineffective, it may be necessary to try alternative treatment methods for the treatment of BV. Based on this, intravaginal gas ozone therapy was applied intravaginal gas ozone as an alternative treatment method. When we look at the literature, the use of gas ozone intravaginally has been described before.[20] As far as we can check from the literature, our study is the first to use intravaginal gas ozone in the treatment of BV. This can be seen as an advantage of our study.

Ozone, a highly reactive molecule, exhibits strong bactericidal, fungicidal, antiviral and anti-protozoal activities. Ozone therapy has previously been used as an anti-inflammatory in the treatment of endometritis and successful results have been obtained.[21] In a study, it was stated that the use of ozone as a therapeutic agent in pelvic inflammatory disease inhibited the necrosis of endometrial epithelial cells and also alleviated inflammatory reactions.[22] This study utilized intravaginal ozone gas in the treatment of BV by taking advantage of this superficial cell protective effect and anti-inflammatory effect.

In an animal study conducted with sheep, antibiotic treatment was applied to one group and ozone foam spray treatment was applied to another group for the placenta fragments remaining in the uterus after birth and the inflammatory process caused by this. This study showed that ozone preparations did not cause any negative side effects and were as effective as, but not statistically better than, antibiotics.[23] In an in vitro study, fifty *Candida albicans* strains were exposed to gaseous ozone at different times. Although ozone is highly effective on the yeast form of *Candida albicans*, ozone therapy appears to induce resistance to amphotericin B.[24] This can be considered a disadvantage of ozone therapy.

Serious complications such as decrease in hemoglobin level, hemolysis, and pancytopenia may occur in systematic intravenous ozone use. Using local gas ozone in our treatment reduces our risk of encountering such serious complications. In a systematic review conducted by Mehta,[25] it was stated that male partner treatment was not statistically beneficial in preventing recurrent bacterial vaginosis and did not reduce recurrence rates. In our study, no additional treatment was applied any special treatment to the spouses of patients who presented to us for the first time or who came again with recurrence. However, it was observed that all patients were recommended to use condoms during intercourse during this period. Another study included couples in which a woman with bacterial vaginosis was in a monogamous relationship with her male partner. In the partner-treatment group, the woman received recommended first-line antimicrobial agents and the male partner received oral and topical antimicrobial therapy (metronidazole 400 mg tablets and 2% clindamycin cream applied to the skin of the penis, both twice daily for 7 days). In the control group, the woman received first-line therapy and the male partner received no treatment. The primary outcome was recurrence of bacterial vaginosis within 12 weeks.[26] In our study, we only made an evaluation for female patients. Another study investigated whether dequalinium chloride, a broad-spectrum antiseptic, was more effective than oral metronidazole in treating BV.[27] Just as this treatment is an alternative to conventional metronidazole treatment, we also investigated what could be possible other than conventional methods in our study.[28] There is a study on the positive effect of vaginal microbiota transplantation in the treatment of bacterial vaginosis. We also believe that the positive results we obtained in our ozone study are due to the positive effect of ozone therapy on the microbiota. Alternative therapies such as probiotics have been increasingly explored in the management of bacterial vaginosis. Probiotic formulations containing *Lactobacillus* species aim to restore the vaginal microbiota by increasing lactobacilli dominance and lowering pH levels, thereby indirectly suppressing anaerobic pathogens.[29,30] However, meta-analyses indicate variable efficacy and highlight the limited impact of probiotics on biofilm-associated bacteria, which are central to recurrent BV. In contrast, intravaginal ozone therapy offers direct antimicrobial and biofilm-disrupting effects in addition to immunomodulatory benefits that support mucosal healing.[9] These multimodal effects likely underlie the favorable recurrence rates observed in our study and suggest that ozone therapy may serve as a valuable alternative or complementary approach to probiotics and antibiotics in future BV management strategies.

Cost and accessibility are important considerations when evaluating intravaginal ozone therapy in comparison to traditional treatments such as metronidazole. While metronidazole is inexpensive and widely available in both oral and intravaginal forms, it often requires repeated courses due to high recurrence rates, potentially increasing cumulative treatment costs and contributing to antimicrobial resistance. Ozone therapy, by contrast, involves a single clinical session but requires specialized devices and trained personnel, which may initially limit its accessibility and increase upfront costs. However, the reduced recurrence rates and shorter overall treatment duration observed in our study suggest that ozone therapy may offer long-term cost benefits for patients with recurrent BV. Wider adoption of ozone technology and potential cost reductions with broader use may further enhance its feasibility in clinical practice.

Future studies should aim to optimize the application of intravaginal ozone therapy for bacterial vaginosis. Dose-ranging and frequency studies are needed to determine the most effective and safest protocol, as the ideal gamma concentration and session number remain to be established.

Combining ozone therapy with other modalities, such as probiotics or conventional antibiotics, could provide complementary benefits by simultaneously disrupting biofilms and reestablishing a protective *Lactobacillus*-dominant microbiota. Moreover, prospective randomized controlled trials with larger cohorts are warranted to confirm efficacy, evaluate recurrence rates over longer follow-up periods, and explore cost-effectiveness and patient-centered outcomes. Such investigations will help define the role of ozone therapy in the evolving therapeutic landscape of BV management.

Among the few studies investigating ozone therapy for bacterial vaginosis, Yarustovskaya et al.[32] conducted a randomized controlled trial comparing standard therapy, local intravaginal ozone, and combined local plus systemic ozone therapy with an interferon inducer. While their findings demonstrated significant benefits for combined therapy, our study uniquely focuses on local intravaginal gas ozone therapy alone, providing evidence for its standalone efficacy and safety.

Similarly, Khairy et al.[33] conducted a randomized clinical trial on intravaginal ozone insufflation for recurrent vulvovaginal candidiasis and reported high clinical and microbiological cure rates, further supporting the antimicrobial and mucosal healing potential of ozone therapy, which may also be relevant for bacterial vaginosis management.

In an ultrastructural study, Alia and Kholoud[34] demonstrated that ozone therapy disrupts the biofilm architecture of *G. vaginalis*, which plays a central role in the persistence and recurrence of bacterial vaginosis, thereby supporting the mechanistic rationale for its therapeutic use.

Limitations

The limitations of our study are the limited number of patients participating in the study, although all patients were diagnosed with BV according to the Amsel criteria, other causes of vaginal infection that may accompany it are not known, and most BV conditions resolve spontaneously over time without requiring treatment, the absence of a control group, and the retrospective design of the study. It may be a disadvantage that our study is not a randomized controlled study comparing conventional methods. This study has several limitations that must be acknowledged. First, the retrospective design inherently carries a risk of selection bias and reliance on pre-existing medical records. Second, the absence of a randomized control group receiving conventional antibiotic therapy limits direct comparisons of treatment efficacy. Third, the single-center nature of the study and relatively small sample size may reduce the generalizability of the results to broader populations. Additionally, potential confounding factors, such as variations in sexual behavior or hormonal status, were not fully assessed. Nevertheless, the use of standardized Amsel criteria for diagnosis, uniform application of the ozone protocol, and regular follow-up visits at 1, 3, 6, 9, and 12 months help to mitigate these biases and provide valuable preliminary data supporting ozone therapy as an alternative treatment for bacterial vaginosis. Our primary aim here was to show the effectiveness of intravaginal ozone therapy in treatment and symptom reduction. Transvaginal ozone treatment does not kill microorganisms that may be located outside the vagina (skin, vestibular glands). This may be the cause of recurrence. In addition, the need to have an ozone device to apply this treatment, the need for a healthcare professional who knows how to use the device, and the cost of the procedure can be considered as limitations of this method for the patient and the healthcare professional. The strengths of the study are that the risk of resistance is reduced since antibiotics are not used and that, as previously stated in the literature, the remodeling effect of ozone is utilized in the treatment.

CONCLUSION

The use of intravaginal gas ozone in the treatment of bacterial vaginosis is a minimally invasive procedure. The fact that we have not seen any major complications, that it is easily applicable, that the treatment success rates are high, and that recurrences are low can be seen as the success of the technique. These findings suggest that with the increase in the use of intravaginal gas ozone, the area of use and the number and frequency of applications in BV and other vaginal infections will increase. Further investigation with a larger number of patients should be conducted to confirm our data.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was carried out with the permission of the Sancaktepe Şehit Prof. Dr. İlhan Varank Trainig Research Hospital Ethics Committee (Date: 28.06.2024, Decision No: 2024/189).

Informed Consent

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

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

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

Author Contributions

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

Availability of data and material

The datasets of the current study are available upon reasonable request.

REFERENCES

1. Coudray MS, Madhivanan P. Bacterial vaginosis-a brief synopsis of the literature. *J Obstet Gynecol Reprod Biol.* 2020;245:143-148. doi:10.1016/j.ejogrb.2019.12.035
2. Gardner HL, Dukes CD. Haemophilus vaginalis vaginitis: a newly defined specific infection previously classified "nonspecific" vaginitis. *Am J Obstet Gynecol.* 1955;69(5):962-976.

3. Ranjit E, Raghubanshi BR, Maskey S, Parajuli P. Prevalence of bacterial vaginosis and ıts association with risk factors among nonpregnant women: a hospital based study. *Int J Microbiol*. 2018;2018:8349601. doi: 10.1155/2018/8349601
4. Martin DH, Marrazzo JM. The vaginal microbiome: current understanding and future directions. *J Infect Dis*. 2016;214(Suppl 1): S36-S41. doi:10.1093/infdis/jiw184
5. Chen JY, Tian H, Beigi RH. Treatment considerations for bacterial vaginosis and the risk of recurrence. *J Womens Health (Larchmt)*. 2009; 18(12):1997-2004. doi:10.1089/jwh.2008.1088
6. Koumans EH, Sternberg M, Bruce C, et al. The prevalence of bacterial vaginosis in the United States, 2001-2004; associations with symptoms, sexual behaviors, and reproductive health. *Sex Transm Dis*. 2007;34(11):864-869. doi:10.1097/OLQ.0b013e318074e565
7. Bradshaw CS, Brotman RM. Making inroads into improving treatment of bacterial vaginosis-striving for long-term cure. *BMC Infect Dis*. 2015; 15:1-12. doi:10.1186/s12879-015-1027-4
8. Zanardi I, Burgassi S, Paccagnini E, Gentile M, Bocci V, Travagli V. What is the best strategy for enhancing the effects of topically applied ozonated oils in cutaneous infections? *Biomed Res Int*. 2013;2013:702949. doi:10.1155/2013/702949
9. Ergani SY, Dilbaz B, Ergani HM, Tekin ÖM. Effect of intrauterine ozone therapy on Asherman syndrome, an experimental rat model. *Eur J Obstet Gynecol Reprod Biol*. 2022;277:90-96. doi:10.1016/j.ejogrb.2022.08.014
10. Sen S, Sen S. Ozone therapy a new vista in dentistry: integrated review. *Med Gas Res*. 2020;10(4):189-192. doi:10.4103/2045-9912.304226
11. Bjarnsholt T. The role of bacterial biofilms in chronic infections. *APMIS Suppl*. 2013;(136):1-51. doi:10.1111/apm.12099
12. Bialoszewski D, Pietruczuk-Padzik A, Kalicinska A, et al. Activity of ozonated water and ozone against *Staphylococcus aureus* and *Pseudomonas aeruginosa* biofilms. *Med Sci Monit*. 2011;17(11):BR339-BR344. doi:10.12659/msm.882044
13. Sagai M, Bocci V. Mechanisms of action ınvolved in ozone therapy: is healing induced via a mild oxidative stress? *Med Gas Res*. 2011;1:29. doi:10.1186/2045-9912-1-29
14. Rowen RJ. Ozone therapy as a primary and sole treatment for acute bacterial infection: case report. *Med Gas Res*. 2018;8(3):121-124. doi:10.4103/2045-9912.241078
15. Bradshaw CS, Morton AN, Hocking J, et al. High recurrence rates of bacterial vaginosis over the course of 12 months after oral metronidazole therapy and factors associated with recurrence. *J Infect Dis*. 2006;193(11):1478-1486. doi:10.1086/503780
16. Bradshaw CS, Vodstrcil LA, Hocking JS, et al. Recurrence of bacterial vaginosis is significantly associated with posttreatment sexual activities and hormonal contraceptive use. *Clin Infect Dis*. 2013;56(6):777-786. doi:10.1093/cid/cis1030
17. Cook RL, Redondo-Lopez V, Schmitt C, Meriwether C, Sobel JD. Clinical, microbiological, and biochemical factors in recurrent bacterial vaginosis. *J Clin Microbiol*. 1992;30(4):870-877. doi:10.1128/jcm.30.4.870-877.1992
18. Marshall AO. Managing recurrent bacterial vaginosis: insights for busy providers. *Sex Med Rev*. 2015;3(2):88-92. doi:10.1002/smrj.45
19. Bagnall P, Rizzolo D. Bacterial vaginosis: a practical review. *JAAPA*. 2017;30(12):15-21. doi:10.1097/01.JAA.0000526770.60197.fa
20. Sobel JD, Ferris D, Schwebke J, et al. Suppressive antibacterial therapy with 0.75% metronidazole vaginal gel to prevent recurrent bacterial vaginosis. *Am J Obstet Gynecol*. 2006;194(5):1283-1289. doi:10.1016/j.ajog.2005.11.041
21. Mandhare M, Jagdale D, Gaikwad P, Gandhi P, Kadam V. Miracle of ozone therapy as an alternative medicine. *Int J Pharm Chem Biol Sci*. 2012;2(1):63-71.
22. Đuričić D, Valpotić H, Samardžija M. Prophylaxis and therapeutic potential of ozone in buiatrics: current knowledge. *Anim Reprod Sci*. 2015;159:1-7. doi:10.1016/j.anireprosci.2015.05.017
23. Wei A, Feng H, Jia XM, Tang H, Liao YY, Li BR. Ozone therapy ameliorates inflammation and endometrial injury in rats with pelvic inflammatory disease. *Biomed Pharmacother*. 2018;107:1418-1425. doi:10.1016/j.biopha.2018.07.137
24. Đuričić D, Valpotić H, Žura Žaja I, Samardžija M. Comparison of intrauterine antibiotics versus ozone medical use in sheep with retained placenta and following obstetric assistance. *Reprod Domest Anim*. 2016; 51(4):538-540. doi:10.1111/rda.12715
25. Zargaran M, Fatahinia M, Zarei Mahmoudabadi A. The efficacy of gaseous ozone against different forms of *Candida albicans*. *Curr Med Mycol*. 2017;3(2):26-32. doi:10.18869/acadpub.cmm.3.2.26
26. Mehta SD. Systematic review of randomized trials of treatment of male sexual partners for improved bacteria vaginosis outcomes in women. *Sex Transm Dis*. 2012;39(10):822-830. doi:10.1097/OLQ.0b013e3182631d89
27. Vodstrcil LA, Plummer EL, Fairley CK, et al. Male-Partner treatment to prevent recurrence of bacterial vaginosis. *N Engl J Med*. 2025;392(10):947-957. doi:10.1056/NEJMoa2405404
28. Raba G, Durkech A, Malík T, Bassfeld D, Grob P, Hurtado-Chong A; Fluomizin Study Group. Efficacy of dequalinium chloride vs metronidazole for the treatment of bacterial vaginosis: a randomized clinical trial. *JAMA Netw Open*. 2024;7(5):e248661. doi:10.1001/jamanetworkopen.2024.8661
29. Luo H, Zhou C, Zhou L, He Y, Xie RH. The effectiveness of vaginal microbiota transplantation for vaginal dysbiosis and bacterial vaginosis: a scoping review. *Arch Gynecol Obstet*. 2024;310(2):643-653. doi:10.1007/s00404-024-07611-1
30. Van de Wijgert JHHM, Jespers V. The role of lactobacilli in vaginal health and bacterial vaginosis: evidence from clinical studies. *FEMS Microbiol Rev*. 2020;44(6):773-792.
31. Machado D, Castro J, Palmeira-de-Oliveira A, Martinez-de-Oliveira J, Cerca N. Bacterial vaginosis biofilms: challenges to current therapies and emerging solutions. *Front Microbiol*. 2016;7:1528. doi:10.3389/fmicb.2015.01528
32. Yarustovskaya OV, Kulikov AG, Shtro LP. Combined treatment of patients with bacterial vaginosis using ozone therapy and an interferon inducer. *Vopr Kurortol Fizioter Lech Fiz Kult*. 2015;92(3):17-21. doi:10.17116/kurort2015545-49
33. Khairy AM, Abdelkhalek AM, Elsharkawy M, et al. Intravaginal ozone insufflation for recurrent vulvovaginal candidiasis: a randomized clinical trial. *Egypt Bullet Women Health J*. 2021;11(3):45-53.
34. Alia A, Kholoud R. The impact of ozone on bacterial vaginosis biofilms: ultrastructural evidence via transmission electron microscopy. *EJ Med*. 2014;6(2):45-52.

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1719842
J Health Sci Med. 2025;8(5):818-824

Assessment of the abdominal aorta regarding its morphology, morphometry, and concomitant pathologies associated with abdominal aorta aneurysm

Hürriyet Çetinok[1,2]

[1]Department of Anatomy, Faculty of Medicine, İstanbul Atlas University, İstanbul, Turkiye
[2]Department of Anatomy and Structural Biology, Albert Einstein College of Medicine, Yeshiva University, New York, USA

Cite this article as: Çetinok H. Assessment of the abdominal aorta regarding its morphology, morphometry, and concomitant pathologies associated with abdominal aorta aneurysm. *J Health Sci Med.* 2025;8(5):818-824.

Received: 05.07.2025 ◆ **Accepted**: 04.08.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: The abdominal aorta (AA) plays a crucial role in both invasive and non-invasive radiological procedures, particularly in hepatorenal and colorectal surgeries. In this study, the abdominal aorta was examined morphologically, and the presence of aneurysms and accompanying pathologies was discussed. Anatomical structures that can guide surgeons and radiologists were defined as landmarks, and a morphometric approach was made.

Methods: This study examined the AA of 20 formalin-embalmed American cadavers (ages 50-96) donated to Albert Einstein College of Medicine's C&DA Department. From supine cadavers, aortic bifurcation (BA) levels were examined for the vertebral column. Linear, longitudinal, and transverse distances were measured between the Left renal artery (LtRA) and branching points of BA, diameters of AA, tortuosity, and morphology were examined.

Results: Four of the 20 cadavers included in the study had AA aneurysm (AAA). Cadaveric examinations showed aortic dilatation, abdominal aorta wall thickness, atherosclerosis, thrombus formation at the Superior Mesenteric Artery (SMA), an abdominal tortuous aorta, L3 vertebrae deviation, multiple AAA, hemivertebrae, L3 compression fractures, and osteodegenerative changes. The average AA transverse diameter is 22.93±2.69 mm. Upon assessment of the correlation between advancing age and the incidence of AAA in the male population, no statistically significant relationship was found (p=0.167).

Conclusion: Although the exact role of atherosclerosis in the development of AAA remains unclear, it may contribute to their occurrence; studies with larger cohorts are needed to better understand their prevalence and associated anatomical changes.

Keywords: Abdominal aorta aneurysm, atherosclerosis, abdominal aorta tortuosity, abdominal aorta, lumbar deviation, hemivertebrae

INTRODUCTION

The abdominal aorta (AA) is situated anterior to the inferior border of the 12th thoracic vertebra and descends anterior to the vertebral column, terminating at the level of the 4th lumbar vertebra (LV), typically to the left of the midline by bifurcating into the two common iliac arteries. Aneurysms typically exhibit localized dilation accompanied by wall thinning of the vessel. Normally, the abdominal aorta measures approximately 2-3 cm in diameter, but these dimensions may vary in the presence of vascular pathologies. Conditions such as abdominal aortic aneurysm (AAA) are characterized by progressive enlargement and structural weakening of the vessel wall. AAA is characterized by abnormal dilation of the abdominal aorta and may displace adjacent anatomical structures such as the inferior vena cava and the third part of the duodenum.

Globally, aortic aneurysms (including both thoracic and abdominal) resulted in approximately 167,249 deaths in 2017-equivalent to an age-standardized death rate of 2.19 per 100,000 persons-placing them among the leading 15 causes of cardiovascular mortality worldwide.[1] The lifetime prevalence of abdominal aortic aneurysm varies according to ethnicity, age, and other factors. The projected incidence ranges from 1.0% to 2.2% in females and from 1.3% to 8.9% in males.[2]

The aorta gradually tapers from the thoracic region toward the aortic bifurcation. With advancing age, the diameter of the abdominal aorta tends to increase, particularly in males. Lederle et al.[3] demonstrated that infrarenal aortic diameter increases with age, is larger in males, and correlates with body size. These factors should be considered when assessing AAA. AAA is strongly associated with advanced age, male

Corresponding Author: Hürriyet Çetinok, hurriyet.cetinok@atlas.edu.tr

sex, and especially smoking-current smokers face nearly a fivefold higher risk of AAA compared to never-smokers, while former smokers have approximately a twofold increased risk. Furthermore, each additional year smoking increases AAA risk by around 4%.[4,5]

Metabolic disorders, such as diabetes and dyslipidemia, are known to affect vascular structure and function, contributing to atherosclerosis and arterial wall remodeling. Although diabetes is a well-known cardiovascular risk factor, evidence indicates an inverse relationship between diabetes and the occurrence and progression of AAA.[6] Although both AAA and arterial tortuosity syndrome (ATS) affect the arterial wall, their underlying histopathological features differ significantly. In ATS, histology typically reveals defects in connective tissue architecture due to mutations in extracellular matrix-related genes, leading to arterial elongation and tortuosity. In contrast, AAA is primarily characterized by chronic inflammation, degradation of elastin and collagen, and protease-mediated destruction of the medial layer.[7] These distinct mechanisms underscore the differing pathogenesis of these two vascular conditions. Peripheral vascular disease and coronary heart disease (CHD), two atherosclerosis-linked diseases, are strong AAA predisposing variables, according to population-based research.[8] Solberg et al.[9] mentioned that numerous epidemiological studies have linked CVRF to AAA. However, few studies have examined the interaction between subclinical ATS and CVRF and the abdominal aortic diameter (AAD) and enlarged AAD (EAAD) 25 mm, which are connected to AAA risk.

A tortuous abdominal aorta (ATA) can also cause inferior vena cava suppression. This abnormality may cause necrosis, centrilobular congestion, liver cirrhosis and fibrosis.[10] ATA can potentially inhibit the inferior vena cava. The AA plays a key role in various procedures. It may predict complications in transfemoral transcatheter aortic valve replacement (TAVR) and serves as a critical marker for AAA repair. Additionally, aortic tortuosity should be considered, as it can impact procedural success in both settings.[11]

Focal stenosis of aortic and/or pulmonary arteries with extensive elongation and tortuosity of the aorta and mid-sized arteries are signs of arterial tortuosity syndrome (ATS).[12] Soft or doughy hyperextensible skin, joint hypermobility, inguinal hernia, and diaphragmatic hernia may indicate a widespread connective tissue problem. Skeletal anomalies include pectus excavatum or carinatum, arachnodactyly, scoliosis, knee/elbow contractures, and camptodactyly.[13]

The cardiovascular system remains a leading contributor to morbidity and mortality across all age groups. Aneurysm formation and dissection, particularly at the aortic root and along the arterial tree, are major concerns. Additionally, ischemic events involving the cerebrovascular and abdominal arterial circulation, such as non-hemorrhagic strokes, further highlight the systemic impact of vascular pathology.

Invasive and noninvasive radiological procedures both rely on the abdominal aorta and abdominal surgeries, especially those involving the hepatorenal and colorectal areas. Aneurysms and related diseases were reviewed after morphological examinations of the abdominal aorta and adjacent structures were performed in this study. A morphometric technique was used to define landmarks to guide radiologists and surgeons during abdominal procedures.

METHODS

A unique combination of two reliable landmarks was tested in this cadaveric study to accurately identify the BA (aortic bifurcation) in three-dimensional coordinates. These landmarks were listed: Vertebral body, Left renal artery (LtRA). This study included anatomical analysis of 20 formalin-embalmed cadavers provided to the Department of Anatomy at Albert Einstein College of Medicine (AECOM), focusing on 20 cases (12 males, 8 females, aged 50-96). All donors authorized for donation and use in clinical trials were accepted in compliance with New York's Anatomical Gift Law. This project is exempt from ethical approval as it employs course cadavers from the Albert Einstein College of Medicine C&DA Department, consistent with the exemption categories specified in Einstein-IRB-citation104(d). A 52-year-old female participant was excluded from the experiment due to metastatic stomach cancer in the para-aortic area. The others have no history of previous abdominal injuries, pathological diseases, or surgical procedures. The AA of the supine cadavers was accessible anteriorly, focusing on the aortic branches.

We used a Mitutoyo Digital Caliper (approved by Mitutoyo America Corp. Calibration Lab.-control number 887014, range 0.000 inch-0006 inch), goniometer, dissection instruments, and an Olympus digital camera for our research. We first analyzed the trajectory of the abdominal aorta, the vertebral column branching levels for BA, morphological changes, tortuosity, aneurysms, wall thickness, variations, osteodegenerative changes, and compression fractures. We grouped vertebral bodies by height: upper, medium, and bottom. However, some AA branches developed from the intervertebral disc.

Coeliac trunk (CT), Superior mesenteric artery (SMA), Inferior mesenteric artery (IMA), LtRA, BA, and their transverse, longitudinal, and linear distances, abdominal aorta diameter were evaluated (**Figure 1**). Statistical analysis was also performed on artery sources and branching angles. We removed the abdominal aorta from the cadaver and vertically dissected it along its major axis from the right and left common iliac arteries to split it into ventral and dorsal segments. Digital photos of the resected ventral surfaces of the AA were taken from the intravascular lumen. The study examined BA morphology, tortuosity, and BA branching locations. Linear(e1), longitudinal(e2), and transverse distances between LtRA and the BA were measured. The BA's transverse diameter at the LtRA level (a1) was also measured. Compare the findings with prior research in the literature.

Statistical Analysis

Data analysis of the data was performed in the SPSS 26.0 program (SPSS, 26.0, Chicago, IL.). The Kolmogorov-Smirnov test validated that the data had a normal distribution ($p<0.05$), while the Levene test results indicated that the

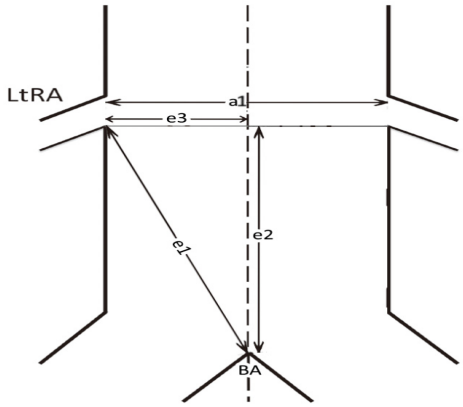

Figure 1. Schematic diagram of measured items of the abdominal aorta at the Lt. RA level; e1 (LtRA to BA linear); e2 (Lt. RA to BA longitudinal); e3 (LtRA to BA transverse distance); a1 (BA's transverse diameter at the LtRA level)

LtRA: Left renal artery, BA: Aortic bifurcation, IVD: Intervertebral disc

variances among groups were equal ($p<0.05$). Given that both assumptions presented in the study were satisfied, the t-test for independent samples was employed for group comparisons, while Pearson correlation analysis was utilized to assess the linear relationship between variables.

RESULTS

The AAin the 19 cadavers examined was divided into the right and left common iliac arteries, facilitating the assessment of BA development levels in relation to the vertebral column, particularly specific vertebral levels. In summary, as depicted in Figure 2, 32% of the BA is located at' the intervertebral disc (IVD) level between the 4th and 5th lumbar vertebrae, 21% from the upper third of the 4th lumbar vertebra, 21% from the middle third of the 4th lumbar vertebra, 16% from the lower third of the 4th lumbar vertebra, 5% from the IVD level between the 3rd and 4th lumbar vertebrae, and 5% from the middle third of the 5th lumbar vertebra. These percentages take into account the age and gender distribution of the cadavers.

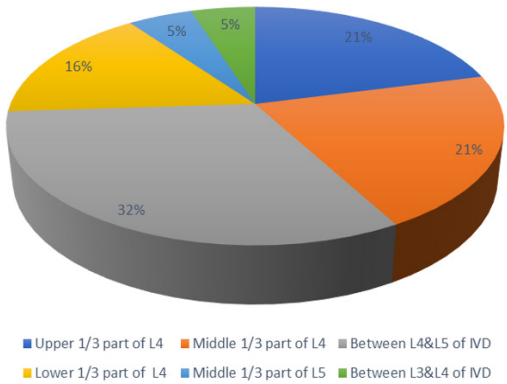

Figure 2. Branching point of the aortic bifurcation from the vertebral column

The transverse diameters of the abdominal aorta within the lumen were evaluated at the bifurcation of the left renal artery from the abdominal aorta. The mean transverse diameter of the abdominal aorta, derived from 19 measurements, is 22.93±2.69 mm.

The measurements of linear, longitudinal, and transverse distances between LtRA and BA were performed accordingly. (e1, e2, e3). The mean linear distance(e1) is 92.67±25.45 mm. The mean longitudinal distance(e2) is 96.24±15.80 mm. The mean transverse distance(e3) between LtRA and BA is 9.08±3.06 mm (Table 1).

Table 1. Descriptive statistics results for males and females

Variable	Males (n=12) M (SD)	Females (n=7) M (SD)	Total (n=19) M (SD)
Age	80.92±10.69	74.71±12.41	78.63±11.43
a1	23.40±2.74	22.10±2.59	22.93±2.69
e1	91.80±28.41	94.18±21.4	92.67±25.45
e2	98.06±12.45	93.11±21.13	96.24±15.80
e3	8.29±3.06	10.43±2.77	9.08±3.06

e1 (Lt. RA to BA linear); e2 (Lt. RA to BA longitudinal); e3 (Lt. RA to BA transverse distance); a1 (BA's transverse diameter of the abdominal aorta at the LtRA level), LtRA: Left renal artery, BA: Aortic bifurcation

91-year-old male (Case 1) had cadaveric examination, which revealed aortic dilatation and aneurysm wall thickness. At the level of the left renal artery, the transverse diameter of the abdominal aorta was 29.04 mm. The aortic bifurcation was located at the middle third of the fifth lumbar vertebra (Table 2).

79-year-old female (Case 2) had an AA aneurysm, atherosclerosis, and SMA thrombus. Aortic bifurcation occurred around the middle third of the fourth lumbar vertebra, with an 18.54 mm transverse abdominal aorta diameter at the left renal artery level (Table 2).

A double right renal artery was found in an 80-year-old man (Case 3). Figure 3, and 4 show AA aneurysm, ATA, and L3 vertebrae deviation in addition to this anatomic variance. The transverse abdominal aorta at the left renal artery level was 24.08 mm wide and bifurcated at the lower third of the fourth lumbar vertebra (Table 2).

Male, 88 years old (Case 4), the cadaveric examination revealed several abdominal aortic aneurysms, hemivertebrae, an L3 compression fracture, increased intervertebral body height at L2-L3, and osteodegenerative alterations (Figure 5, and 6). The aortic bifurcation was situated in the upper third of the fourth lumbar vertebra, with the diameter of the transverse abdominal aorta at the left renal artery level, measuring 26.83 mm (Table 2).

Table 2. Descriptive measures related abdominal aorta in four cases

Cases	Age/gender	a1	e1	e2	e3	Branching point of BA
Case I	91, M	29.04 mm	115.95 mm	114.0 mm	8.7 mm	L5M1/3
Case II	79, F	18.54 mm	61.46 mm	60.77 mm	8.63 mm	L4 M1/3
Case III	80, M	24.08 mm	105.79 mm	102.91mm	93 mm	L4 L1/3
Case IV	88, M	26.83 mm	10.63 mm	101.59 mm	11.61mm	L4 U1/3

e1 (Lt. RA to BA linear); e2 (Lt. RA to BA longitudinal); e3 (Lt. RA to BA transverse distance); a1 (BA's transverse diameter of the abdominal aorta at the LtRA level), LtRA: Left renal artery, BA: Aortic bifurcation, U1/3: Upper third, M1/3: Middle third, L1/3: Lower 1/3

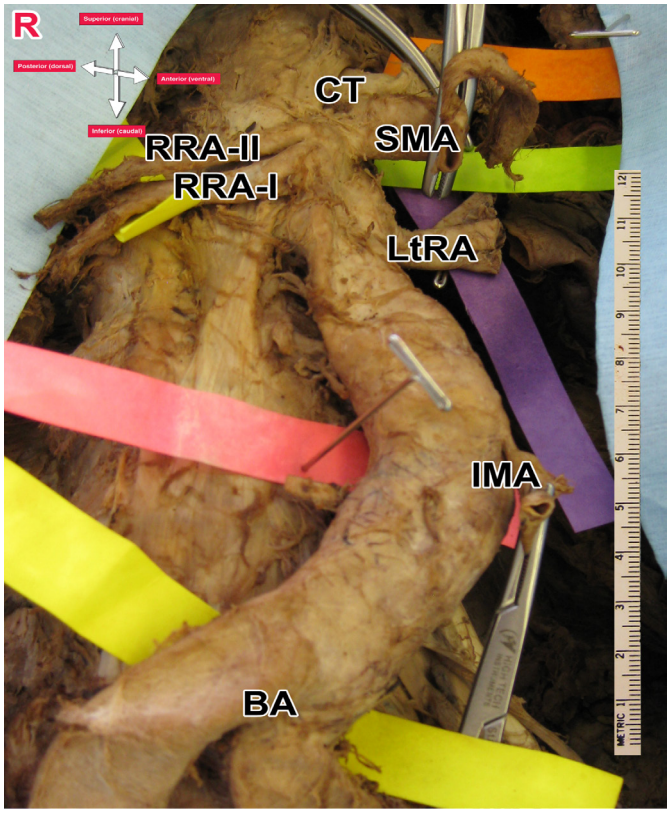

Figure 3. The aspect of the superior side of the supine-positioned cadaver. In addition to double renal arteries, AA aneurysm (*), ATA, and L3 vertebrae deviation were also found.

AA: Abdominal aorta, ATA: Tortuous abdominal aorta, LtRA: Left renal artery, BA: Aortic bifurcation, CT: Coeliac trunk, SMA: Superior mesenteric artery, IMA: Inferior mesenteric artery

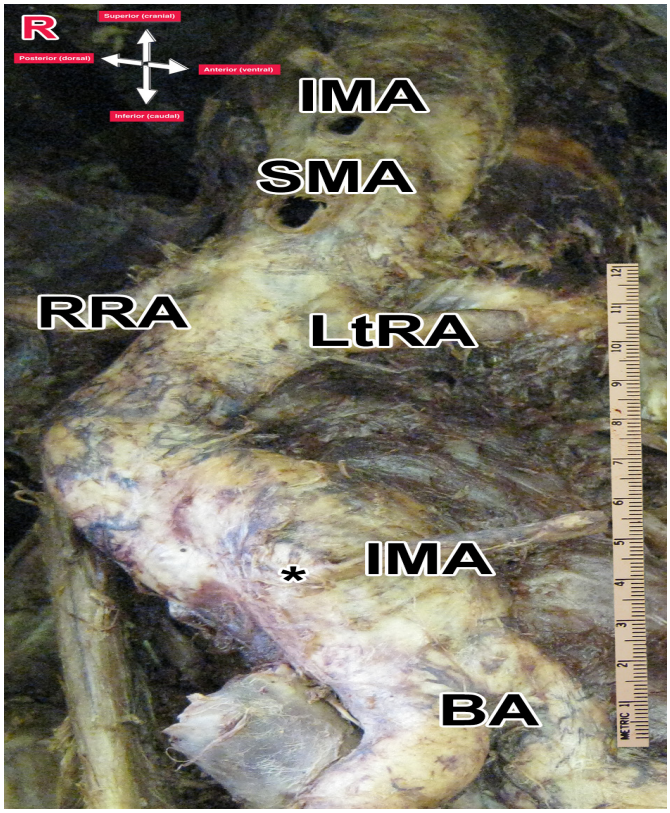

Figure 5. AA with branches related to this study, removed from the cadaver to take measurements. As a result of this cadaveric examination, multiple AA aneurysms, hemivertebrae, ATA, L3 compressor fracture, increased intervertebral body height (L2-L3), and osteodegenerative changes were described

AA: Abdominal aorta, LtRA: Left renal artery, ATA: Tortuous abdominal aorta, BA: Aortic bifurcation, CT: Coeliac trunk, SMA: Superior mesenteric artery, IMA: Inferior mesenteric artery, RRA: Right renal artery

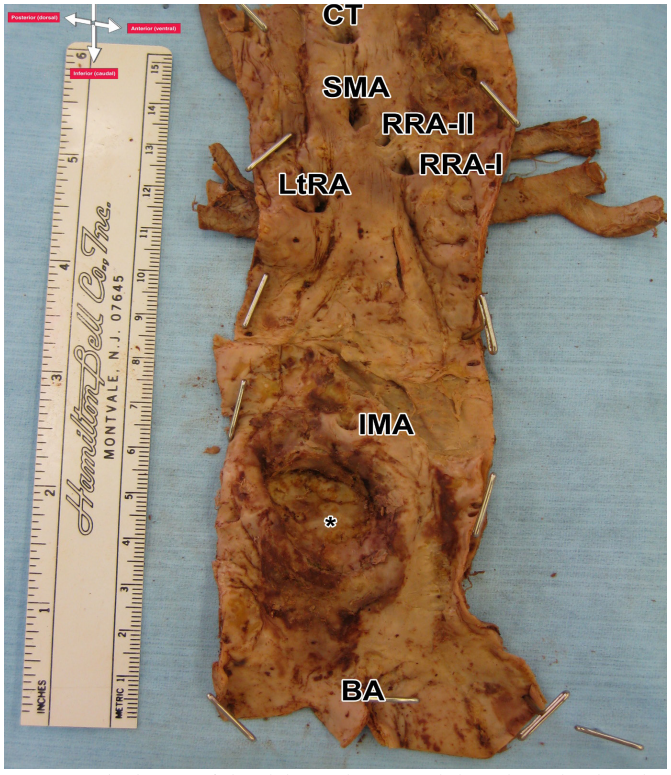

Figure 4. The lumen of the abdominal aorta with branches related to this study was dissected ventrally from the line that crosses the BA. Two openings of renal arteries (RRA-I, RRA-II) and the abdominal aorta aneurysm (*) are clearly

LtRA: Left renal artery, BA: Aortic bifurcation, CT: Coeliac trunk, SMA: Superior mesenteric artery, IMA: Inferior mesenteric artery, RRA: Right renal artery

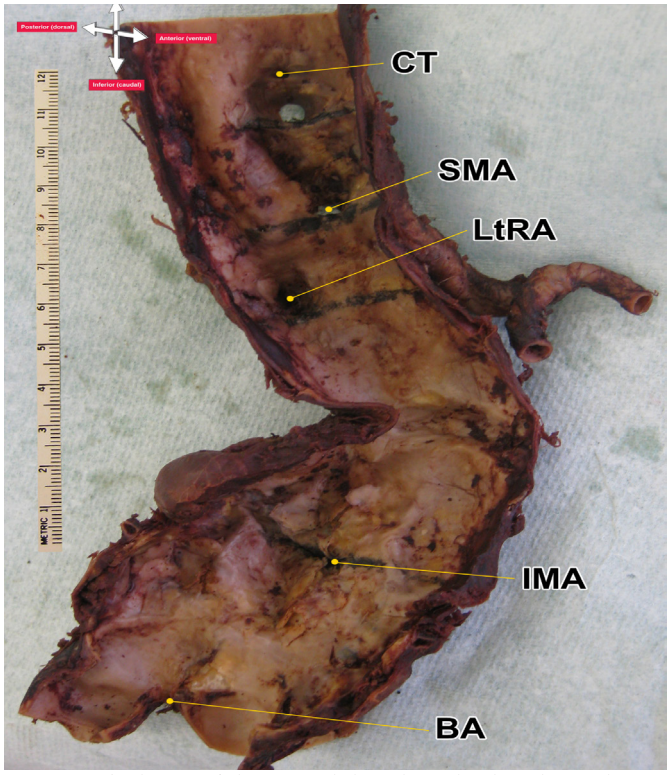

Figure 6. The lumen of the AA with branches related to this study, was dissected ventrally from the line that crosses the Aortic bifurcation. In addition findings from this case, ATA, Multiple AA aneurysms, and atherosclerotic changes were described

AA: Abdominal aorta, LtRA: Left renal artery, ATA: Tortuous abdominal aorta, BA: Aortic bifurcation, CT: Coeliac trunk, SMA: Superior mesenteric artery, IMA: Inferior mesenteric artery, RRA: Right renal artery

In a statistical analysis of the male population, no significant link was identified between the incidence of AAA and advancing age (p=0.167>0.05).

There is no statistically significant correlation between advancing age and aortic diameter in males (p=0.138>0.05).

DISCUSSION

The primary finding of this study is that in 32% of the cadavers, the aortic bifurcation was located at the intervertebral disc (IVD) level between the 4th and 5th lumbar vertebrae, indicating this level as the most common site of bifurcation. In the 19 cadavers examined, the abdominal aorta was observed to divide into the right and left common iliac arteries at variable levels along the lower lumbar spine. Specifically, 21% bifurcated at the upper third of L4, 21% at the middle third of L4, 16% at the lower third of L4, 5% at the IVD between L3 and L4, and 5% at the middle third of L5 (**Figure 2**). These findings provide updated anatomical reference data on aortic bifurcation levels and support prior studies reporting L4 as the most common site, while also highlighting inter-individual variability potentially influenced by age, sex, and degenerative changes. These results align closely with large cadaveric and imaging studies in the literature, which consistently report L4 as the most common bifurcation level-such as the systematic review by Greek investigators identifying L4 body in 42.2% of 3537 specimens,[14] and a Thai cadaver study finding L4 body bifurcation in 70.1% of 187 specimens.[15] MRI studies similarly report L4 in about two-thirds of cases (e.g., 67% in a 441-patient series).[16] Compared to these benchmarks, our 32% at the L4-L5 disc reflects the variability evident across populations and supports the broad range-from upper L3 to upper S1-reported in anatomo-radiological studies.[14] Finally, we acknowledged potential contributors to variability in bifurcation level in our sample-including age-related vascular dilation, aneurysms, degenerative spinal pathology, and lumbosacral transitional vertebrae-and noted the limitations related to sample size and heterogeneity.

Powell et al.[17] claim that hereditary defects in proteolytic enzymes cause AAA, implying that atherosclerosis may be coincidental. Our goal was to understand the association between atherosclerosis and aortic enlargement, particularly in the abdominal region; thus, we eliminated aortas with visible aneurysms. The examination supports that atherosclerotic arterial wall degradation causes aneurysms, notably in the. Atherosclerosis may not cause AAA, but it is undoubtedly a major component.

The Society for Vascular Surgery and the International Society for Cardiovascular Surgery defined abdominal AAA as an infrarenal segment, with a reported mean diameter of 16.6-21.6 mmfor females and 19.9-23.9 mm for males, as determined by computed tomography and intravenous arteriography.[18] Conversely, certain researchers have defined AAA as an infrarenal aortic diameter of 30 mm.[19] There is still no consensus on the standard length of aortic diameter or other mathematical methods to characterize a AAA; however, Hirsch et al.[19] recommended 30.0 mm. More importantly, postmortem investigations show that AAAs with lower diameters rupture, whereas bigger AAAs do not.[12] This suggests that even smaller aorta diameters and expansion rates may be therapeutically relevant.

Our findings indicate a possible association between medial degeneration, aortic size, and atherosclerosis, although this relationship was not statistically significant. The strongest correlation was observed between AA plaque production and wall erosion and significant media microarchitecture degeneration. These data revealed that AA alterations may predispose to aneurysm formation. Plaque composition and progression affect aortic size.[12] Due to their larger necrotic cores, abdominal aortic plaques rupture or ulcerate. Plaque-induced artery wall weakening and atrophy define atherosclerosis. This thinning and the basic bulging beneath the plaques may produce atherosclerosis-related enlargement. Thus, this method can maintain a wide lumen for long periods. This compensatory expansion may avoid or delay AA stenosis or produce aneurysms. Much more research is needed on how risk variables affect aortic plaque progression.[20]

Although not statistically significant, the observed trends may support the hypothesis that atherosclerotic wall changes play a role in abdominal aortic aneurysm development. Atherosclerosis may not cause all AAA, but it is likely to play a key role. Few articles have described the abdominal tortuous aorta.[21-23] There is no literature on the safety of complications and abdominal surgery in this condition. An ATA predicted problems in transfemoral transcatheter aortic valve replacement patients. Kinnel et al.[11] found that non-ruptured AAA patients had more aortic tortuosity than ruptured AAA patients of equal aneurysm size, suggesting that it may lessen rupture risk, and the abdominal aorta's lateral displacement can be misconstrued as an aneurysm when palpated through the abdominal wall as a pulsatile mass, according to Feller and Woodburne.[24]

The breadth and degree of abdominal aorta tortuosity can cause aortic lumen obstruction, hypertension, discomfort, and insufficiency.

A complete preoperative radiological assessment and skills are needed for surgical and endovascular interventions to be successful and safe.

Multiple renal arteries are associated with abdominal aorta tortuosity and aneurysm, according to Cetinok.[25]

Since the LtRA was the reference point in prior studies on the AA and its branches, the catheter often lodges there first. Catheter implantation is easier when LtRA and vascular distances are estimated.[26,27] Considering LtRA changes, we estimated BA's position on the AA relative to the columna vertebralis in this study.

Takahashi et al.[27] found that the lateral diameter of the aorta at the inferior border of the left renal artery was 20.1±2.9 (14.7-25.1) mm, as ascertained from intravascular observations. Sonesson et al.[28] indicated that the internal diameter of the aorta at the LtRA level, as assessed by ultrasonography, was 20.4±2.4 mm in a 70-year-old male and 17.3±2.0 mm in a female. Cauldwell and Anson[29] measured a distance of 100.0±13.6 mm from the outer surface of the blood vessels to the middle of the LtRA. Many articles solely provided

quantifiable values. The dimensions and caliber of blood arteries might differ across individuals due to anatomical and physiological traits, especially height, as well as the existence of vascular pathologies. In our research, we examined the transverse diameters of the AA within the lumen at the site where the LtRA bifurcates from the AA. The diameters mean for male gender was 23.40±2.74 mm; for females22.10±2.59 mm, among each gender was 22.93±2.69 mm.

Pirró et al.[30] revealed the following findings from cadaver examinations for the vertebral body level of the AB: L3 2%, L4 50%, L4/5 intervertebral disk 7%, L5 39%, and S1 2%. The results presented by Prakash et al.[31] were as follows: L3 20%, L4 54%, and L5 26%. Chithriki et al.[16] conducted MRI studies that segmented the vertebral body into three sections, yielding the following results: L3 upper 0.9%, L3 middle 1.4%, L3 lower 7.0%, L3/4 intervertebral disk 13.4%, L4 upper 19.1%, L4 middle 24.0%, L4 lower 23.8%, L4/5 intervertebral disk 7.7%, L5 upper 1.6%, L5 middle 0.9%, and L5/S1. intervertebral disk 0.9%. Our study reveals that 32% of aortic bifurcations arise at the intervertebral disc level between the 4th and 5th lumbar vertebrae, 21% from the upper third of the 4th lumbar vertebra, 21% from the middle third of the 4th lumbar vertebra, 16% from the lower third of the 4th lumbar vertebra, 5% from the intervertebral disc level between the 3rd and 4th lumbar vertebrae, and 5% from the middle third of the 5th lumbar vertebra, considering the ages and genders of the cadavers analyzed (Table 1, Graphic 1).

Our investigation delineated various findings, including aortic dilatation, concomitant aortic aneurysm wall thickness, atherosclerosis, and thrombus at the superior mesenteric artery (SMA). A double right renal artery was noted, accompanied by multiple aortic aneurysms, hemivertebrae, aortic deviation at the L3 vertebra, a compression fracture at L3, increased intervertebral body height between L2 and L3, and osteodegenerative alterations.

It is common for the LtRA to be the initial location of catheter entrapment during abdominal angiography, which involves passing the catheter upward from the femoral artery to the AA in a real clinical environment. In order to place catheters more easily, it is necessary to determine the distance between the LtRA and other blood vessels.[28]

We expressed the relative positional relationships of LtRA in an indexed format, referencing the longitudinal, linear, and transverse distances of the aorta to mitigate the impact of variations in individual concurrent pathologies associated with AA Aneurysm.

No statistically significant relationship was found between age and the transverse diameter of the abdominal aorta at the level of the left renal artery. While a weak positive trend was observed, it does not permit any definitive interpretation due to the absence of statistical significance in the current dataset. Future studies with larger and more homogeneous samples are required to further investigate this potential relationship.

Limitations

One limitation of this study is the relatively small sample size, which may have reduced the statistical power of some analyses and limited the ability to detect significant associations. Consequently, results with p-values above 0.05 should be interpreted with caution. Larger, multicenter studies are needed to confirm these findings and provide more robust conclusions regarding the morphometric characteristics and related vascular pathologies.

CONCLUSION

This cadaveric study demonstrated variability in aortic bifurcation levels, most commonly at the L4-L5 intervertebral disc. Although no statistically significant correlation was found between age and aortic diameter at the level of the left renal artery, a weak positive trend was observed. AAA cases frequently presented with additional anatomical variations such as vertebral deformities, aortic tortuosity, and accessory renal arteries. These rare combinations may have implications for surgical and endovascular planning. The main limitation of the study is the small sample size, which may have limited statistical power. Further research with larger cohorts and radiological or histological correlation is needed to clarify the relationship between aortic morphology, atherosclerosis, and aneurysm development. A better understanding of these variations could enhance the safety of vascular procedures.

ETHICAL DECLARATIONS

Ethics Committee Approval

This project is exempt from ethical approval as it employs course cadavers from the Albert Einstein College of Medicine C&DA Department, consistent with the exemption categories specified in Einstein-IRB-citation104(d).

Informed Consent

Since the study was conducted without the participation of any living being, no written consent form was obtained.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

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

Author Contributions

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

Acknowledgements

I would like to express my endless gratitude to all the donors who donated their bodies for education and science in this study and to the Albert Einstein College of Medicine, Department of Anatomy team.

REFERENCES

1. Wei L, Bu X, Wang X, Liu J, Ma A, Wang T. Global burden of aortic aneurysm and attributable risk factors from 1990 to 2017. *Glob Heart*. 2021;16(1):35. doi:10.5334/gh.920

2. Lederle FA. Aneurysm diameter. *Eur J Vasc Endovasc Surg.* 2000;20(3): 316. doi:10.1053/ejvs.1999.1124
3. Lederle FA, Johnson GR, Wilson SE, et al. Relationship of age, gender, race, and body size to infrarenal aortic diameter. The aneurysm detection and management (ADAM) veterans affairs cooperative study investigators. *J Vasc Surg.* 1997;26(4):595-601. doi:10.1016/s0741-5214 (97)70057-0
4. Wilmink TB, Quick CR, Day NE. The association between cigarette smoking and abdominal aortic aneurysms. *J Vasc Surg.* 1999;30(6):1099-1105. doi:10.1016/s0741-5214(99)70049-2
5. Aune D, Schlesinger S, Norat T, Riboli E. Tobacco smoking and the risk of abdominal aortic aneurysm: a systematic review and meta-analysis of prospective studies. *Sci Rep.* 2018;8(1):14786. doi:10.1038/s41598-018-32100-2
6. Golledge J, Clancy P, Jamrozik K, Norman PE. Obesity, adipokines, and abdominal aortic aneurysm: health in men study. *Circulation.* 2007; 116(20):2275-2279. doi:10.1161/CIRCULATIONAHA.107.717926
7. Hobbs SD, Claridge MW, Quick CR, Day NE, Bradbury AW, Wilmink AB. LDL cholesterol is associated with small abdominal aortic aneurysms. *Eur J Vasc Endovasc Surg.* 2003;26(6):618-622. doi:10.1016/s1078-5884(03)00412-x
8. MacSweeney ST, O'Meara M, Alexander C, O'Malley MK, Powell JT, Greenhalgh RM. High prevalence of unsuspected abdominal aortic aneurysm in patients with confirmed symptomatic peripheral or cerebral arterial disease. *Br J Surg.* 1993;80(5):582-584. doi:10.1002/bjs.1800800510
9. Solberg S, Forsdahl SH, Singh K, Jacobsen BK. Diameter of the infrarenal aorta as a risk factor for abdominal aortic aneurysm: the Tromsø study, 1994-2001. *Eur J Vasc Endovasc Surg.* 2010;39(3):280-284. doi: 10.1016/j.ejvs.2009.10.017
10. Wadhwa A, Soni S. An accessory renal artery from a tortuous abdominal aorta: a case report. *J Clin Diagn Res.* 2012;6:1292-1293.
11. Kinnel M, Faroux L, Villecourt A, et al. Abdominal aorta tortuosity on computed tomography identifies patients at risk of complications during transfemoral transcatheter aortic valve replacement. *Arch Cardiovasc Dis.* 2020;113(3):159-167. doi:10.1016/j.acvd.2019.10.006
12. Darling RC, Messina CR, Brewster DC, Ottinger LW. Autopsy study of unoperated abdominal aortic aneurysms. The case for early resection. *Circulation.* 1977;56(3 suppl):II161-II164.
13. Callewaert B, De Paepe A, Coucke P. Arterial Tortuosity Syndrome. In: Adam MP, Feldman J, Mirzaa GM, Pagon RA, Wallace SE, Amemiya A, eds. GeneReviews®. Seattle (WA): University of Washington, Seattle; November 13, 2014.
14. Panagouli E, Antonopoulos I, Tsoucalas G, et al. Case series and a systematic review concerning the level of the aortic bifurcation. *Folia Morphol (Warsz).* 2021;80(2):302-309. doi:10.5603/FM.a2020.0064
15. Khamanarong K, Sae-Jung S, Supa-Adirek C, Teerakul S, Prachaney P. Aortic bifurcation: a cadaveric study of its relationship to the spine. *J Med Assoc Thai.* 2009;92(1):47-49.
16. Chithriki M, Jaibaji M, Steele RD. The anatomical relationship of the aortic bifurcation to the lumbar vertebrae: a MRI study. *Surg Radiol Anat.* 2002;24(5):308-312. doi:10.1007/s00276-002-0036-3
17. Powell J, Greenhalgh RM. Cellular, enzymatic, and genetic factors in the pathogenesis of abdominal aortic aneurysms. *J Vasc Surg.* 1989;9(2):297-304. doi:10.1067/mva. 1989.vs0090297
18. Johnston KW, Rutherford RB, Tilson MD, Shah DM, Hollier L, Stanley JC. Suggested standards for reporting on arterial aneurysms. Subcommittee on Reporting Standards for Arterial Aneurysms, Ad Hoc Committee on Reporting Standards, Society for Vascular Surgery and North American Chapter, International Society for Cardiovascular Surgery. *J Vasc Surg.* 1991;13(3):452-458. doi:10.1067/mva.1991.26737
19. Hirsch AT, Haskal ZJ, Hertzer NR, et al. ACC/AHA guidelines for the management of patients with peripheral arterial disease (lower extremity, renal, mesenteric, and abdominal aortic): a collaborative report from the American Associations for Vascular Surgery/Society for Vascular Surgery, Society for Cardiovascular Angiography and Interventions, Society for Vascular Medicine and Biology, Society of Interventional Radiology, and the ACC/AHA Task Force on Practice Guidelines (writing committee to develop guidelines for the management of patients with peripheral arterial disease)-summary of recommendations. *J Vasc Interv Radiol.* 2006;17(9):1383-1398. doi:10.1097/01.RVI.0000240426.53079.46
20. Zarins CK, Xu CP, Glagov S. Aneurysmal enlargement of the aorta during regression of experimental atherosclerosis. *J Vasc Surg.* 1992; 15(1):90-101. doi:10.1067/mva.1992.34045
21. Xu H, Huang J, Zeng Y, et al. Network pharmacology and molecular analysis of mechanisms underlying the therapeutic effects of Rhubarb in treating atherosclerosis and abdominal aortic aneurysm. *Heliyon.* 2025;11(4):e41906. doi:10.1016/j.heliyon.2025.e41906
22. Jahangir E, Lipworth L, Edwards TL, et al. Smoking, sex, risk factors and abdominal aortic aneurysms: a prospective study of 18 782 persons aged above 65 years in the Southern Community Cohort Study. *J Epidemiol Community Health.* 2015;69(5):481-488. doi:10.1136/jech-2014-204920
23. Van de Luijtgaarden KM, Rouwet EV, Hoeks SE, Stolker RJ, Verhagen HJ, Majoor-Krakauer D. Risk of abdominal aortic aneurysm (AAA) among male and female relatives of AAA patients. *Vasc Med.* 2017;22(2):112-118. doi:10.1177/1358863X16686409
24. Feller I, Woodburne RT. Surgical anatomy of the abdominal aorta. *Ann Surg.* 1961;154(Suppl 6):239-252. doi:10.1097/00000658-196112000-00033
25. Çetinok EH. Renal artery variations with clinical significance: multiple renal arteries and their coexistent anatomical abnormalities: a cadaveric study. *Maltepe Tıp Derg.* 2024;16(2):46-54. doi: 10.35514/mtd.2024.111
26. Pennington N, Soames RW. The anterior visceral branches of the abdominal aorta and their relationship to the renal arteries. *Surg Radiol Anat.* 2005;27(5):395-403. doi:10.1007/s00276-005-0026-3
27. Takahashi T, Takeuchi K, Ito T, Hayashi S, Qu N, Itoh M. Positional relationships of abdominal aorta landmarks for angiography: observations from the intravascular space. *Surg Radiol Anat.* 2014;36(7): 681-688. doi:10.1007/s00276-013-1249-3
28. Sonesson B, Länne T, Hansen F, Sandgren T. Infrarenal aortic diameter in the healthy person. *Eur J Vasc Surg.* 1994;8(1):89-95. doi:10.1016/s0950-821x(05)80127-6
29. Cauldwell EW, Anson BJ. The visceral branches of the abdominal aorta: topographical relationships. *Am J Anat.* 1943;73(1):27-57. doi:10.1002/aja.1000730103
30. Pirró N, Ciampi D, Champsaur P, Di Marino V. The anatomical relationship of the iliocava junction to the lumbosacral spine and the aortic bifurcation. *Surg Radiol Anat.* 2005;27(2):137-141. doi:10.1007/s00276-004-0301-8
31. Prakash, Mokhasi V, Rajini T, Shashirekha M. The abdominal aorta and its branches: anatomical variations and clinical implications. *Folia Morphol (Warsz).* 2011;70(4):282-286.

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1743257
J Health Sci Med. 2025;8(5):825-831

AI vs AI: clinical reasoning performance of language models in orthopedic rehabilitation

Ertuğrul Safran[1], Yusuf Yaşasın[2]

[1]Department of Physiotherapy and Rehabilitation, Faculty of Health Sciences, Bezmialem Vakıf University, İstanbul, Turkiye
[2]Department of Physiotherapy and Rehabilitation, Institute of Graduate Studies, Bahçeşehir University, İstanbul, Turkiye

Cite this article as: Safran E, Yaşasın Y. AI vs AI: clinical reasoning performance of language models in orthopedic rehabilitation. *J Health Sci Med.* 2025;8(5):825-831.

Received: 15.07.2025 ◆ **Accepted**: 06.08.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: This study aimed to compare the clinical reasoning and treatment planning performance of three advanced large language models (LLMs)-ChatGPT-4o, Gemini 2.5 Pro, and DeepSeek-V3-in orthopedic rehabilitation. Their responses to standardized clinical scenarios were evaluated to determine alignment with evidence-based physiotherapy practices, focusing on relevance, accuracy, completeness, applicability, and safety awareness.

Methods: Three fictional but clinically realistic scenarios involving rotator cuff tendinopathy, lumbar disc herniation with radiculopathy, and anterior cruciate ligament (ACL) reconstruction were developed by an experienced physiotherapist. These scenarios were independently queried on the same day by three AI models using identical prompts. A blinded expert physiotherapist evaluated each model's detailed responses using a 5-point Likert Scale across five domains: clinical accuracy, relevance, completeness, applicability, and safety awareness. Mean scores and descriptive statistics were calculated.

Results: DeepSeek-V3 was consistently rated highest (5/5) across all domains and scenarios, demonstrating comprehensive and clinically rigorous plans. ChatGPT-4o showed strong performance overall, with total scores ranging from 19 to 20 out of 25, though it exhibited lower completeness scores due to less specific milestones. Gemini 2.5 Pro scored lower overall (average total score 18/25), with particular weaknesses in applicability and clinical relevance in complex cases such as lumbar disc herniation. All models provided evidence-based treatment approaches emphasizing pain management, postural correction, gradual strengthening, and return-to-activity progression. Differences arose in emphasis on lifestyle modification, patient education depth, and integration of psychosocial factors, with Gemini uniquely addressing psychological readiness in ACL rehabilitation.

Conclusion: AI-generated rehabilitation plans show substantial concordance with current physiotherapy guidelines but vary in detail and clinical practicality. DeepSeek-V3 outperformed the other models in consistency and safety considerations, while ChatGPT-4o balanced clinical accuracy with moderate completeness. Gemini 2.5 Pro's inclusion of biopsychosocial components offers valuable insights but may require further refinement for clinical applicability. These findings highlight the potential and current limitations of AI tools in orthopedic rehabilitation, suggesting careful model selection based on clinical context and user needs.

Keywords: Artificial Intelligence, clinical reasoning, musculoskeletal diseases, rehabilitation, physical therapy modalities

INTRODUCTION

In recent years, the integration of Artificial Intelligence (AI) into healthcare has accelerated, driven largely by the development of large language models (LLMs) such as OpenAI's ChatGPT, Google's Gemini, and DeepSeek.[1,2] These models are capable of generating human-like responses to diverse queries, including those related to clinical reasoning, patient education, and therapeutic guidance.[3-6] As their accessibility increases, physiotherapy students, clinicians, and even patients are beginning to consult AI tools to gain rapid information on musculoskeletal conditions and rehabilitation strategies.[7,8] However, concerns regarding the reliability and clinical relevance of AI-generated responses particularly in complex decision-making contexts persist and warrant systematic investigation.

Orthopedic rehabilitation represents a central and foundational pillar of physiotherapy practice. It involves the assessment and management of musculoskeletal injuries and disorders such as low back pain, rotator cuff tears, ligament injuries, and post-operative recovery following orthopedic surgeries.[9] Effective orthopedic rehabilitation demands accurate clinical reasoning, individualized treatment planning, and ongoing adaptation based on the patient's response and risk profile.[10,11] For these reasons, it serves as an ideal context

Corresponding Author: Ertuğrul Safran, esafran@bezmialem.edu.tr

in which to evaluate the potential and limitations of AI-generated clinical content.

Despite the growing use of AI in health education and preliminary clinical decision support, comparative analyses of different language models are scarce. Existing studies have largely focused on general medical information delivery[12] or single-model evaluations[13] while no published research has systematically compared multiple AI models in orthopedic physiotherapy. Given their differing architectures and data sources, it is plausible that these models may vary in the accuracy, depth, and clinical applicability of their responses.[14,15]

This pilot study was conducted to compare the clinical reasoning performance of ChatGPT-4o, Gemini 2.5 Pro, and DeepSeek-V3 in orthopedic rehabilitation. Their responses to standardized clinical scenarios were analyzed for relevance, accuracy, completeness, applicability, and safety awareness. By identifying strengths and limitations of each model, the study aims to inform the potential role of AI in physiotherapy education and practice while clarifying its current boundaries.

METHODS

Study Design

This study employed a structured comparative design to evaluate the clinical reasoning performance of three large language models-OpenAI's ChatGPT, Google's Gemini, and DeepSeek-V3-in the context of orthopedic rehabilitation. The models' responses to standardized fictional clinical scenarios were independently rated by expert physiotherapists using a multi-domain scoring rubric (assessing accuracy, clinical reasoning, clarity, and safety) that was developed for this study. This study did not involve human or animal participants. Therefore, ethical approval was not required according to national guidelines. All procedures were carried out in accordance with the ethical rules and the principles.

Clinical Scenario Development and AI Response Collection

Three structured clinical scenarios were developed to reflect common musculoskeletal conditions encountered in orthopedic physiotherapy practice, deliberately covering one spinal condition (lumbar disc herniation), one upper extremity condition (rotator cuff tendinopathy), and one lower extremity post-surgical case (anterior cruciate ligament reconstruction). These scenarios were entirely fictional but constructed by an experienced physiotherapist to ensure clinical realism without involving any identifiable patient data.

Importantly, the process of querying AI models was conducted by a different researcher, independent of the scenario author. This separation ensured neutrality and minimized the risk of bias in prompt formulation or interpretation. Each case was submitted to the AI systems using the following standardized prompt:

"Based on the case details below, outline an appropriate physiotherapy treatment plan, including clinical priorities, precautions, and justification of therapeutic choices."

ChatGPT (GPT-4, accessed via OpenAI), Gemini (accessed via Google AI), and DeepSeek-V3 (accessed via DeepSeek AI platform) were queried with each scenario under identical conditions and on the same day to ensure consistency. No prior context or model-specific optimization was applied.

The three case scenarios, designed to reflect real-life clinical situations and reviewed by the artificial intelligence, are presented below.

Scenario 1: A 52-year-old right-handed female presents with a 6-month history of progressive right shoulder pain. The pain began insidiously without any specific trauma and is localized over the anterolateral aspect of the shoulder. She describes the pain as dull and aching, worsened by overhead activities, reaching behind her back (such as fastening her bra), and sleeping on the affected side. Morning stiffness is minimal, and there is no reported numbness, tingling, or neck pain. She works as an office administrator but is also an amateur tennis player, playing twice a week. Over the past few months, she has reduced her tennis activities due to increasing discomfort. NSAIDs provide partial relief. She reports occasional night pain, particularly after increased shoulder use during the day. Past medical history includes mild hypertension, well controlled with medication. No history of diabetes or rheumatoid disease. No prior shoulder injuries or surgeries. On physical examination, her posture shows mild protraction of the right shoulder and forward head posture. Active range of motion reveals painful arc between 70-120° of abduction and flexion. Passive range of motion is mostly preserved, with mild discomfort at end ranges. Strength testing shows mild weakness (4+/5) in right shoulder abduction and external rotation. Special tests:Positive Neer impingement test, Positive Hawkins-Kennedy test, Jobe (empty can) test positive for pain, minimal weakness, Negative external rotation lag sign, Negative drop arm test. Palpation reveals tenderness over the supraspinatus tendon and subacromial region. No signs of gross instability or labral pathology. Ultrasound report shows hypoechoic areas and mild thickening of the supraspinatus tendon without full-thickness tear; mild subacromial-subdeltoid bursitis noted. Functional impact includes difficulty with overhead reaching, carrying moderate loads, and participating in tennis. Patient expresses concern about long-term shoulder function and returning to recreational sports.

Scenario 2: A 38-year-old right-handed male presents with a 3-month history of low back pain radiating to the right leg. The pain is described as a dull ache in the lumbar region, with burning/shooting radiation to the right gluteal area and posterior thigh. Prolonged sitting and forward bending exacerbate the symptoms, while walking provides partial relief. There is no morning stiffness, and pain decreases with rest. The patient reports a pain intensity of 7 out of 10 on the Visual Analog Scale (VAS 7/10). He works as an office employee, spending approximately 8-10 hours per day at a computer. There is no prior history of lumbar disc herniation or low back pain. His medical history is unremarkable for systemic diseases. He leads a sedentary lifestyle and does not engage in regular exercise. He has been smoking about 10 cigarettes per day for the past 10 years. He occasionally uses

nonsteroidal anti-inflammatory drugs (NSAIDs) for symptom relief. Physical examination reveals mild lumbar hypolordosis and mild atrophy of the right gluteal muscles. Lumbar flexion is restricted and painful; right lateral flexion is painful, left lateral flexion is limited; extension is slightly painful. There is tenderness over the right L4-L5 paravertebral area and right gluteal region with noted myofascial tightness. On orthopedic and neurological testing, the right straight leg raise (SLR) is painful at 45°, and the right Lasègue test is positive. Patellar and Achilles reflexes are normal. Muscle strength testing shows 4/5 right ankle dorsiflexion and 5/5 plantarflexion. There is mild hypoesthesia in the right L5 dermatome. Gait analysis reveals an antalgic pattern with right-side unloading. Lumbar magnetic resonance imaging (MRI) shows a right posterolateral disc herniation at the L4-L5 level, consistent with nerve root compression. Clinically, the patient is diagnosed with L4-L5 disc herniation associated with right L5 radiculopathy. There is pain, mild motor weakness, and sensory alteration. Chronicity risk is high due to the 3-month duration, sedentary lifestyle, and smoking habit.

Scenario 3: A 25-year-old, male, six weeks post-right knee anterior cruciate ligament (ACL) reconstruction, presents for physiotherapy. The injury occurred during an amateur football game when his right knee twisted, resulting in immediate sharp pain, a giving-way sensation, and swelling. Surgery was performed using a hamstring graft, and postoperative instructions included two weeks of bracing and crutch use. He now presents at week 6 post-surgery for rehabilitation. The patient reports occasional pain and a sense of swelling in the right knee, with feelings of instability when going up and down stairs. He experiences fatigue during prolonged walking. He has partially resumed daily activities but is anxious about returning to sports. His rehabilitation goals include achieving full knee range of motion, regaining muscle strength, and ensuring a safe return to sports.

Past medical history: No systemic diseases, non-smoker, normal BMI, previously active lifestyle.

Physical examination findings: Mild right knee swelling, well-healed surgical scar, knee flexion approximately 0-110°, near full extension, quadriceps and hamstring muscle atrophy, notable weakness of the right vastus medialis obliquus (VMO), slightly antalgic gait, full weight-bearing, stability tests; negative Lachman and pivot shift (postoperative), negative valgus/varus stress tests, single-leg stance shows poor balance, minimal pain on palpation, especially around the patellofemoral region.

Functional status: Experiences fatigue during long walks, difficulty climbing stairs, avoids squatting and jumping activities. He is highly motivated with the goal of returning to sports.

Clinical summary: Right ACL reconstruction, postoperative week 6, early-to-mid rehabilitation phase.

Evaluation Criteria and Expert Scoring

To minimize bias and ensure blinding during the evaluation process, the study employed a two-phase approach. One licensed physiotherapist with seven years of clinical experience was responsible for designing the clinical scenarios, inputting them into the AI systems, and collecting the generated responses. A second physiotherapist, with 12 years of experience in the same field and blinded to the identity of the AI model that produced each response, independently scored the answers using a predefined evaluation rubric. This procedure was implemented to reduce potential evaluator bias and maintain objectivity in the scoring process.

A 5-point Likert scale (1 = Very poor, 5 = Excellent) was used across five domains, with explicit scoring anchors provided for each rating level (Table 1). This rubric was adapted from prior frameworks used in studies evaluating AI-generated health content[2,16] and refined for orthopedic rehabilitation by physiotherapy experts.[2,16]

Clinical accuracy: Consistency with evidence-based physiotherapy approaches.

Applicability: Practical relevance and feasibility of the proposed treatment.

Safety awareness: Consideration of contraindications, red flags, and patient-specific risks.

Relevance: The extent to which the model's response directly addresses the clinical scenario and aligns with the presented case context.

Completeness: The degree to which the response provides all essential components of assessment, reasoning, and management without omitting key information.

Domain-specific and total average scores were calculated separately for ChatGPT, Gemini, and Deepseek (Table 1).

Table 1. Multi-domain evaluation rubric for AI-generated responses

Domain	1-very poor	2-poor	3-fair	4-good	5-excellent
Clinical accuracy	Provides incorrect or misleading information; contradicts evidence-based practice.	Mostly inaccurate; several critical errors present.	Mixed accuracy; includes some correct elements but notable gaps.	Largely accurate with only minor errors or omissions.	Fully accurate and aligned with current evidence-based physiotherapy guidelines.
Applicability	No practical application; advice is unusable in clinical context.	Limited relevance; many suggestions unrealistic or impractical.	Moderately relevant; some feasible elements but lacks clarity on implementation.	Mostly practical and relevant with minor feasibility issues.	Fully relevant and readily applicable to clinical physiotherapy practice.
Safety awareness	Ignores major red flags, contraindications, or patient risks.	Acknowledges some risks but omits important safety considerations.	Basic safety elements present but lacks depth or misses subtler risks.	Covers most safety considerations; only minor omissions.	Thoroughly integrates safety, contraindications, and patient-specific precautions.
Relevance	Does not address the case; largely off-topic or generic.	Partially relevant; addresses only some aspects of the case.	Moderately relevant; case-related but missing key contextual links.	Mostly relevant; minor digressions but generally on point.	Fully relevant, directly addressing all aspects of the presented scenario.
Completeness	Major components (assessment, reasoning, management) missing.	Many important aspects missing; very superficial response.	Partially complete; includes key components but lacks depth.	Almost complete with only minor omissions.	Fully complete, covering assessment, reasoning, and management comprehensively.

AI: Artificial Intelligence

Statistical Analysis

Mean scores were calculated for each evaluation domain and case based on the single rater's assessments. Descriptive statistics (mean±standard deviation) are reported. All analyses were performed using IBM SPSS Statistics version 25.

RESULTS

The responses of the three different AI models to the three clinical scenarios are summarized below. Therapist ratings were calculated based on the detailed, full-length responses.

Quantitative assessment revealed significant performance differences across AI models (Table 1). DeepSeek-V3 achieved perfect scores (5/5) in all parameters across all cases, totaling 25/25 for each clinical scenario. ChatGPT-4o demonstrated consistent performance with total scores ranging 19-20/25, though it showed relative weakness in Completeness (average: 3.0/5). Gemini 2.5 Pro scored lowest in Applicability (average: 3.0/5) and total scores (18/25 in all cases), with notable vulnerability in Lumbar Disc Herniation (Relevance: 3/5).

Scenario 1: Shoulder Pathology (Rotator Cuff Tendinopathy & Postural Correction)

ChatGPT-4o provided a comprehensive, phased physiotherapy treatment plan focusing on reducing pain and inflammation associated with supraspinatus tendinopathy and subacromial bursitis. The plan emphasized restoring pain-free shoulder range of motion, correcting rotator cuff and scapular muscle imbalances, and addressing postural abnormalities such as forward head posture and shoulder protraction. Strengthening progressed from pain-free isometrics to functional and sports-specific drills with a controlled return to tennis.

DeepSeek-V3 outlined a pain-guided approach with gradual progression of exercise intensity based on patient tolerance. The plan included manual therapy interventions, neuromuscular control exercises, ergonomic adjustments, and patient education emphasizing self-management. Early management targeted inflammation and pain with careful monitoring to avoid aggravation.

Gemini 2.5 Pro proposed a detailed multi-phase rehabilitation program with a strong focus on symptom control, postural correction, progressive strengthening, and tennis-specific retraining. The approach incorporated manual therapy, neuromuscular education, and technique analysis to optimize biomechanical function and prevent recurrence.

Scenario 2: Lumbar Disc Herniation with Radiculopathy

ChatGPT-4o recommended a conservative treatment plan centered on alleviating radicular pain and decompressing the L5 nerve root. Key elements included restoring lumbar mobility, correcting posture, strengthening core and gluteal muscles, and modifying lifestyle factors such as reducing sedentary behavior and smoking cessation. Neural mobilization techniques and gradual progression of exercises were emphasized.

DeepSeek-V3 presented a comprehensive rehabilitation strategy addressing symptom centralization, neurological function restoration, and lifestyle modification. The plan incorporated ergonomics education, postural retraining, core stabilization, and gradual functional activity retraining, with detailed progression criteria and patient education.

Gemini 2.5 Pro outlined a phase-based approach involving acute symptom management, motor control enhancement, and advanced functional strengthening. Lifestyle interventions, including smoking cessation and home exercise compliance, were integral components to ensure long-term success and recurrence prevention.

Scenario 3: ACL Reconstruction Rehabilitation

ChatGPT-4o emphasized restoring full knee extension and flexion, managing swelling and mild pain, and progressively strengthening the quadriceps and hamstrings. Proprioceptive and neuromuscular control exercises were integrated to improve gait and prevent compensatory patterns. Return to sport was carefully phased with milestones and precautions to protect the graft.

DeepSeek-V3 focused on symmetrical range of motion recovery, dynamic stability, and graft protection through progressive closed-chain strengthening and neuromuscular retraining. Balance and functional mobility exercises were advanced gradually, with consideration for sport-specific demands.

Gemini 2.5 Pro-proposed a criteria-based rehabilitation plan incorporating pain and swelling control, progressive strength training, plyometric and agility drills, and psychological readiness assessment. The program included detailed milestones for return to football, emphasizing safe progression through non-contact to full-contact activities (Table 2).

DISCUSSION

This study compared the clinical reasoning performance of three large language models (LLMs)-ChatGPT-4o, DeepSeek-V3, and Gemini 2.5 Pro-using three scenarios that reflected key areas of orthopedic rehabilitation; one spinal, one upper extremity, and one lower extremity post-surgical case. Across all scenarios, the models produced rehabilitation plans largely consistent with physiotherapy guidelines, but important differences emerged in how each model approached clinical reasoning, integrated psychosocial elements, and managed safety considerations. These findings highlight both the potential and the current boundaries of AI in supporting rehabilitation practice. Although DeepSeek-V3 achieved the highest possible scores across domains, this reflects its alignment with guideline-based benchmarks rather than flawless real-world applicability. In clinical practice, factors such as patient variability, contextual judgment, and therapist-patient interaction remain essential and cannot be fully replicated by an AI-generated plan.Recent advances in LLM have sparked growing interest in their application across various healthcare domains, including physiotherapy.[17-19] As these models, such as ChatGPT, become increasingly integrated into clinical and educational contexts, questions arise regarding their reliability, consistency, and alignment with professional guidelines. In the field of musculoskeletal rehabilitation, where clinical reasoning and individualized decision-making are critical, evaluating the performance

Table 2. AI model performance scores across clinical case

Case	AI model	Relevance	Accuracy	Completeness	Applicability	Safety awareness	Total score
1 (Rotator cuff tendinopathy)	ChatGPT-4o	5	4	3	4	4	20
	Deepseek V3	5	5	5	5	5	25
	Gemini 2.5 Pro	4	3	4	3	4	18
2 (Lumbar disc herniation)	ChatGPT-4o	4	4	3	4	4	19
	Deepseek V3	5	5	5	5	5	25
	Gemini 2.5 Pro	3	4	4	3	4	18
3 (ACL reconstruction)	ChatGPT-4o	4	4	3	4	4	19
	Deepseek V3	5	5	5	5	5	25
	Gemini 2.5 Pro	4	4	3	3	4	18
Overall average	ChatGPT-4o	4.3	4	3	4	4	19.3
	Deepseek V3	5	5	5	5	5	25
	Gemini 2.5 Pro	3.7	3.7	3.7	3	4	18

AI: Artificial Intelligence, ACL: Anterior cruciate ligament

of AI-based tools has become both timely and essential. Hao et al.[20] investigated the alignment of ChatGPT-4's responses with clinical practice guidelines (CPGs) across 30 physical therapy-related questions and found that 80% of its outputs were consistent with established recommendations. The model performed best in upper extremity conditions (100% accuracy) and showed lower performance in spinal cases (60%). Inter-rater reliability was very good (κ=0.847), supporting the model's consistency in clinical interpretation. Bilika et al.[21] evaluated the clinical reasoning capacity of ChatGPT-3.5 in both simple and complex physiotherapy scenarios. While the model demonstrated a reasonable ability to generate evidence-based reasoning in straightforward cases, key assessment components-such as re-assessment and subjective evaluation-were missing in up to 40% of responses. The authors emphasized the variability in output quality and the strong influence of prompt phrasing on model performance.

As interest in the clinical applications of LLMs grows, recent research has shifted from examining single models like ChatGPT to conducting comparative analyses involving multiple platforms-including ChatGPT, Google's Gemini, and DeepSeek-to evaluate their reasoning accuracy, consistency, and alignment with evidence-based guidelines.[14,15,22-25] These comparisons offer a broader perspective on the capabilities and limitations of different language models in healthcare contexts.[12,26] Marcaccini et al.[15] evaluated the diagnostic and management capabilities of ChatGPT-4o, DeepSeek-V3, and Gemini 1.5 in 58 real-world hand fracture cases. While ChatGPT-4o achieved high diagnostic accuracy (98.28%) and recall, DeepSeek-V3 showed moderate performance, and Gemini 1.5 failed to deliver clinically reliable outputs. These findings underscore significant variability in model reliability across complex orthopedic scenarios. Wu et al.[27] investigated the alignment of ChatGPT-4.0 and DeepSeek-R1 with Chinese clinical practice guidelines for knee osteoarthritis (KOA). Among 17 therapeutic recommendations, DeepSeek-R1 demonstrated slightly higher concordance (71%) compared to ChatGPT-4.0 (59%). However, both models showed inconsistent outputs for treatments such as ozone therapy and arthroscopy. Despite high inter-rater agreement, the study emphasized the need for critical scrutiny of AI-generated medical advice. In contrast to the above study focused on KOA, the present research explores AI performance in broader orthopedic rehabilitation contexts with structured scenarios and multi-criteria scoring. Gültekin et al.[28] compared ChatGPT-4o and DeepSeek R1-each equipped with real-time internet access features (Deep Research and DeepThink, respectively)-in generating educational responses to frequently asked questions about ACL surgery. While both models demonstrated high accuracy and consistency, ChatGPT-4o provided more comprehensive answers, whereas DeepSeek R1 produced significantly clearer and more readable content, as supported by Flesch-Kincaid metrics. These findings emphasize model-specific strengths relevant to evidence-based patient education. Saglam et al.[29] compared the clinical decision-making performance of GPT-4 and GPT-3.5 in the fields of sports surgery and physiotherapy using standardized scenarios evaluated by clinicians. GPT-4 outperformed GPT-3.5 across all parameters, including diagnostic accuracy, treatment planning, and rehabilitation protocols. While the findings support the potential of LLMs in medical decision support, the study focused solely on GPT-based models, limiting generalizability to the broader AI landscape. In contrast, the present study expands the comparative framework to include diverse AI architectures-ChatGPT-4o, Gemini 2.5 Pro, and DeepSeek-V3-thus providing a more comprehensive assessment of language model performance in orthopedic rehabilitation contexts. Recent work by Karaagac and Carkit[30] evaluated the performance of four AI-based chatbots-GPT, DeepSeek, Gemini, and Copilot-in providing public health information on liver cancer. Using established quality assessment tools such as the DISCERN instrument and guideline concordance checks (AASLD, NCCN, ESMO), they found that while all models generally offered relevant content, there were notable differences in actionability, clarity, and adherence to guidelines. GPT and DeepSeek were more action-oriented, whereas Gemini and Copilot excelled in user understandability. However, none fully met the standards of consistent, guideline-referenced medical communication. Although their focus was on oncology-related public education, this line of research reinforces the importance

of evaluating AI performance across diverse domains. From a practical perspective, model selection may depend on the user profile. Clinicians might prefer DeepSeek-V3 for its strong emphasis on safety and detailed planning, while students and trainees may benefit from ChatGPT-4o's structured, phased rehabilitation frameworks that support learning. Gemini's inclusion of psychosocial elements and accessible explanations could be particularly valuable for patient education or wellness-oriented applications. This differentiation underscores the importance of matching the model's strengths to the specific needs of its intended user. To our knowledge, the present study is the first to systematically compare ChatGPT-4o, Gemini 2.5 Pro, and DeepSeek-V3 in the specialized context of orthopedic rehabilitation, where clinical reasoning and treatment planning accuracy are critical.

Only one expert physiotherapist rated the AI responses, which prevented assessment of inter-rater reliability. Future studies should involve multiple raters to enhance scoring robustness and allow reliability testing. In addition, there is a possibility of prior familiarity with the models, which, although minimized by separating the scenario author from the evaluator, cannot be completely ruled out. The scenarios themselves were fictional and, despite being designed for clinical realism, inherently involve a degree of subjectivity that may have influenced how responses were interpreted and scored. Finally, because only three standardized scenarios were included for each model, the dataset was too small for meaningful inferential statistical testing. Future studies with larger and more diverse datasets should incorporate appropriate statistical analyses to allow robust comparisons between models.

CONCLUSION

Overall, the AI-generated plans demonstrated substantial alignment with evidence-based physiotherapy principles but varied in the integration of psychosocial components, lifestyle advice, and patient education depth. These differences may inform the choice of AI tool depending on clinical context and user expertise. These findings underscore the growing potential of AI-generated rehabilitation plans to support clinical decision-making; however, as their integration into healthcare settings increases, it is crucial to critically assess their consistency and adherence to established professional guidelines.

ETHICAL DECLARATIONS

Ethics Committee Approval

This study did not require ethical approval as it did not involve any human subjects or animal experiments.

Informed Consent

Because the study did not involve human participants, no written consent was obtained.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

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

Author Contributions

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

Data availability

Data are available upon request.

REFERENCES

1. Zhou Z. Evaluation of ChatGPT's capabilities in medical report generation. *Cureus*. 2023;15(4):e37589. doi:10.7759/cureus.37589
2. Onan D, Arıkan H, Can İ, Güven Ş, Işıkay L, Ozge A. Examining the ability of artificial intelligence with ChatGPT-4.0 to create an exercise program: case scenario examples" lumbar disc herniation, chronic migraine, and urge urinary incontinence". *Turk J Kinesiol*. 2025;11(1): 28-44.
3. Li J, Dada A, Puladi B, Kleesiek J, Egger J. ChatGPT in healthcare: a taxonomy and systematic review. *Comput Method Progr Biomed*. 2024; 245:108013. doi:10.1016/j.cmpb.2024.108013
4. Tustumi F, Andreollo NA, Aguilar-Nascimento JEd. Future of the language models in healthcare: the role of ChatGPT. *Arq Bras Cir Dig*. 2023;36:e1727. doi:10.1590/0102-672020230002e1727
5. Vaishya R, Misra A, Vaish A. ChatGPT: is this version good for healthcare and research? *Diabetes Metabol Syndr*. 2023;17(4):102744. doi:10.1016/j.dsx.2023.102744
6. Wang C, Liu S, Yang H, Guo J, Wu Y, Liu J. Ethical considerations of using ChatGPT in health care. *J Med Internet Res*. 2023;25:e48009. doi: 10.2196/48009
7. Giorgino R, Alessandri-Bonetti M, Del Re M, Verdoni F, Peretti GM, Mangiavini L. Google Bard and ChatGPT in orthopedics: which is the better doctor in sports medicine and pediatric orthopedics? The role of AI in patient education. *Diagnostics*. 2024;14(12):1253. doi:10.3390/ diagnostics14121253
8. Artioli E, Veronesi F, Mazzotti A, et al. Assessing ChatGPT responses to common patient questions regarding total ankle arthroplasty. *J Exp Orthop*. 2025;12(1):e70138. doi:10.1002/jeo2.70138
9. Giangarra CE, Manske RC. Clinical orthopaedic rehabilitation: a team approach E-book: Elsevier Health Sciences; 2017.
10. Sueki D, Brechter J. Orthopedic rehabilitation clinical advisor: Elsevier Health Sciences; 2009.
11. Powden CJ, Hoch MC, Hoch JM. Examination of response shift after rehabilitation for orthopedic conditions: a systematic review. *J Sport Rehabil*. 2018;27(5):469-479. doi:10.1123/jsr.2017-0104
12. Sahin Ozdemir M, Ozdemir YE. Comparison of the performances between ChatGPT and Gemini in answering questions on viral hepatitis. *Sci Rep*. 2025;15(1):1712. doi:10.1038/s41598-024-83575-1
13. Erol E, Arıkan H. Does ChatGPT provide comprehensive and accurate information regarding the effects, types and programming of core exercises? *Turk J Kinesiol*. 2024;10(3):178-182. doi:10.31459/turkjkin. 1516614
14. Rossettini G, Bargeri S, Cook C, et al. Accuracy of ChatGPT-3.5, ChatGPT-4o, Copilot, Gemini, Claude, and Perplexity in advising on lumbosacral radicular pain against clinical practice guidelines: crosssectional study. *Front Digit Health*. 2025;7:1574287. doi:10.3389/ fdgth.2025.1574287
15. Marcaccini G, Seth I, Xie Y, et al. Breaking bones, breaking barriers: ChatGPT, DeepSeek, and Gemini in hand fracture management. *J Clin Med*. 2025;14(6):1983. doi:10.3390/jcm14061983
16. Boone Jr HN, Boone DA. Analyzing likert data. *J Extension*. 2012;50(2): 48. doi:10.34068/joe.50.02.48
17. Sidiq M, Sharma J. ChatGPT in physiotherapy research and education: a boon or bane?-an overview. *Eur J Physiother*. 2025:1-3. doi:10.1080/21 679169.2025.2513354

18. Safran E, Yildirim S. A cross-sectional study on ChatGPT's alignment with clinical practice guidelines in musculoskeletal rehabilitation. *BMC Musculoskelet Disord*. 2025;26(1):411. doi:10.1186/s12891-025-08650-8
19. Naqvi WM, Shaikh SZ, Mishra GV. Large language models in physical therapy: time to adapt and adept. *Front Public Health*. 2024;12:1364660. doi:10.3389/fpubh.2024.1364660
20. Hao J, Yao Z, Tang Y, Remis A, Wu K, Yu X. Artificial intelligence in physical therapy: evaluating ChatGPT's role in clinical decision support for musculoskeletal care. *Ann Biomed Eng*. 2025;53(1):9-13. doi:10.1007/s10439-025-03676-4
21. Bilika P, Stefanouli V, Strimpakos N, Kapreli EV. Clinical reasoning using ChatGPT: is it beyond credibility for physiotherapists use? *Physiother Theory Pract*. 2024;40(12):2943-2962. doi:10.1080/09593985.2023.2291656
22. Arbel Y, Gimmon Y, Shmueli L. Evaluating the potential of large language models for vestibular rehabilitation education: a comparison of ChatGPT, Google Gemini, and Clinicians. *Phys Ther*. 2025;105(4):pzaf010. doi:10.1093/ptj/pzaf010
23. Kleebayoon A, Wiwanitkit V. On "evaluating the potential of large language models for vestibular rehabilitation education: a comparison of ChatGPT, Google Gemini, and Clinicians." *Phys Ther*. 2025;105(4):pzaf010. doi:10.1093/ptj/pzaf010
24. Gürses ÖA, Özüdoğru A, Tuncay F, Kararti C. The role of artificial intelligence large language models in personalized rehabilitation programs for knee osteoarthritis: an observational study. *J Med Syst*. 2025;49(1):73. doi:10.1007/s10916-025-02207-x
25. Puce L, Bragazzi NL, Curra A, Trompetto C. Harnessing generative artificial intelligence for exercise and training prescription: applications and implications in sports and physical activity-a systematic literature review. *Appl Sci (2076-3417)*. 2025;15:7.
26. Usen A, Kuculmez O. Evaluation of the performance of large language models in the management of axial spondyloarthropathy: analysis of EULAR 2022 recommendations. *Diagnostics*. 2025;15(12):1455. doi:10.3390/diagnostics15121455
27. Wu H, Yao S, Bao H, Guo Y, Xu C, Ma J. ChatGPT-4.0 and DeepSeek-R1 does not yet provide clinically supported answers for knee osteoarthritis. *Knee*. 2025;56:386-396. doi:10.1016/j.knee.2025.06.007
28. Gültekin O, Inoue J, Yilmaz B, et al. Evaluating DeepResearch and DeepThink in anterior cruciate ligament surgery patient education: ChatGPT-4o excels in comprehensiveness, DeepSeek R1 leads in clarity and readability of orthopaedic information. *Knee Surg Sports Traumatol Arthrosc*. 2025. doi:10.1002/ksa.12711
29. Saglam S, Uludag V, Karaduman ZO, Arıcan M, Yücel MO, Dalaslan RE. Comparative evaluation of artificial intelligence models GPT-4 and GPT-3.5 in clinical decision-making in sports surgery and physiotherapy: a cross-sectional study. *BMC Med Inform Decis Mak*. 2025;25(1):163. doi:10.1186/s12911-025-02996-8
30. Karaagac M, Carkit S. Evaluation of AI-based chatbots in liver cancer information dissemination: a comparative analysis of GPT, DeepSeek, Copilot, and Gemini. *Oncology*. 2025. doi:10.1159/000546726

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1732198
J Health Sci Med. 2025;8(5):832-838

Comparison of clinical outcome, cost and union times in pediatric ankle fractures: a retrospective case series

Okan Ateş, Mustafa Altıntaş
Department of Orthopaedics and Traumatology, Diyarbakır Gazi Yaşargil Training and Research Hospital, Diyarbakır, Turkiye

Cite this article as: Ateş O, Altıntaş M. Comparison of clinical outcome, cost and union times in pediatric ankle fractures: a retrospective case series. *J Health Sci Med.* 2025;8(5):832-838.

Received: 01.07.2025 ◆ **Accepted**: 08.08.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: The aim of this study was to evaluate the effects of fracture type and implant used in ankle-related fractures in pediatric patients on functional outcome, union time, hospital stay and cost.

Methods: A total of 58 cases operated on due to ankle fractures between 2017 and 2023 were reviewed. Patients were divided into 4 groups according to diagnosis. The first group included patients with isolated malleolus fractures, the second group included patients with bimalleolar or trimalleolar fractures, the third group included patients with distal tibia fractures, and the fourth group included patients with distal tibia fractures with a lateral malleolus. Patients were divided into five groups according to the implants used. These were determined as: K-wire, cannulated screw, plate screw, cannulated screw and plate screw, cannulated screw and plate screw+fixator. Demographic characteristics, union times, implant costs, and the American Orthopaedic Foot and Ankle Society (AOFAS) score were examined between the groups.

Results: Union time was significantly shorter in patients with isolated malleolar fractures compared to other groups (p=0.042). Cost analysis by diagnosis showed that costs in the range of $100-300 were more common in isolated malleolar fractures, whereas costs exceeding $300 were more common in distal tibia fractures (p=0.001). Post-recovery AOFAS (American Orthopaedic Foot & Ankle Society) scores were significantly higher in the isolated malleolar fracture group compared to the bimalleolar/trimalleolar and distal tibia fracture groups (p=0.001). Comparison between all groups; implant comparison, the mean union time (in months) was significantly longer in the cannulated screw+plate and screw+fixator group compared to the K-wire, cannulated screw, plate and screw, and cannulated screw+plate and screw groups (p=0.0001). The plate and screw group also had significantly longer union times than the K-wire and cannulated screw groups (p=0.002, p=0.0001). Post-recovery AOFAS scores were significantly higher in the K-wire group compared to the plate and screw and cannulated screw+plate and screw+fixator groups (p=0.018).

Conclusion: In isolated malleolar fractures, due to the more minor nature of the trauma, shorter hospital stays, quicker fracture union, better functional outcomes, and lower implant costs were observed. Implant costs did not affect functional outcomes or time to union.

Keywords: Ankle fracture, cost analysis, surgical outcomes, pediatric trauma

INTRODUCTION

Tibial fractures rank as the third most common long bone injuries in the pediatric population.[1] Moreover, pediatric ankle fractures account for approximately 18% of all physeal injuries.[2] These fractures are observed twice as frequently in boys compared to girls. The highest incidence occurs between the ages of 8 and 15, with the majority being associated with sports-related activities.[3] The primary objective of treatment is to restore joint congruency and functional alignment in order to prevent the development of osteoarthritis in the long term. Another critical goal is to avoid limb length discrepancies, given that the distal tibial physis contributes to approximately 45% of the overall length of the ankle.[4]

Fractures that have been anatomically reduced are managed with immobilization using a hanging cast and monitored for six weeks with non-weight-bearing mobilization.[4] Non-reduced and rotationally unstable fractures are treated surgically. Medial malleolar fractures, bimalleolar and trimalleolar fractures, distal tibial fractures are treated surgically with a combination of K-wire, cannulated screws, and plates, taking age into account.[5]

Given the high incidence rates observed in adult populations and the increasing costs associated with surgical interventions, a study conducted in the United States has reported that the annual economic burden of ankle fracture treatment amounts to approximately 11 billion USD. Of this, nearly 1.2 billion USD is attributed to direct healthcare costs, including physician fees, surgical supplies, and operating room time.[6] It has been suggested that by understanding the economic implications

Corresponding Author: Okan Ateş, atesokan@msn.com

of different treatment strategies and diagnostic approaches, it may be possible to optimize resource allocation, minimize unnecessary expenditures, and ultimately enhance patient outcomes.[7]

The aim of this study is to evaluate the impact of fracture type and the implant used on functional outcomes, time to bone union, length of hospital stay, and treatment cost in pediatric patients (ages 0-16) with fractures around the ankle region of the lower extremity.

METHODS

Ethical Approval and Study Design

Approval was obtained from the institution and researchers prior to the study. Ethical approval was granted by the Diyarbakır Gazi Yaşargil Training and Research Hospital Non-interventional Clinical Researches Ethics Committee where all imaging and patient procedures were conducted as a single-center study (Date: 25.10.2024, Decision No: 105). The study did not receive any financial support. All procedures were performed in accordance with ethical guidelines and the principles of the Declaration of Helsinki.

Patient Selection and Study Population

This study retrospectively evaluated pediatric patients aged 0-16 years who underwent surgery due to ankle region fractures of the lower extremity between January 2017 and January 2023 at a level 1 trauma center. Patients older than 16 years, those with tibial shaft or foot fractures, patients lost to follow-up, or those without regular radiographic monitoring were excluded. Of the 76 patients who presented to the emergency department during the study period, 14 were excluded due to irregular follow-up, and an additional 4 patients were excluded due to lack of regular radiographic imaging. Ultimately, 58 patients were included. Four patients with soft tissue defects underwent two-stage external fixation treatment, while the remaining received single-stage treatment. Demographic data, length of hospital stay, fracture location, time to union, surgical methods, post-union functional scores, and treatment costs were recorded.

Grouping and Treatment Protocol

Patients were classified into four diagnostic groups: Group 1 included isolated malleolar fractures (**Figure 1**) (n: 16); group 2 comprised bimalleolar or trimalleolar fractures (**Figure 2**) (n: 14); group 3 consisted of distal tibial fractures (**Figure 3**) (n: 19); and group 4 included patients with lateral malleolus fractures combined with distal tibial fractures (n: 9). According to the implants used, patients were divided into five groups: K-wire (**Figure 1c, d**), cannulated screw, plate-screw (**Figure 2 c, d**), cannulated screw with plate-screw, and cannulated screw with plate-screw plus external fixator. Preoperative and postoperative radiological evaluations were performed for all patients. Postoperative immobilization with a cast was applied for one month except for patients with isolated K-wire and cannulated screws. Fracture union was monitored via anteroposterior (AP) and lateral radiographs taken biweekly during the first three months post-discharge, followed by monthly radiographs thereafter. Radiological and

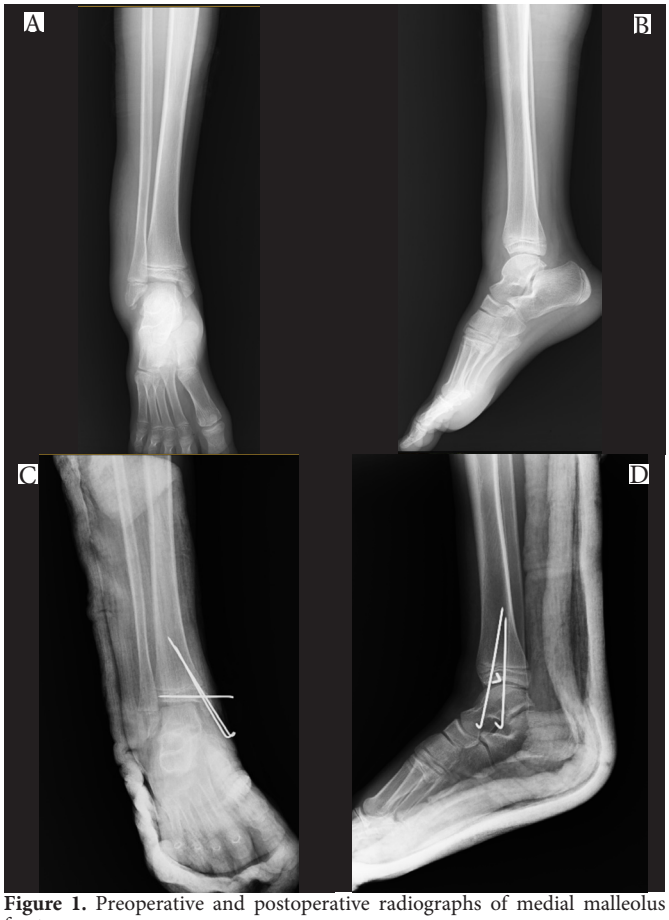

Figure 1. Preoperative and postoperative radiographs of medial malleolus fracture

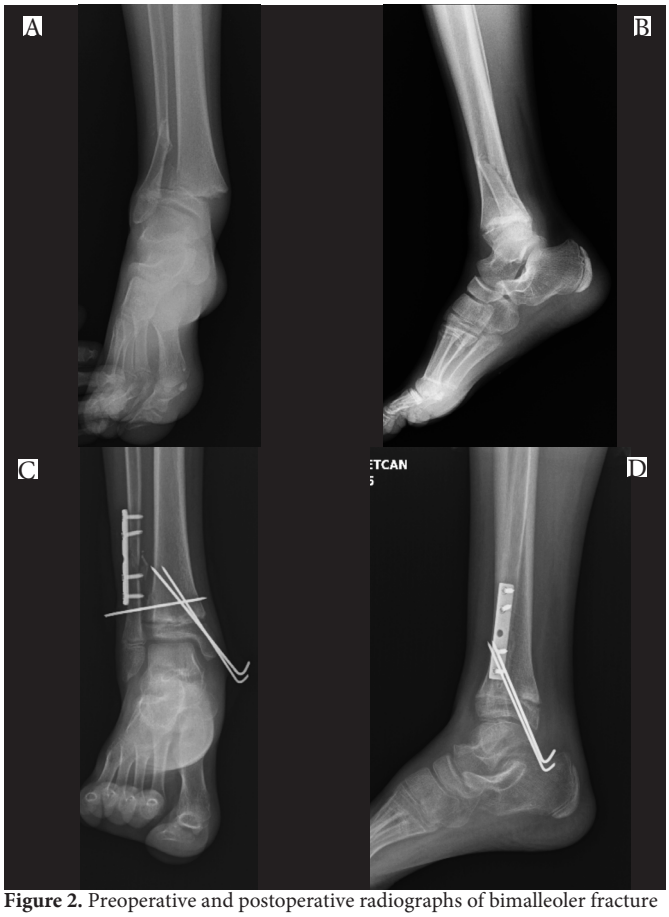

Figure 2. Preoperative and postoperative radiographs of bimalleoler fracture

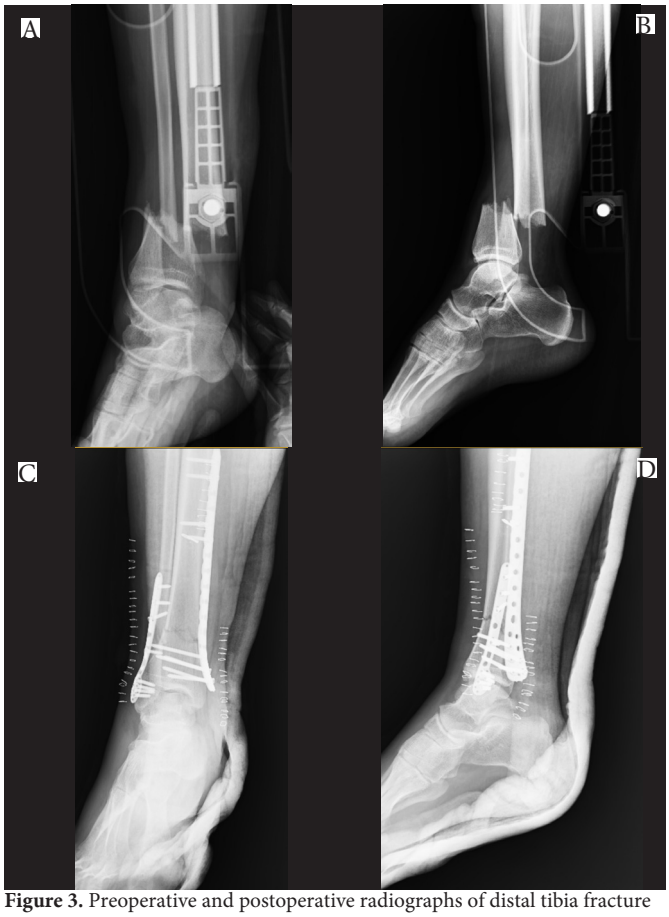

Figure 3. Preoperative and postoperative radiographs of distal tibia fracture

clinical signs of union were defined as the absence of pain or tenderness upon palpation of the fracture line, ability to bear weight painlessly, formation of a hard callus around the fracture line on radiographs, and union of at least three out of four cortices. The American Orthopaedic Foot and Ankle Society (AOFAS)[8] score was assessed after fracture union. Implant costs were evaluated in US dollars according to the exchange rate at the time of treatment and categorized as <100, 100-300, and >300 dollars.

Statistical Analysis

Data analyses were performed using NCSS (Number Cruncher Statistical System) 2007 Statistical Software (Utah, USA). Descriptive statistics such as mean, standard deviation, median, and interquartile range were calculated. Between-group normality of distribution was assessed using the Shapiro-Wilk test. For normally distributed variables, comparisons between groups over time were made using one-way ANOVA, followed by Tukey's post-hoc test for subgroup analyses. Variables with non-normal distribution (those with heterogeneity between groups) were compared using the Kruskal-Wallis test and Dunn's multiple comparison test for subgroup analyses. Categorical data between groups were analyzed using the chi-square test. A p value of less than 0.05 was considered statistically significant.

RESULTS

The demographic characteristics, hospital stay duration, time to fracture union, AOFAS scores, and costs of patients grouped according to diagnosis and implant type are presented separately in Table 1, 2. When grouped into four diagnostic categories, no significant differences were observed between groups in terms of age and gender distribution (Table 1). Regarding the general treatment approach, patients operated on for isolated malleolar fractures were primarily treated with K-wires or cannulated screws, whereas bimalleolar and trimalleolar fractures were treated with cannulated screws and plate-screw fixation, and distal tibial fractures with plate-screw fixation; this distribution showed a statistically significant difference (p=0.002). Analysis of hospital stay duration revealed that patients with isolated malleolar fractures had significantly shorter hospital stays compared to other groups (p=0.042), while no significant differences were noted among the other groups (Table 1, 3, and Figure 4).

Table 1. Outcomes according to fracture diagnosis

		Isolated malleolar n: 16		Bimalleoler-trimalleoler n: 14		Distal Tibia fracture n: 19		Lateral malleolar+ distal tibial n: 9		p
Age (year)	Mean±SD	13.64±2.81		13±3.43		12±2.83		11.38±3.98		0.277*
Gender	Male	9	56.25%	10	71.43%	13	68.42%	4	44.44%	0.519+
	Female	7	43.75%	4	28.57%	6	31.58%	5	55.56%	
Implant type	K-wire	8	50.00%	7	50.00%	4	21.05%	4	44.44%	0.002+
	Cannulated screw	4	25.00%	1	7.14%	0	0.00%	0	0.00%	
	Plate-screw	2	12.50%	1	7.14%	14	73.68%	4	44.44%	
	Cannulated+plate-screw	2	12.50%	3	21.43%	0	0.00%	0	0.00%	
	Cannulated+plate-screw+fixator	0	0.00%	2	14.29%	1	5.26%	1	11.11%	
Hospital stay (days)	mean±SD	2±1.46		6.64±11.29		3.47±1.71		2.67±1.87		0.042‡
	Median (IQR)	2 (1-2)		2 (1-7.5)		3 (2-4)		2 (1.5-3.5)		
Time to union (months)		3.44±0.51		4.36±1.55		5.05±1.35		4.89±0.93		0.001*
Implant cost	<100$	8	50.00%	8	57.14%	4	21.05%	4	44.44%	0.001+
	100-200$	8	50.00%	4	28.57%	2	10.53%	1	11.11%	
	>300$	0	0.00%	2	14.29%	13	68.42%	4	44.44%	
AOFAS score after recovery	Mean±SD	87.88±3.16		81.21±9.07		79.26±6.12		81.33±5.83		0.002*

*One-way analysis of variance, ‡Kruskal-Wallis test, +Chi-square test, SD: Standard deviation, IQR: Interquartile range, AOFAS: American Orthopaedic Foot and Ankle Society

Table 2. Multiple comparison tests based on diagnosis

Multiple comparison test	Dunn's	Tukey	
	Hospital stay	Time to union	AOFAS
Isolated vs bimalleolar-trimalleolar	0.048	0.156	0.029
Isolated vs distal tibial	0.033	0.001	0.001
Isolated malleolus fracture/lateral malleolus fracture+distal tibia fracture	0.245	0.023	0.075
Bimalleolar-trimalleolar/distal tibia fracture	0.631	0.346	0.818
Bimalleolar-trimalleolar/lateral malleolus fracture+distal tibia fracture	0.649	0.717	0.999
Distal tibia fracture/lateral malleolus fracture+distal tibia fracture	0.145	0.986	0.850

AOFAS: American Orthopaedic Foot and Ankle Society

Table 3. Outcomes according to implant type

		K-wire n: 23	Cannulated screw n: 5	Plate-screw n: 21	Cannulated+ plate-screw n: 5	Cannulated+plate-screw+fixator n: 4	p
Age (year)	Mean±SD	10.61±2.87	13.9±3	13.31±2.61	16.18±1.61	14.15±3.39	0.0001*
Gender	Male	12 52.17%	4 80.00%	14 66.67%	4 80.00%	2 50.00%	0.599+
	Female	11 47.83%	1 20.00%	7 33.33%	1 20.00%	2 50.00%	
Diagnosis	Isolated malleolar	8 34.78%	4 80.00%	2 9.52%	2 40.00%	0 0.00%	0.002+
	Bimalleoler-trimalleoler	7 30.43%	1 20.00%	1 4.76%	3 60.00%	2 50.00%	
	Distal tibia fracture	4 17.39%	0 0.00%	14 66.67%	0 0.00%	1 25.00%	
	Lateral malleolar+distal tibial	4 17.39%	0 0.00%	4 19.05%	0 0.00%	1 25.00%	
Hospital stay (days)	Mean±SD	4.61±9.02	1.80±0.45	3.43±1.66	3.20±3.35	3.00±2.16	0.104‡
	Median (IQR)	2 (1-3)	2 (1.5-2)	3 (2-4)	3 (1-6)	2.5 (1.25-5.25)	
Time to union (months)		3.74±0.69	3.2±0.45	5.10±1.26	4.00±0.71	6.75±1.26	0.0001*
Implant cost	<100$	23 100.00%	1 20.00%	0 0.00%	0 0.00%	0 0.00%	0.0001+
	100-200$	0 0.00%	4 80.00%	6 28.57%	5 100.00%	0 0.00%	
	>300$	0 0.00%	0 0.00%	15 71.43%	0 0.00%	4 100.00%	
AOFAS score after recovery	Mean±SD	85.91±7.29	84.8±3.11	79.71±4.79	81.6±9.32	74.75±6.99	0.004*

*One-way analysis of variance, ‡Kruskal-Wallis test, +Chi Square test, SD: Standard deviation, IQR: Interquartile range

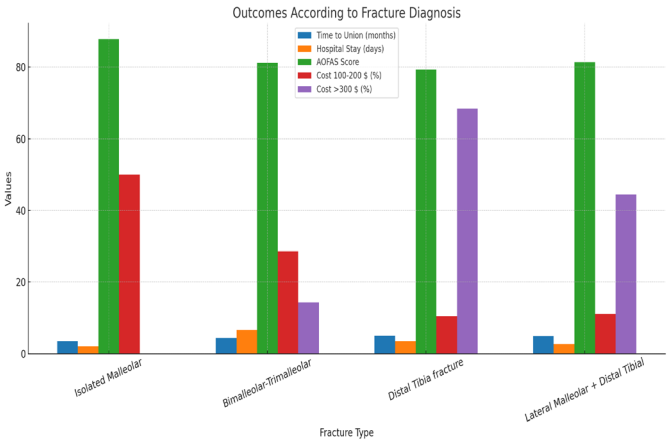

Figure 4. Outcomes and cost by diagnosis

Similarly, time to fracture union was significantly shorter in the isolated malleolar fracture group compared to others (p=0.023) (Table 3, Figure 4). Cost analysis based on diagnosis showed a higher prevalence of costs in the range of $100-300 for isolated malleolar fractures, while distal tibial fractures were associated with costs exceeding $300 (p=0.001). Post-union AOFAS scores were significantly higher in the isolated malleolar fracture group compared to the bimalleolar-trimalleolar and distal tibial fracture groups (p=0.001) (Table 3, Figure 5), with no statistically significant differences observed among the other groups (p>0.05).

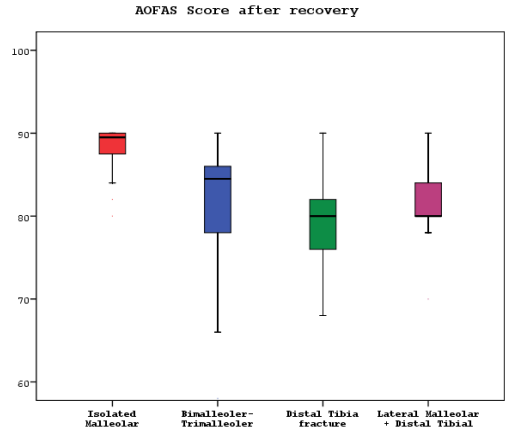

Figure 5. AOFAS score after recovert by diagnosis

AOFAS: American Orthopaedic Foot and Ankle Society

When patients were classified into five groups according to implant type, those treated with K-wires were significantly younger than patients treated with other implants (p=0.0001) (Table 2, Figure 6). No statistically significant difference was found in mean hospital stay duration across implant groups (p=0.104). The mean time to union was significantly longer in the cannulated screw and plate-screw plus external fixator group compared to the K-wire, cannulated screw, plate-screw, and cannulated screw plus plate-screw groups (p=0.0001) (Figure 6). Additionally, the mean union time in the plate-screw group was significantly longer than in the K-wire and cannulated screw groups (p=0.002, p=0.0001, Table 2, 4), no significant differences observed other than the comparisons in previous sentence (p>0.05) (Table 4).

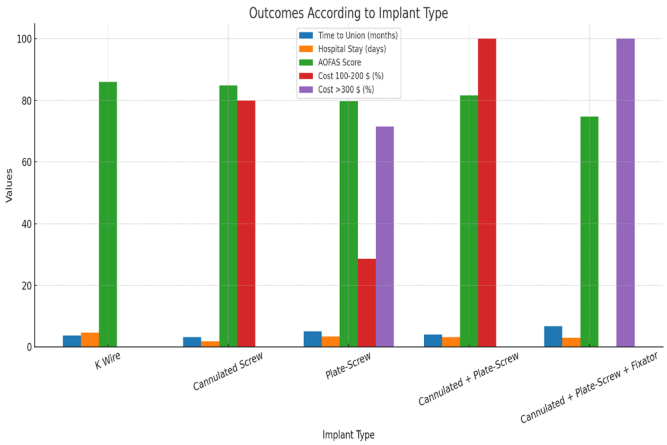

Figure 6. Outcomes and cost by implant type

Table 4. Tukey multiple comparison test based on implant type

Tukey multiple comparison test	Age	Time to union	AOFAS
K-wire/cannulated screw	0.123	0.791	0.997
K-wire/plate screw	0.095	0.0001	0.018
K-wire/cannulated screw and plate screw	0.001	0.982	0.650
K-wire/cannulated screw and plate screw+fixator	0.016	0.0001	0.018
Cannulated screw/plate screw	0.993	0.002	0.504
Cannulated screw/cannulated screw and plate screw	0.682	0.689	0.932
Cannulated screw/cannulated screw and plate screw+fixator	0.999	0.0001	0.147
Plate screw/cannulated screw and plate screw	0.233	0.171	0.975
Plate screw/cannulated screw and plate screw+ fixator	0.981	0.023	0.615
Cannulated screw and plate screw/cannulated screw and plate screw+fixator	0.801	0.001	0.505

AOFAS: American Orthopaedic Foot and Ankle Society

Statistically significant differences were found in post-union AOFAS scores among the implant groups (p=0.0001). The K-wire group demonstrated significantly higher post-union AOFAS scores compared to the plate-screw and cannulated screw plus plate-screw plus external fixator groups (p=0.018), while no significant differences were observed among other groups (p>0.05) (Table 4, Figure 7).

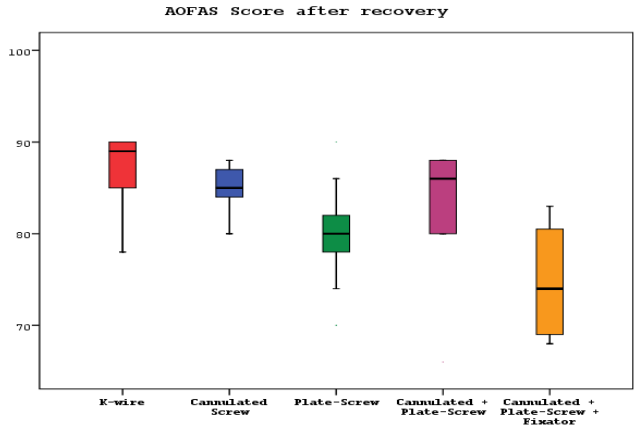

Figure 7. AOFAS score after recovery, grouping the results according to implants

AOFAS: American Orthopaedic Foot and Ankle Society

DISCUSSION

Our study provides valuable insights into the differences in treatment protocols, implant costs, and functional outcomes in operatively treated pediatric ankle fractures, with a focus on diagnostic and therapeutic approaches. Our study offers a different perspective on pediatric ankle fractures by separately evaluating both diagnosis and implants used in the pediatric age group. Consistent with existing literature,[9-12] our findings indicate that patients with isolated malleolar fractures demonstrated superior outcomes in terms of union time, functional results, and length of hospital stay compared to other fracture types. Cost analysis based on implant type showed, as expected, that K-wire fixation was associated with lower costs, and higher AOFAS scores were observed in these patients, likely due to the less invasive nature of both the trauma and the surgical procedure.

Pediatric ankle fractures are frequently encountered and are of particular importance as they often involve the growth plate (epiphysis). Swapnil M. Keny,[13] in his 2024 review of current concepts, emphasized the increasing relevance of modern approaches, including 3D printing and AI (artificial intelligence)-assisted applications in both diagnosis and treatment. In a systematic review by Talaski et al.,[14] the use of CT scans was recommended, particularly in triplane fractures, to assess the need for surgical intervention. Kang et al.[15] reported favorable outcomes in their study involving 46 patients, recommending surgical fixation for fractures with displacements greater than 4 mm. Similarly, Ayas et al.[16] reported positive results in their study of 25 patients, recommending surgery for triplane fractures displaced more than 2 mm. Roberts et al.,[17] in their cohort of 261 patients, emphasized that anatomical reduction was critical for optimal healing and minimizing complications; inadequately reduced fractures were associated with deformity and other issues. In our study, we also observed good outcomes in surgically treated unstable fractures involving the ankle region with displacement greater than 2 mm among 58 patients. Considering the critical role of the epiphyseal region in skeletal development, we recommend detailed evaluation with CT if necessary, and surgical intervention for unstable and displaced fractures.

Ankle region fractures involving the epiphysis are classified as malleolar and metaphyseal fractures based on the Salter-Harris classification.[18] In our study, we categorized patients based on fracture location and implant type. In their guideline on the management of pediatric ankle fractures, Venkatadass et al.[19] recommended K-wires and screws for isolated malleolar fractures, and plate-screw fixation for bimalleolar, trimalleolar, and distal tibia fractures. Their findings are in agreement with ours and with other literature reporting favorable outcomes following closed reduction and percuta neous fixation. Onay et al.,[20] in a study of 39 patients, compared closed reduction with percutaneous fixation, open reduction with screw fixation, and open reduction with plate fixation, and found no significant differences in union or functional outcomes. In our study, K-wires or cannulated screws used for isolated malleolar fractures resulted in better union times and outcomes compared to other fracture types and implant groups. This is likely due to the lower energy of trauma and the relatively smaller anatomical involvement in these cases. No significant differences were found among other fracture types or implant combinations.

In orthopedic surgery, implant cost constitutes a major component of overall expenses. Understanding the clinical and economic implications of different surgical and diagnostic approaches allows surgeons to adopt more efficient and cost-effective strategies. Stull, et al.[21] reported mean costs of $12,920 for isolated bimalleolar injuries and $18,613 for trimalleolar fractures, emphasizing the importance of cost containment. Rainey et al.,[7] in a systematic review of pediatric ankle fractures, found only 7 eligible studies out of 131 reviewed, none of which directly addressed cost, instead focusing on diagnostic strategies and outcome improvement. This highlights a notable gap in the literature and the need for further research. Barfield et al.[18] found that outpatient procedures involving ankle surgery were less costly when performed via ORIF (open reduction internal fixation) compared to non-orthopedic ankle surgery. Given the high frequency and cost burden of these procedures, effective resource utilization and selection of the most efficient treatment options represent an ethical imperative for orthopedic surgeons.

Pediatric ankle fractures currently hold a significant place in pediatric fractures, but if not treated appropriately, they can lead to future deformities and morbidity. Therefore, patients should undergo a thorough evaluation and, if necessary, consultation with several individuals before deciding on surgical options. It should be remembered that any mistake could result in a person with walking impairment in the future, and due diligence should be exercised, taking into account the costs involved.

Limitations

This study has several limitations. First, it was a retrospective study relying on previously recorded data. Although data entry was conducted prospectively, which may mitigate some recall and collection biases, these cannot be entirely eliminated. Further randomized controlled studies with larger sample sizes are needed to validate our findings. Another limitation is the relatively small sample size, as data were collected from a single trauma center. Surgeries were also performed by different surgeons, introducing procedural variability. In addition, the number of screws used per patient was not standardized, which may have affected cost analyses and outcomes.

CONCLUSION

Surgical treatment of isolated malleolar, bimalleolar, trimalleolar, and distal tibial fractures in the pediatric population yielded favorable results. Isolated malleolar fractures, due to their lower-energy nature, were associated with shorter hospital stays, quicker union times, better functional outcomes, and lower implant costs. Although distal tibial fractures were linked to higher implant costs, However, the comparable clinical results across non-malleolar fracture types suggest that implant costs do not significantly impact overall functional outcomes. These findings support the ethical and clinical necessity for efficient and cost-effective treatment planning in pediatric ankle fractures.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was carried out with the permission of the Diyarbakır Gazi Yaşargil Training and Research Hospital Non-interventional Clinical Researches Ethics Committee (Date: 25.10.2024, Decision No: 105).

Informed Consent

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

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

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

Author Contributions

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

REFERENCES

1. Cruz AI, Raducha JE, Swarup I, Schachne JM, Fabricant PD. Evidence-based update on the surgical treatment of pediatric tibial shaft fractures. *Curr Opin Pediatr*. 2019;31(1):92-102. doi:10.1097/MOP.0000000000000704
2. Su AW, Larson AN. Pediatric ankle fractures: concepts and treatment principles. *Foot Ankle Clin*. 2015;20(4):705-719. doi:10.1016/j.fcl.2015.07.004
3. Cancino B, Sepúlveda M, Birrer E. Ankle fractures in children. *EFORT Open Rev*. 2021;6(7):593-606. doi:10.1302/2058-5241.6.200042
4. Olgun ZD, Maestre S. Management of pediatric ankle fractures. *Curr Rev Musculoskelet Med*. 2018;11(3):475-484. doi:10.1007/s12178-018-9510-3
5. Cancino B, Sepúlveda M, Birrer E. Ankle fractures in children. *EFORT Open Rev*. 2021;6(7):593-606. doi:10.1302/2058-5241.6.200042

6. Belatti DA, Phisitkul P. Economic burden of foot and ankle surgery in the US Medicare population. *Foot Ankle Int*. 2014;35(4):334-340. doi:10.1177/1071100713519777
7. Rainey M, Ramirez C, Dayoub E, Sanchez D, Childs B, Bacak C. The economic ramifications of pediatric distal tibial physeal fractures. *J Orthopaed Business*. 2025;5:1.
8. Madeley NJ, Wing KJ, Topliss C, Penner MJ, Glazebrook MA, Younger AS. Responsiveness and validity of the SF-36, Ankle Osteoarthritis Scale, AOFAS ankle hindfoot score, and Foot Function Index in end stage ankle arthritis. *Foot Ankle Int*. 2012;33(1):57-63. doi:10.3113/FAI.2012.0057
9. Verhage SM, Schipper IB, Hoogendoorn JM. Long-term functional and radiographic outcomes in 243 operated ankle fractures. *J Foot Ankle Res*. 2015;8(1):45. doi:10.1186/s13047-015-0098-1
10. Pina G, Fonseca F, Vaz A, Carvalho A, Borralho N. Unstable malleolar ankle fractures: evaluation of prognostic factors and sports return. *Arch Orthop Trauma Surg*. 2021;141(1):99-104. doi:10.1007/s00402-020-03650-w
11. Broos PLO, Bisschop APG. Operative treatment of ankle fractures in adults: correlation between types of fracture and final results. *Injury*. 1991;22(5):403-406. doi:10.1016/0020-1383(91)90106-O
12. Tejwani NC, McLaurin TM, Walsh M, Bhadsavle S, Koval KJ, Egol KA. Are outcomes of bimalleolar fractures poorer than those of lateral malleolar fractures with medial ligamentous injury? *J Bone Joint Surg*. 2007;89(7):1438-1441. doi:10.2106/JBJS.F.01006
13. Keny SM. Pediatric and adolescent ankle fractures: current concepts and advances in management. *J Clin Orthopaed*. 2024;9(2):44-46. doi:10.13107/jcorth.2024.v09i02.660
14. Talaski GM, Sleem B, Luo EJ, et al. Diagnosis, treatment, and potential complications of triplane ankle fractures in pediatric patients: a systematic review. *J Clin Med*. 2025;14(5):1578. doi:10.3390/jcm14051578
15. Kang H, Kang T, Lim C. The incidence and treatment outcome of atypical triplane fractures in adolescents. *Indian J Orthop*. 2022;56(12):2133-2140. doi:10.1007/s43465-022-00679-4
16. Ayas MS, Kalkışım M, Turgut MC, et al. Analysis of clinical outcomes in pediatric distal tibia triplanar fractures treated surgically and conservatively. *Cureus*. 2021;13(12):e20723. doi:10.7759/cureus.20723
17. Roberts V, Mason LW, Harrison E, Molloy AP, Mangwani J. Does functional outcome depend on the quality of the fracture fixation? Mid to long term outcomes of ankle fractures at two university teaching hospitals. *Foot Ankle Surg*. 2019;25(4):538-541. doi:10.1016/j.fas.2018.04.008
18. Dias LS, Tachdjian MO. Physeal injuries of the ankle in children: classification. *Clin Orthop Relat Res*. 1978;(136):230-233.
19. Venkatadass K, Sangeet G, Prasad VD, Rajasekaran S. Paediatric ankle fractures: guidelines to management. *Indian J Orthop*. 2021;55(1):35-46. doi:10.1007/s43465-020-00270-9
20. Onay T, Çelen ZE, Bayhan M, Kandemir İ, Kiliç NC, Kayaalp ME. A more conservative approach in the surgical management of pediatric physeal ankle fractures should be preferred: mid to long-term functional outcomes of three different surgical techniques for salter-harris type II and triplane distal tibial fractures. *J Pediatr Orthop*. 2023;43(9):e734-e741. doi:10.1097/BPO.0000000000002471
21. Stull JD, Bhat SB, Kane JM, Raikin SM. Economic burden of ınpatient admission of ankle fractures. *Foot Ankle Int*. 2017;38(9):997-1004. doi:10.1177/1071100717709576

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1746396

J Health Sci Med. 2025;8(5):839-846

Prognostic value of left ventricular remodeling phenotypes in patients undergoing transcatheter aortic valve implantation

Murat Can Güney

Department of Cardiology, Medicana International Ankara Hospital, Faculty of Medicine, Atılım University, Ankara, Turkiye

Cite this article as: Güney MC. Prognostic value of left ventricular remodeling phenotypes in patients undergoing transcatheter aortic valve implantation. *J Health Sci Med.* 2025;8(5):839-846.

Received: 19.07.2025 ◆ **Accepted**: 10.08.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: This study aimed to evaluate the impact of baseline left ventricular (LV) remodeling phenotypes on clinical and echocardiographic outcomes in patients undergoing transcatheter aortic valve implantation (TAVI).

Methods: A total of 413 patients with aortic stenosis (AS) who underwent TAVI between July 2011 and January 2024 were retrospectively analyzed. Based on echocardiographic parameters, patients were classified into concentric remodeling (CR, 7%), concentric hypertrophy (CH, 84.5%), and eccentric hypertrophy (EH, 8.5%) groups.

Results: Patients in the EH group were significantly younger (mean age: 76, p<0.001) and predominantly male (p<0.001). Prior myocardial infarction (MI) (p<0.001) and coronary artery bypass grefting (CABG) (p=0.003) were more common in this group. EH patients had the lowest baseline ejection fraction (EF) (p<0.001), highest left ventricular end diastolic dimension (LVEDD) and left ventricular end systolic dimension (LVESD) (both p<0.001), increased prevalence of low flow low gradient (LFLG) AS (p<0.001), and lower frequency of very severe aortic stenosis (VSAS) (p=0.005). At one-year follow-up, EH patients showed the most pronounced improvement in EF (+17.6%, p=0.002) and reduction in LVEDD (-7.3%, p=0.006). Permanent pacemaker implantation was highest in the EH group (28.6%) and significantly greater than in the CH group (p=0.022). No significant differences in in-hospital or one-year mortality were observed between groups (p>0.05).

Conclusion: LV remodeling patterns are strongly associated with reverse remodeling and conduction-related complications after TAVI. While EH patients show greater structural recovery, they are also at higher risk for post-procedural pacemaker implantation.

Keywords: Transcatheter aortic valve implantation, left ventricular remodeling, concentric hypertrophy, eccentric hypertrophy, permanent pacemaker, aortic stenosis

INTRODUCTION

Severe aortic stenosis (AS) leads to chronic left ventricular (LV) pressure overload, resulting in various patterns of myocardial remodeling such as concentric remodeling (CR), concentric hypertrophy, (CH) and eccentric hypertrophy (EH). These adaptations, while initially compensatory, may become maladaptive over time, contributing to myocardial fibrosis, systolic and diastolic dysfunction, and heart failure.[1-3]

Transcatheter aortic valve implantation (TAVI) effectively relieves afterload and may promote reverse remodeling-manifested by reductions in LV mass and improvements in systolic function.[4] However, the degree and prognostic relevance of reverse remodeling vary significantly among patients.[5]

Recent studies suggest that baseline LV geometry influences both the extent of reverse remodeling and clinical outcomes after TAVI.[6] Notably, concentric remodeling has emerged as the least favorable LV geometric pattern in terms of prognosis. In patients undergoing TAVI, this phenotype has been independently associated with a significantly higher risk of all-cause mortality at one year, despite comparable procedural outcomes.[7] These findings underline the prognostic relevance of preprocedural LV geometry in risk stratification.

Despite growing interest, limited data exist on how specific LV remodeling phenotypes affect reverse remodeling and outcomes after TAVI. This study aims to assess the association between baseline remodeling patterns, the extent of reverse remodeling, and their impact on clinical and echocardiographic outcomes in patients undergoing TAVI.

METHODS

Ethics

The study was carried out with the permission of the Gazi University President's Office Ethics Committee (Date: 14.07.2025, Decision No: 2025-1243). All procedures were

Corresponding Author: Murat Can Güney, murat_guney@msn.com

carried out in accordance with the ethical rules and the principles of the Declaration of Helsinki.

Study Design and Population

This retrospective observational study included patients who underwent TAVI for severe symptomatic aortic stenosis between July 2011 and January 2024. All procedures were performed via the transfemoral route. Patients were included if they had baseline and follow-up transthoracic echocardiographic (TTE) data available and had complete clinical outcome documentation. Exclusion criteria included valve-in-valve procedures, prior aortic valve surgery, and presence of incomplete echocardiographic data. The study has received approval from the local ethics committee and adheres to the Declaration of Helsinki.

Patient Evaluation and Procedure

All patients were evaluated by a multidisciplinary heart team including interventional cardiologists, cardiac imaging specialists, cardiovascular surgeons, and anesthesiologists. Preprocedural assessments included TTE and/or transesophageal echocardiography, coronary angiography, and multi-slice computed tomography to determine anatomical suitability for TAVI. The procedure was performed under fluoroscopic and echocardiographic guidance. Percutaneous closure systems (e.g., ProGlide or Prostar) were used in most cases, and patients received dual antiplatelet therapy or tailored anticoagulation regimens post-procedure depending on comorbid conditions.

Echocardiographic Analysis

TTE was performed before the procedure and repeated at follow-up (typically at 1 year) to assess cardiac structural and functional changes. LV dimensions, wall thickness, and left ventricular ejection fraction (LVEF) were measured according to current guideline recommendations. Changes in LVEF, left ventricular end-diastolic dimension (LVEDD), left ventricular end systolic dimension (LVESD) were calculated according to the baseline and follow up echocardiographic data. Changes in echocardiographic parameters at one year visit are calculated as delta left ventricular end-diastolic dimension (d-LVEDD) and d-LVEF separately according to the following equations;

d-LVEDD (%)=[(LVEDD at one year after TAVI-baseline LVEDD)/baseline LVEDD]*100

d-LVEF (%)=[(LVEF at one year after TAVI-baseline LVEF)/baseline LVEF]*100

Relative wall thickness (RWT) was calculated as RWT=[2 × posterior wall thickness (PWT)]÷LVEDD and Left Ventricular Mass Index (LVMI) as LVMI = 0.8 × [1.04 × (LVEDD + PWT + IVS)3 – LVEDD3)] + 0.6 and indexed to body surface area (BSA).

LV geometry was classified into three remodeling types based on LVMI and relative wall thickness (RWT): Concentric remodeling (normal LVMI, RWT >0.42), Concentric hypertrophy (increased LVMI, RWT >0.42), Eccentric hypertrophy (increased LVMI, RWT ≤0.42). Increased LVMI is defined as LVMI >95 g/m^2 for women and LVMI > 115 g/m^2 for men.

Outcomes and Follow-Up

Clinical and echocardiographic follow-up data were obtained at 1-year post-TAVI. The primary outcome was all-cause mortality. Secondary outcomes included symptomatic improvement in New York Heart Association (NYHA) class, rehospitalization for heart failure, and echocardiographic parameters of reverse remodeling.

Statistical Analysis

The distributional characteristics of continuous variables were initially assessed using the Kolmogorov-Smirnov test to determine conformity to normality. Homogeneity of variances was evaluated with Levene's test. Descriptive statistics were reported as mean±standard deviation, median (minimum-maximum), or median (25^{th}-75^{th} percentiles) for continuous variables, and as frequency (n) and percentage (%) for categorical variables. For group comparisons involving continuous variables that satisfied parametric assumptions, One-Way Analysis of Variance (ANOVA) was applied. In cases where assumptions of normality or homogeneity of variances were not met, the non-parametric Kruskal-Wallis test was employed. When a significant overall difference was detected, post-hoc pairwise comparisons were performed using Tukey's HSD or Dunn-Bonferroni tests, depending on the nature of the data. For categorical variables, Pearson's chi-square (χ^2) test was used unless otherwise specified. When more than 25% of cells in a 2x2 contingency table had an expected frequency below 5, Fisher's exact test was applied. In cases where expected frequencies ranged between 5 and 25, the continuity-corrected chi-square test was preferred. For RxC tables where one or both categorical variables had more than two levels, and expected frequencies fell below acceptable thresholds, the Fisher-Freeman-Halton test was used. All statistical analyses were conducted using IBM SPSS Statistics version 25.0 (IBM Corp., Armonk, NY, USA). A two-sided p-value of <0.05 was considered statistically significant. However, to control for type I error in multiple comparisons, Bonferroni correction was applied where appropriate.

RESULTS

A total of 413 patients aged between 52 and 103 years were included in the study. Of these, 7% (n=29) had CR, 84.5% (n=349) had CH, and 8.5% (n=35) had EH. All subsequent analyses and interpretations were conducted according to this grouping.

Table 1 presents descriptive statistics of baseline clinical characteristics. A significant difference was observed in mean age across the groups (p<0.001), primarily due to lower mean age in EH group compared to CR and CH groups (p<0.001). Gender distribution also differed significantly among groups (p<0.001), driven by a higher proportion of males and a lower proportion of females in EH group compared to CH group (p<0.001).

No significant differences were found among groups in anthropometric measurements (p>0.05). Similarly, there were no statistically significant differences in NYHA class, The Society of Thoracic Surgeons (STS) score, prior valve surgery, chronic obstructive pulmonary disease (COPD) severity,

Table 1. Baseline characteristics of the patients according to groups

	CR group (n=29)	CH group (n=349)	EH group (n=35)	p-value
Age (years)*	80.0±6.3[A]	77.9±7.7[B]	72.4±7.8[A,B]	<0.001[a]
Gender				<0.001[b]
Male	17 (58.6%)	142 (40.7%)[B]	26 (74.3%)[B]	
Female	12 (41.4%)	207 (59.3%)[B]	9 (25.7%)[B]	
Anthropometric characteristics				
Weight (kg)*	74.7±13.4	73.1±13.4	72.8±10.9	0.814[a]
Height (m)*	1.66±0.072	1.65±0.073	1.65±0.065	0.642[a]
Body-mass index (kg/m^2)*	27.1±4.4	26.9±4.4	26.7±3.8	0.940[a]
NYHA				0.538[c]
2	9 (31.0%)	98 (28.1%)	6 (17.1%)	
3	17 (58.6%)	190 (54.4%)	22 (62.9%)	
4	2 (6.9%)	54 (15.5%)	6 (17.1%)	
Pulmonary edema	1 (3.4%)	7 (2.0%)	1 (2.9%)	
STS score**	6.0 (3.3-7.5)	6.7 (5.3-8.4)	6.7 (5.5-8.5)	0.136[d]
Previous valve surgery	0 (0.0%)	8 (2.3%)	3 (8.6%)	0.113[c]
COPD				0.543[c]
None	20 (69.0%)	177 (50.7%)	16 (45.7%)	
Mild	1 (3.4%)	23 (6.6%)	1 (2.9%)	
Medium	6 (20.7%)	101 (28.9%)	11 (31.4%)	
Severe	2 (6.9%)	48 (13.8%)	7 (20.0%)	
Other co-morbidities				
CVA	1 (3.4%)	15 (4.3%)	2 (5.7%)	0.874[c]
PAH	4 (13.8%)	26 (7.4%)	3 (8.6%)	0.376[c]
DM	7 (24.1%)	106 (30.4%)	14 (40.0%)	0.363[b]
HT	24 (82.8%)	292 (83.7%)	29 (82.9%)	0.98
HPL	17 (58.6%)	181 (51.9%)	23 (65.7%)	0.249[b]
AF	5 (17.2%)	80 (22.9%)	9 (25.7%)	0.711[b]
Basal GFR**	67.0 (51.5-83.0)	65.0 (50.0-81.0)	65.0 (56.0-84.0)	0.553[d]
MI history	1 (3.4%)[A]	40 (11.5%)[B]	13 (37.1%)[A,B]	<0.001[c]
PCI history	6 (20.7%)	76 (21.8%)	12 (34.3%)	0.234[b]
CABG history	6 (20.7%)[A]	79 (22.6%)[B]	17 (48.6%)[A,B]	0.003[b]

Descriptive statistics are presented as * mean ± standard deviation or ** median (25th-75th percentile). a One-Way Analysis of Variance (ANOVA), b Pearson's χ^2 test, c Fisher-Freeman-Halton test, d Kruskal-Wallis test. A: The difference between CR group and EH group is statistically significant (p<0.05); B: The difference between CH group and EH group is statistically significant (p<0.01).
CR: Concentric remodeling, CH: Concentric hypertrophy, EH: Eccentric hypertrophy, NYHA: New York Hear Association, STS: Society of Thoracic Surgeons, COPD: Chronic obstructive pulmonary disease, CVA: Cerebrovascular accident, PAH: Peripheral artery disease, DM: Diabetes mellitus, HT: Hypertension, HPL: Hyperlipidemia, AF: Atrial fibrillation, GFR: Glomerular filtration rate, MI: Myocardial infarction, PCI: Percutaneous coronary intervention, CABG: Coronary artery bypass grefting

stroke, pulmonary hypertension, diabetes, hypertension, hyperlipidemia, or history of atrial fibrillation (p>0.05).

Baseline GFR levels were comparable across groups (p=0.553). However, there was a significant difference in the history of myocardial infarction (MI) (p<0.001), with EH group showing a higher prevalence compared to CR group (p=0.003) and CH group (p<0.001). No significant difference was observed in the history of stenting (p=0.234), but the prevalence of prior CABG was significantly higher in EH group than in CR group (p=0.040) and CH group (p=0.002) (p=0.003 overall).

Table 2 summarizes baseline echocardiographic findings. Baseline LVEF was significantly lower in both CH group and EH group compared to CR group (p<0.001), with EH group also having significantly lower LVEF than CH group (p<0.001).

LVEDD was significantly higher in CH group and EH group compared to CR group (p<0.001), and also significantly higher in EH group than CH group (p<0.001). LVESD showed a similar pattern, being significantly larger in CH group and EH group versus CR group (p<0.001), and significantly higher in EH group compared to CH group (p<0.001).

Septal and posterior wall thicknesses were significantly higher in CH group than in both CR and EH groups (p<0.001). Left atrial (LA) diameter was significantly greater in both CH group (p=0.008) and EH group (p<0.001) compared to CR group, and also significantly higher in EH group than in CH group (p<0.001).

Table 2. Baseline echocardiographic findings of the patients according to groups				
	CR group (n=29)	**CH group (n=349)**	**EH group (n=35)**	**p-value**
Basal EF*	65.0 (60.0-65.0)[A,B]	60.0 (45.0-65.0)[A,C]	30.0 (25.0-40.0)[B,C]	<0.001[a]
Basal LVEDD*	4.1 (3.8-4.2)[A,B]	4.6 (4.4-4.9)[A,C]	5.9 (5.5-6.4)[B,C]	<0.001[a]
Basal LVESD*	2.4 (2.2-2.6)[A,B]	2.9 (2.6-3.4)[A,C]	4.7 (4.0-5.3)[B,C]	<0.001[a]
Basal septum*	1.10 (1.05-1.30)[A]	1.40 (1.30-1.50)[A,C]	1.20 (1.10-1.30)[C]	<0.001[a]
Basal posterior*	1.10 (1.00-1.20)[A]	1.30 (1.20-1.40)[A,C]	1.10 (1.00-1.20)[C]	<0.001[a]
LVM*	157.4 (140.8-178.7)	251.4 (220.5-284.8)	276.3 (246.5-338.8)	n/a
RWT*	0.55 (0.48-0.63)	0.56 (0.51-0.62)	0.37 (0.32-0.40)	n/a
LVMI*	92.6 (82.8-105.1)	147.9 (129.7-167.6)	162.5 (145.0-199.3)	n/a
Basal LA*	4.3 (4.0-4.6)[A,B]	4.5 (4.3-4.9)[A,C]	4.9 (4.5-5.4)[B,C]	<0.001[a]
AS group				<0.001[b]
HG	24 (82.8%)	231 (66.2%)	25 (71.4%)	
LFLG	0 (0.0%)[B]	11 (3.2%)[C]	8 (22.9%)[B,C]	
Paradoxical LFLG	0 (0.0%)	4 (1.1%)	0 (0.0%)	
VSAS	5 (17.2%)	103 (29.5%)[C]	2 (5.7%)[C]	
Basal AoVel*	4.4 (4.1-4.8)[B]	4.5 (4.2-5.0)[C]	4.2 (3.8-4.3)[B,C]	<0.001[a]
Basal AVA*	0.68 (0.56-0.75)[B]	0.68 (0.53-0.80)[C]	0.80 (0.67-0.90)[B,C]	0.002[a]
Basal mean gradient*	49.0 (41.0-55.0)[B]	49.0 (42.0-60.0)[C]	41.0 (35.0-45.0)[B,C]	<0.001[a]
Basal PASB*	40.0 (31.0-55.0)	40.0 (34.0-55.0)	40.0 (30.0-55.0)	0.973[a]
Basal MR				0.766[b]
None	0 (0.0%)	4 (1.1%)	0 (0.0%)	
Trivial	7 (24.1%)	75 (21.5%)	8 (22.9%)	
1	14 (48.3%)	161 (46.1%)	15 (42.9%)	
2	7 (24.1%)	72 (20.6%)	5 (14.3%)	
3	1 (3.4%)	32 (9.2%)	7 (20.0%)	
4	0 (0.0%)	5 (1.4%)	0 (0.0%)	
Bazal AR				0.474[b]
None	8 (27.6%)	44 (12.6%)	5 (14.3%)	
Trivial	4 (13.8%)	51 (14.6%)	4 (11.4%)	
1	13 (44.8%)	181 (51.9%)	15 (42.9%)	
2	3 (10.3%)	59 (16.9%)	9 (25.7%)	
3	1 (3.4%)	9 (2.6%)	1 (2.9%)	
4	0 (0.0%)	5 (1.4%)	1 (2.9%)	
Bicuspid aorta	1 (3.4%)[B]	53 (15.2%)	9 (25.7%)[B]	0.048[c]

Descriptive statistics are presented as *median (25^{th}-75^{th} percentile). a Kruskal-Wallis test, b Fisher-Freeman-Halton test, c Pearson's χ^2 test. n/a: Not assessed. A: The difference between CR group and CH group is statistically significant (p<0.01); B: The difference between CR group and EH group is statistically significant (p<0.05); C: The difference between CH group and EH group is statistically significant (p<0.01).
EF: Ejection fraction, LVEDD: Left ventricular end-diastolic dimension, LVESD: Left ventricular end-systolic dimension, LVM: Left ventricular mass, RWT: Relative wall thickness, LVMI: Left ventricular mass index, LA: Left atrium, AS: Aortic stenosis, HG: High gradient, LFLG: Low flow low gradient, VSAS: Very severe aortic stenosis, AoVel: Aortic Velocity, AVA: Aortic valve area, PASP: Pulmonary artery systolic pressure, MR: Mitral regurgitation, AR: Aortic regurgitation

The prevalence of low-flow, low-gradient (LFLG) AS was significantly higher in EH group compared to CR group (p=0.006) and CH group (p<0.001). Conversely, very severe aortic stenosis (VSAS) was less common in EH group than in CH group (p=0.005).

Baseline aortic valve ejection velocity differed significantly among the groups (p<0.001), being lower in EH group compared to both CR group (p=0.017) and CH group (p<0.001). Aortic valve area (AVA) was significantly greater in EH group than in CR group (p=0.036) and CH group (p=0.002) (p=0.002overall). Mean transvalvular gradient was also significantly lower in EH group compared to CR group (p=0.004) and CH group (p<0.001) (p<0.001 overall). Pulmonary artery pressure (PAP) was similar across groups (p=0.973).

There were no significant differences among the groups in the distribution of baseline mitral or aortic regurgitation severity (p=0.766 and p=0.474, respectively). However, bicuspid aortic valve prevalence differed significantly between groups (p=0.048), primarily due to a higher rate in EH group compared to CR group (p=0.017).

Table 3 compares procedural outcomes by groups. There were no significant differences among groups in predilatation, postdilatation, valve type used, or valve-in-valve procedures

(p>0.05). However, valve size distribution differed significantly (p=0.032); the 23 mm valve was used less frequently in EH group compared to CR group (p=0.008) and 2 (p=0.006), while the 29 mm valve was used more often in EH group than in CR group (p=0.009) and CH group (p=0.008). Device success did not differ between groups (p>0.999).

Table 3. Procedural outcomes of the patients according to groups

	CR group (n=29)	CH group (n=349)	EH group (n=35)	p-value
Predilatation	23 (79.3%)	282 (80.8%)	28 (80.0%)	0.976[a]
Postdilatation	1 (3.4%)	11 (3.2%)	0 (0.0%)	0.690[b]
Valve size				0.032[b]
20	0 (0.0%)	1 (0.3%)	0 (0.0%)	
23	15 (51.7%)[A]	148 (42.4%)[B]	6 (17.1%)[A,B]	
25	0 (0.0%)	7 (2.0%)	0 (0.0%)	
26	13 (44.8%)	149 (42.7%)	18 (51.4%)	
27	0 (0.0%)	4 (1.1%)	1 (2.9%)	
29	1 (3.4%)[A]	40 (11.5%)[B]	10 (28.6%)[A,B]	
Valve type				0.926[b]
Sapien XT	27 (93.1%)	302 (86.5%)	31 (88.6%)	
Edwards Sapien 3	2 (6.9%)	30 (8.6%)	3 (8.6%)	
Lotus	0 (0.0%)	17 (4.9%)	1 (2.9%)	
Valv in valv	0 (0.0%)	1 (0.3%)	0 (0.0%)	n/a
Procedural success	28 (96.6%)	337 (96.6%)	34 (97.1%)	>0.999[b]

a: Pearson's χ^2 test, b: Fisher-Freeman-Halton test. n/a: Not assessed. A: The difference between CR group and EH group is statistically significant (p<0.01); B: The difference between CH group and EH group is statistically significant (p<0.01). CR: Concentric remodeling, CH: Concentric hypertrophy, EH: Eccentric hypertrophy

Table 4 presents in-hospital events, discharge duration, and mortality. While in-hospital complication rates were similar across groups (p>0.05), the incidence of post-TAVI pacemaker implantation was significantly higher in EH group compared to CH group (p=0.022). Median length of hospital stay was comparable (p=0.289). No significant differences were observed in in-hospital mortality (p=0.283) or 1-year all-cause mortality rates (p=0.301).

Table 4. In-hospital and mortality outcomes by groups

	CR group (n=29)	CH group (n=349)	EH group (n=35)	p-value
Stroke	1 (3.4%)	1 (0.3%)	0 (0.0%)	n/a
Pericardial tamponade	0 (0.0%)	4 (1.1%)	1 (2.9%)	0.571[a]
Arrhytmias	4 (13.8%)	43 (12.3%)	6 (17.1%)	0.619[a]
Peripheral complications	2 (6.9%)	22 (6.3%)	4 (11.4%)	0.419[a]
Pacemaker implantation	2 (6.9%)	20 (5.7%)[A]	6 (17.1%)[A]	0.036[a]
Lenght of hospital stay (days)*	4 (2-10)	4 (2-15)	4 (2-12)	0.289[b]
In hospital mortality	0 (0.0%)	7 (2.0%)	2 (5.7%)	0.283[a]
Mortality at first year	6 (20.7%)	42 (12.0%)	3 (8.6%)	0.301[a]

Descriptive statistics are presented as *median (minimum-maximum). a: Fisher-Freeman-Halton test, b: Kruskal-Wallis test. n/a: Not assessed. A: The difference between CH group and EH group is statistically significant (p=0.022). CR: Concentric remodeling, CH: Concentric hypertrophy, EH: Eccentric hypertrophy

Finally, Table 5 presents follow-up comparisons for EF, LVEDD, mitral regurgitation (MR), and PAP. In CR group, EF did not significantly change from baseline to 1-year (p=0.273), whereas CH group (p<0.001) and EH group (p=0.002) showed significant improvement. The percentage change in EF at 1 year was significantly greater in CH group (p=0.020) and EH group (p=0.004) compared to CR group.

LVEDD significantly increased in CR group at 1 year (p<0.001), whereas it significantly decreased in EH group (p=0.006), and remained stable in CH group (p=0.141). The intergroup

Table 5. Multiple comparisons of EF, LVEDD, MR grade, and PAP by group and follow-up time points

	Basal	1-year	p-valuea	Change	p-valueb
EF					0.005
CR group	65.0 (60.0-65.0)	65.0 (60.0-65.0)	0.273	0.0 (0.0-0.0)[A,B]	
CH group	55.0 (50.0-65.0)	60.0 (50.0-65.0)	<0.001	0.0 (0.0-8.3)[A]	
EH group	32.5 (25.0-43.7)	40.0 (30.0-45.0)	0.002	2.9 (0.0-27.7)[B]	
LVEDD					<0.001
CR group	4.1 (3.7-4.2)	4.3 (3.9-4.5)	<0.001	2.9 (0.0-9.3)[A,B]	
CH group	4.6 (4.4-4.9)	4.7 (4.4-5.0)	0.141	0.0 (-1.9-2.2)[A,C]	
EH group	5.9 (5.5-6.4)	5.7 (5.1-6.3)	0.006	0.0 (-6.8-0.0)[B,C]	
MR degree					0.184
CR group	1 (0-1)	0 (0-1)	0.007	-1 (-1-0)	
CH group	1 (1-2)	1 (0-1)	<0.001	0 (-1-0)	
EH group	1 (1-2)	0 (0-1)	<0.001	-1 (-2-0)	
PAP					0.771
CR group	40.0 (20.0-55.0)	30.0 (23.7-40.0)	0.056	1.2 (0.0-5.0)	
CH group	40.0 (30.0-55.0)	35.0 (25.0-45.0)	<0.001	0.0 (-5.0-0.0)	
EH group	40.0 (30.0-55.0)	35.0 (30.0-45.0)	0.003	-5.0 (-20.0-0.0)	

Descriptive statistics are presented as median (25^{th}-75^{th} percentile). a: Within-group comparisons between baseline and 1-year values were performed using the Wilcoxon signed-rank test; results were considered statistically significant at p<0.0167 after Bonferroni correction. b: Between-group comparisons of the changes from baseline to 1 year were performed using the Kruskal-Wallis test; p<0.05 was considered statistically significant. A: The difference between CR group and CH group is statistically significant (p<0.05); B: The difference between CR group and EH group is statistically significant (p<0.01); C: The difference between CH group and EH group is statistically significant (p=0.006).
EF: Ejection fraction, LVEDD: Left ventricular end diastolic dimension, MR: Mitral regurgitation, PAP: Pulmonary artery pressure, CR: Concentric remodeling, CH: Concentric hypertrophy, EH: Eccentric hypertrophy

comparison showed a significant difference in percentage change in LVEDD (p<0.001), with a greater increase in CR group and a greater reduction in EH group compared to CH group (p=0.006).

All three groups experienced significant reductions in MR severity at 1 year (CR group: p=0.007; CH and EH groups: p<0.001), but the magnitude of change did not differ significantly among groups (p=0.184).

Regarding PAP, CR group showed no significant change (p=0.056), while CH group (p<0.001) and EH group (p=0.003) showed significant decreases. However, the intergroup comparison of PAP changes revealed no significant difference (p=0.771).

DISCUSSION

In this study, we demonstrated that baseline LV remodeling phenotypes were significantly associated with distinct clinical, echocardiographic, and procedural profiles, as well as differential reverse remodeling outcomes following TAVI. Patients in EH group were significantly younger and predominantly male. They also had higher rates of prior MI and CABG. Baseline echocardiographic evaluation revealed that EH group had the lowest EF and the highest LVEDD, LVESD, and LA. In addition, the prevalence of LFLG AS was highest and VSAS lowest in this group. EH group also showed increased frequency of bicuspid valve morphology and more frequent use of 29 mm valve prostheses. Notably, post-TAVI PPM implantation was significantly more common in this group. Regarding reverse remodeling, EH group showed significant improvement in EF and a reduction in LVEDD, while CR group demonstrated an increase in LVEDD and no meaningful change in EF. These findings emphasize the prognostic importance of baseline LV geometry in predicting myocardial response and clinical outcomes after TAVI.

EH group displayed a distinct baseline profile, both clinically and echocardiographically. These patients were significantly younger and more often male. The higher prevalence of prior MI and CABG in this group reflects a substantial ischemic burden, which may contribute to the development of eccentric remodeling through chronic myocardial injury and volume overload. This finding supports the hypothesis that adverse ventricular remodeling in EH patients may result from ischemia-related structural changes rather than isolated pressure overload. The younger age profile in this group might therefore be partially explained by earlier manifestation of ischemic cardiomyopathy. EH is typically characterized by increased LV cavity size with relatively reduced wall thickness, representing a maladaptive response to combined pressure and volume stress, unlike concentric patterns that primarily reflect pressure overload.[8] In our cohort, this pathophysiological profile was supported by echocardiographic findings: patients in EH group had the lowest EF and the largest LVEDD, LVESD, and LA diameters among all groups at baseline. These parameters are suggestive of more advanced structural decompensation and LA remodeling, possibly secondary to long-standing diastolic dysfunction or volume expansion from ischemic myocardial remodeling.[9] Furthermore, the increased prevalence of LFLG AS and the lower frequency of VSAS in this group support the notion of reduced LV contractile reserve and compromised hemodynamics.[10,11] The higher frequency of bicuspid aortic valves and the need for larger prosthetic valve sizes (e.g., 29 mm) further differentiate this phenotype anatomically.[12] Taken together, these findings indicate that EH patients undergoing TAVI represent a subgroup with more pronounced ventricular dilation, functional impairment, and ischemic history, all of which may influence their procedural risks and post-intervention recovery trajectories.

Previous studies have demonstrated that both concentric hypertrophy and concentric remodeling are associated with increased mortality following TAVI.[7, 13] Despite the significant differences in baseline clinical profiles, echocardiographic parameters, and myocardial remodeling patterns, overall mortality-both in-hospital and at one year-did not differ significantly among the three groups in our study. This finding suggests that, while LV geometry influences myocardial structure and functional recovery, it may not independently dictate survival in the short to intermediate term following TAVI(14). One possible explanation is that the procedure itself effectively alleviates the primary hemodynamic burden of aortic stenosis across all remodeling types, allowing comparable early survival benefits regardless of baseline ventricular phenotype.[15] Additionally, the standardized nature of the transfemoral approach and improved device performance may have contributed to consistent procedural success across groups. These results are in line with previous studies indicating that while EH and CH may be associated with greater structural compromise, TAVI can mitigate adverse prognostic features when performed successfully.[13] Nonetheless, the absence of mortality differences should be interpreted cautiously, as long-term follow-up might reveal divergence in outcomes influenced by residual ventricular dysfunction, comorbid conditions, and the degree of reverse remodeling achieved over time.

An important finding of our study was the more pronounced reverse remodeling observed in EH group compared to other remodeling phenotypes. Patients in this group demonstrated significant improvement in EF and a notable reduction in LVEDD at one year, indicating a favorable myocardial response following the relief of afterload with TAVI.[12] These results suggest that despite having more advanced structural dilation and lower baseline systolic function, EH patients retain considerable myocardial plasticity and may benefit more from ventricular unloading.[16] However, this anatomical configuration may also explain the higher rate of post-TAVI PPM implantation observed in this group. The combination of larger LV dimensions, altered septal geometry, and frequent use of larger prosthetic valves (e.g., 29 mm) may increase mechanical interaction with the conduction system, particularly the atrioventricular node and left bundle branch. The indications for pacemaker implantation were predominantly related to early post-procedural conduction disturbances. No additional device implantations were required due to late-onset high-degree AV block or symptomatic bradycardia during follow-up. While previous studies have reported a numerically higher-

but not statistically significant-rate of permanent pacemaker implantation in patients with eccentric hypertrophy, our study is the first to demonstrate a statistically significant increase in PPM incidence in this subgroup.[13,17] While improved reverse remodeling may support long-term functional recovery, the increased need for PPM represents a relevant procedural risk in this subgroup and may influence post-intervention quality of life and long-term rhythm outcomes. This duality shows the importance of individualized procedural planning and follow-up strategies in patients with EH undergoing TAVI.

Limitations

This study has several limitations that should be acknowledged. First, its retrospective and observational design may introduce selection bias and limit the ability to infer causality. Second, the single-timepoint echocardiographic classification of LV remodeling patterns does not account for dynamic changes that may occur prior to or following TAVI. Third, although echocardiographic assessments were performed according to standard guidelines, interobserver variability was not formally evaluated. Additionally, the relatively small number of patients in the concentric remodeling and eccentric hypertrophy groups may reduce the statistical power to detect subtle differences in outcomes. Finally, the follow-up period was limited to one year, which may not fully capture the long-term impact of remodeling patterns on survival, heart failure progression, and device-related complications. Prospective studies with longer follow-up and larger, multicenter cohorts are warranted to validate these findings.

CONCLUSION

As a result, our findings highlight the prognostic relevance of baseline LV remodeling phenotypes in patients undergoing TAVI. Eccentric hypertrophy was associated with more advanced ventricular dilation and impaired baseline function but demonstrated the greatest potential for reverse remodeling following valve intervention. Despite these structural and functional differences, short- and mid-term mortality rates were similar across all groups, emphasizing the procedural efficacy of TAVI regardless of remodeling pattern. However, the higher incidence of PPM implantation in the EH group indicates that anatomical considerations may still influence procedural risks. Overall, systematic assessment of LV geometry prior to TAVI may enhance patient stratification, guide valve selection, and inform post-procedural management, particularly in individuals with advanced remodeling patterns.

Take-home Messages

- Left ventricular remodeling patterns significantly influence the degree of structural recovery after TAVI.
- Patients with eccentric hypertrophy demonstrated the most pronounced reverse remodeling, reflected by greater improvements in EF and reductions in LVEDD.
- The incidence of permanent pacemaker implantation was significantly higher in the eccentric hypertrophy group.
- Despite anatomical and functional differences, 1-year mortality rates were similar across all remodeling groups.
- Pre-procedural assessment of LV geometry may enhance risk stratification and guide clinical decision-making in TAVI candidates.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was carried out with the permission of the Gazi University President's Office Ethics Committee (Date: 14.07.2025, Decision No: 2025-1243).

Informed Consent

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

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

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

Author Contributions

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

REFERENCES

1. Grossman W, Jones D, McLaurin LP. Wall stress and patterns of hypertrophy in the human left ventricle. *J Clin Invest*. 1975;56(1):56-64. doi:10.1172/JCI108079
2. Cioffi G, Faggiano P, Vizzardi E, et al. Prognostic effect of inappropriately high left ventricular mass in asymptomatic severe aortic stenosis. *Heart*. 2011;97(4):301-307. doi:10.1136/hrt.2010.192997
3. Bonow RO, Dodd JT, Maron BJ, et al. Long-term serial changes in left ventricular function and reversal of ventricular dilatation after valve replacement for chronic aortic regurgitation. *Circulation*. 1988;78(5 Pt 1): 1108-1120. doi:10.1161/01.cir.78.5.1108
4. Lindman BR, Stewart WJ, Pibarot P, et al. Early regression of severe left ventricular hypertrophy after transcatheter aortic valve replacement is associated with decreased hospitalizations. *JACC Cardiovasc Interv*. 2014;7(6):662-673. doi:10.1016/j.jcin.2014.02.011
5. Gaasch WH, Zile MR. Left ventricular structural remodeling in health and disease: with special emphasis on volume, mass, and geometry. *J Am Coll Cardiol*. 2011;58(17):1733-1740. doi:10.1016/j.jacc.2011.07.022
6. Une D, Mesana L, Chan V, et al. Clinical impact of changes in left ventricular function after aortic valve replacement: analysis from 3112 patients. *Circulation*. 2015;132(8):741-747. doi:10.1161/CIRCULATIONAHA.115.015371
7. Sato K, Kumar A, Jones BM, et al. Reversibility of cardiac function predicts outcome after transcatheter aortic valve replacement in patients with severe aortic stenosis. *J Am Heart Assoc*. 2017;6(7):e005798. doi:10.1161/JAHA.117.005798
8. Dweck MR, Boon NA, Newby DE. Calcific aortic stenosis: a disease of the valve and the myocardium. *J Am Coll Cardiol*. 2012;60(19):1854-1863. doi:10.1016/j.jacc.2012.02.093
9. Capoulade R, Clavel MA, Dumesnil JG, et al. Insulin resistance and LVH progression in patients with calcific aortic stenosis: a substudy of the ASTRONOMER trial. *JACC Cardiovasc Imaging*. 2013;6(2):165-174. doi:10.1016/j.jcmg.2012.11.004
10. Gerdts E. Left ventricular structure in different types of chronic pressure overload. *Eur Heart J Suppl*. 2008;10(suppl_E):E23-E30. doi:10.1093/eurheartj/sun015

11. Pibarot P, Dumesnil JG. Low-flow, low-gradient aortic stenosis with normal and depressed left ventricular ejection fraction. *J Am Coll Cardiol*. 2012;60(19):1845-1853. doi:10.1016/j.jacc.2012.06.051
12. Clavel MA, Berthelot-Richer M, Le Ven F, et al. Impact of classic and paradoxical low flow on survival after aortic valve replacement for severe aortic stenosis. *J Am Coll Cardiol*. 2015;65(7):645-653. doi:10.1016/j.jacc.2014.11.047
13. Rymuza B, Zbroński K, Scisło P, et al. Left ventricular remodelling pattern and its relation to clinical outcomes in patients with severe aortic stenosis treated with transcatheter aortic valve implantation. *Postepy Kardiol Interwencyjnej*. 2017;13(4):288-294. doi:10.5114/aic.2017.71609
14. Beach JM, Mihaljevic T, Rajeswaran J, et al. Ventricular hypertrophy and left atrial dilatation persist and are associated with reduced survival after valve replacement for aortic stenosis. *J Thorac Cardiovasc Surg*. 2014;147(1):362-369.e8. doi:10.1016/j.jtcvs.2012.12.016
15. La Manna A, Sanfilippo A, Capodanno D, et al. Left ventricular reverse remodeling after transcatheter aortic valve implantation: a cardiovascular magnetic resonance study. *J Cardiovasc Magn Reson*. 2013;15(1):39. doi:10.1186/1532-429X-15-39
16. Petrov G, Dworatzek E, Schulze TM, et al. Maladaptive remodeling is associated with impaired survival in women but not in men after aortic valve replacement. *JACC Cardiovasc Imaging*. 2014;7(11):1073-1080. doi:10.1016/j.jcmg.2014.06.017
17. Helder MR, Ugur M, Bavaria JE, et al. The effect of postoperative medical treatment on left ventricular mass regression after aortic valve replacement. *J Thorac Cardiovasc Surg*. 2015;149(3):781-786. doi:10.1016/j.jtcvs.2014.10.034

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1730961
J Health Sci Med. 2025;8(5):847-851

Prevalence of the neuropathic pain component in rheumatoid arthritis patients and its relationship to disease activity: a cross-sectional study

Derya Çırakoğlu
Department of Physical Therapy and Rehabilitation, Faculty of Medicine, Ordu University, Ordu, Turkiye

Cite this article as: Çırakoğlu D. Prevalence of the neuropathic pain component in rheumatoid arthritis patients and its relationship to disease activity: a cross-sectional study. *J Health Sci Med.* 2025;8(5):847-851.

Received: 30.06.2025 ◆ **Accepted**: 14.08.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: The purpose of this study is to examine the prevalence of the neuropathic pain (NP) component in patients with rheumatoid arthritis (RA) and its effects on disease activity, functional status, and quality of life.

Methods: The sample of the study consisted of 120 individuals, including 60 patients diagnosed with RA and 60 age- and sex-matched controls. The disease activity score-28 (DAS-28) for disease activity, the Visual Analog Scale (VAS) for pain, the painDETECT questionnaire (PDQ) for NP, the Rheumatoid Arthritis Quality of Life Questionnaire (RA-QoL) for quality of life, and the Health Assessment Questionnaire (HAQ) for functional status were used to collect data.

Results: The prevalence of NP was 63.3% (n=38) in the patient group and 6.7% (n=4) in the control group, and the difference between the two groups was statistically significant ($p<0.001$). The patient group was divided based on their PDQ scores into two subgroups consisting of those with unclear or likely NP symptoms (PDQ≥13) and those without NP symptoms (PDQ<13). The VAS-pain, DAS-28, HAQ, and RA-QoL scores of the subgroups with and without symptoms showed statistically significant differences ($p<0.001$). The PDQ scores of the patients were positively correlated with their DAS-28, VAS-pain, RA-QoL, and HAQ scores ($p<0.001$).

Conclusion: In addition to nociceptive pain, RA patients also have non-negligible rates of NP. NP is associated with high disease activity, low quality of life, and limited functional capacity.

Keywords: Rheumatoid arthritis, neuropathic pain, pain-DETECT questionnaire, HAQ

INTRODUCTION

Rheumatoid arthritis (RA) is a chronic, progressive, and inflammatory disease that starts in synovial tissue and leads to damage in the joints at later stages.[1] In RA cases, keeping the disease under control and in remission requires strict follow-up and management.[2] In most patients, inflammation can be effectively managed using disease modifying anti-rheumatic drugs (DMARDs) and the latest generation of agents.[1,3] However, in some patients, pain may persist despite objective signs such as a decrease in acute phase reactant levels and the number of swollen joints.[4] In addition to this, pain scores of some patients decline after starting treatment but plateau at a certain point.[5] Both situations indicate the presence of different pain mechanisms other than the nociceptive pain that develops due to synovial inflammation.[1,3] Recent studies have provided evidence of the possibility of a neuropathic component of pain in rheumatic diseases.[6]

Neuropathic pain (NP) is a form of pain caused by a dysfunction in the somatosensory nervous system due to a lesion or disease. While describing NP, patients report symptoms such as burning, stabbing, tingling, and numbness.[6] In recent studies, it was reported that pain could have a neuropathic component in some rheumatic diseases including RA.[7,8] The underlying factors of the etiology of NP in rheumatic diseases have not yet been identified completely.[9] Some researchers stated that in RA, NP could be caused by central sensitization.[4,6] Moreover, in electroneuromyography examinations of RA patients with NP conducted in a previous study, neuropathies were identified in 48.5% of the patients, and it was concluded that peripheral mechanisms could be effective in the etiology of NP.[8] Considering that pain persists even in RA patients with completely suppressed inflammation and halted joint damage, central or peripheral mechanisms may contribute to this pain.[4,10]

In this study, we aimed to determine the prevalence of the NP component of pain in patients diagnosed with RA using the painDETECT questionnaire (PDQ) and examine the relationships between NP and disease activity, functional status, and quality of life.

Corresponding Author: Derya Çırakoğlu, drderya79@gmail.com

METHODS

Participants

Approval to conduct the study was obtained from the Ordu University Non-interventional Scientific Researches Ethics Committee (Date: 11.04.2025, Decision No: 2025/117). All procedures were carried out in accordance with the ethical rules and the principles of the Declaration of Helsinki.

This study was planned with a cross-sectional design in which the data of patients diagnosed with RA were obtained from their patient files. Patients who had been diagnosed with RA according to the ACR/EULAR 2010 classification criteria and were being followed up as outpatients at the Physical Medicine and Rehabilitation Clinics between January 2023 and March 2025 were included in the study.[11] Disease duration, sex, age, swollen joint count (SJC), tender joint count (TJC), complete blood count results, erythrocyte sedimentation rate (ESR) values, and C-reactive protein (CRP) values were recorded.

Inclusion and exclusion criteria: Patients who had a follow-up duration of at least one year with the diagnosis of RA and were 18-65 years old were included. The sample of the study excluded those with diabetes mellitus, thyroid disease, peripheral or central nervous system diseases or lesions, or psychiatric disorders and those who were using antidepressant or antiepileptic drugs. The control group included individuals who were 18-65 years old, did not have any health problems, were not using any medication, and agreed to participate in the study.

II. Clinical Assessments

Disease activity score in 28 joints (DAS-28): DAS-28 was used to evaluate disease activity in the patient group. DAS-28 scores are calculated based on the number of tender and swollen joints, CRP or ESR values, and the global self-assessment of the patient.[12]

Pain DETECT Questionnaire (PDQ): Neuropathic symptoms were evaluated using PDQ. The questionnaire consists of seven sensory items reflecting the characteristics of pain, one item reflecting the course of the pain, and one item reflecting the direction of the pain. PDQ scores vary from -1 to 38. Scores are categorized as unlikely NP for <13, uncertain NP for 13-18, and likely NP for >18.[13] The severity of pain was evaluated based on a 10-cm version of the Visual Analog Scale (VAS).

Health Assessment Questionnaire (HAQ): The functional statuses of the patients were evaluated using HAQ. The scale consists of 20 items and the following categories; dressing and grooming, arising, eating, walking, hygiene, reach, grip, and common daily activities. Its maximum score is 60, and higher scores indicate poorer functional status.[14]

Rheumatoid Arthritis Quality of Life (RA-QoL) Questionnaire: This disease-specific scale was used to identify the quality of life of the patients. RA-QoL consists of 30 questions with the response options of yes (1) and no (0). Scores range from 0 to 30, with higher scores indicating a poorer quality of life for the patient.[15]

Statistical Analysis

The sample size was determined based on the literature on the prevalence of NP and associated factors in patients with RA. With 90% power, a type I error rate (α) of 0.05, and a large effect size (1.07) according to Cohen's criteria, the estimated sample size was 25 participants per group.

All statistical analyses were carried out using the SPSS (20.0, Inc., Chicago, Illinois, USA) program. The Kolmogorov-Smirnov test was used to determine whether the data were normally distributed. Mean and standard deviation values were calculated. The Mann-Whitney U test for the non-normally distributed data and student's T test for the normally distributed data were used to compare two independent groups of variables. Chi-squared tests were used to compare categorical variables. Relationships between variables were analyzed using Spearman's correlation analysis. p-values of <0.05 were accepted as statistically significant.

RESULTS

The patient group consisted of 50 women (83.3%) and 10 men (16.7%), while the control group consisted of 47 women (78.3%) and 13 (21.7%) men. There was no statistically significant difference between the patient and control groups in terms of their sex distributions (p=0.487). The mean ages of the participants were 45.82±12.66 (19-65) in the patient group and 42.02±12.20 (22-65) in the control group, and the groups did not differ significantly from each other in terms of age (p=0.097). While 38 of the 60 patients (63.3%) were found to have NP, 4 participants in the control group (6.7%) had NP. The difference between the groups was significant (p<0.001). There was also a significant difference between the VAS-pain scores of the patient and control groups (p<0.001). The characteristics and scale scores of the patient and control groups are shown in Table 1. The demographic and disease-related characteristics of the participants are presented in Table 2.

Table 1. Descriptive characteristics and scale scores of the participants

	Patient group	Control group	p
Age (years)	45.82±12.66	42.02±12.20	0.487
Sex (F/M)	50/10	47/13	0.097
VAS-pain score	4.85±2.82	1.32±1.77	<0.001
PDQ score	14.32±8.64	3.10±4.22	<0.001
NP rate	63.3%	6.7%	<0.001

F/M: Female/male, VAS: Visual Analog Scale, PDQ: Pain-DETECT Questionnaire, NP: Neuropathic pain

The patient group was divided into two subgroups consisting of those with unclear or likely NP symptoms (PDQ≥13) and those without NP symptoms (PDQ<13), and there was no statistically significant difference between the two subgroups in terms of their age, sex, or disease duration (respectively, p=0.096, p=0.539, p=0.872). Primary-middle school graduates constituted 63.2% of the patients with NP. The rate of high school graduates was higher among those without NP. The difference between the two subgroups based on their education levels was significant (p=0.034). The VAS-

Table 2. Characteristics of patients with RA (n=60)	
Sex (F/M)	50/10
Age (years)	45.82±12.66
Disease duration (years)	9.97±7.69
Education level, n (%)	
Primary-middle school	51.7 (31)
High school	33.3 (20)
University	15.0 (9)
RA medication, n (%)	
DMARD	53.3 (32)
Biologics	21.7 (13)
DMARD+biologics	25.0 (15)
Prednisolone dose (mg/day)	3.95±3.09
DAS-28	3.61±1.28
VAS-pain	4.85±2.82
SJC	1.62±2.26
TJC	2.43±3.12
PDQ	14.32±8.64
HAQ	8.3±7.57
RA-QoL	12.08±8.24

RA: Rheumatoid arthritis, F/M: Female/male, DMARD: Disease-modifying antirheumatic drug, DAS-28: Disease activity score, VAS: Visual Analog Scale, SJC: Swollen joint count, TJC: Tender joint count, PDQ: Pain-DETECT Questionnaire, HAQ: Health Assessment Questionnaire, RA-QoL: Rheumatoid Arthritis Quality of Life- Questionnaire

Table 3. Comparison of descriptive characteristics and scale scores of the patient subgroups with uncertain or likely NP and with unlikely NP				
		PDQ≥13 (n=38)	PDQ<13 (n=22)	p
Age (years)		48.03±11.37	42.00±14.08	0.096
Sex (F/M)		32/6	18/4	0.539
Disease duration (years)		10.05±7.82	9.82±7.63	0.872
Education level n (%)	Primary-middle school	24 (63.2%)	7 (31.8%)	0.034
	High school	11 (28.9%)	9 (40.9%)	
	University	3 (7.9%)	6 (27.3%)	
RA medication n (%)	DMARD	17 (44.7%)	15 (68.2%)	0.004
	Biologic	9 (23.7%)	4 (18.2%)	
	DMARD+biologics	12 (31.6%)	3 (13.6%)	
Prednisolone dose (mg/day)		4.27±3.28	3.40±2.73	0.453
SJC		2.24±2.51	0.55±1.14	0.451
TJC		3.08±3.15	1.32±2.80	0.013
VAS-pain		6.34±2.13	2.27±1.85	<0.001
DAS-28		4.14±1.12	2.68±0.99	<0.001
HAQ		11.18±7.80	3.32±3.59	<0.001
RA-QoL		15.11±7.99	6.86±5.74	<0.001

NP: Neuropathic pain, PDQ: Pain-DETECT Questionnaire, F/M: Female/male, RA: Rheumatoid arthritis, DMARD: Disease-modifying antirheumatic drug, SJC: Swollen joint count, TJC: Tender joint count, VAS: Visual Analog Scale, DAS-28: Disease activity score, HAQ: Health Assessment Questionnaire, RA-QoL: Rheumatoid Arthritis Quality of Life Questionnaire

pain, DAS-28, HAQ, and RA-QoL scores of the subgroups with and without symptoms showed statistically significant differences (all p<0.001). In the subgroup with NP, the patients receiving a combination of biologic agents and DMARDs were significantly more prevalent than those receiving monotherapy (p=0.004). No significant difference was observed between the groups in terms of their prednisolone dosages (p=0.453). While no significant difference was found between the subgroups in terms of SJC (p=0.451), there was a significant difference in terms of TJC (p=0.013). The results of the analyses of the scale scores and subgroups of the patient group are shown in Table 3.

The PDQ scores of the patients had positive correlations with their DAS-28 (p<0.001), VAS-pain (p<0.001), RA-QoL (p<0.001), and HAQ (p<0.001) scores, while there was no significant correlation between their PDQ scores and disease durations (p=0.613). The degrees of relationships between the PDQ scores of the patients and their clinical parameters, as well as correlation coefficients, are given in Table 4.

Table 4. Relationships between PDQ scores and clinical parameters in the patient group			
	r	p	
DAS-28	0.582	<0.001	Moderate
VAS-pain	0.735	<0.001	Strong
Disease duration (years)	-0.067	0.613	No correlation
HAQ	0.650	<0.001	Strong
RA-QoL	0.620	<0.001	Strong

PDQ: Pain-DETECT Questionnaire, DAS-28: Disease activity score, VAS: Visual Analog Scale, HAQ: Health Assessment Questionnaire, RA-QoL: Rheumatoid Arthritis Quality of Life Questionnaire, r: Spearman's rank correlation coefficient

DISCUSSION

In this study, we showed that the pain experienced by RA patients had a neuropathic component. The presence of NP in the patients was also associated with high disease activity, low quality of life, and poor functional status. The prevalence of NP in RA patients was reported by Koca et al.[1] as 60.3%. The authors concluded that both central sensitization and NP were associated with disease activity. In our study, in agreement with the literature, the prevalence of NP in RA patients was found to be 63.3%, and this rate was much higher than the rate in the control group. In a recent study, the presence of NP in RA and osteoarthritis patients was investigated, and it was seen that there was an NP component in both patient groups.[9] On the other hand, while some studies showed NP presence in RA cases, these rates were reported to be low. These varying results may have originated from differences in the racial or regional characteristics of participants. Such that, as opposed to the case in Turkish patients demonstrated in this study, studies conducted with Japanese patients revealed low rates of NP.[6,16] The most important problem of RA patients is chronic pain. The International Association for the Study of Pain categorizes pain as nociceptive, neuropathic, and nociplastic pain. This classification aims to increase the likelihood of success in treatment by performing the appropriate treatment for patients with chronic pain.[17] In RA cases, nociceptive pain is caused by inflammation and/or varying degrees of damage in the joint. In addition to this, damage in nociceptive nerve endings may lead to NP by affecting the central or peripheral nervous system.[18]

The results of our study demonstrated that factors such as disease duration, sex, and age in RA cases were not associated with the NP component. This result was compatible with the literature.[6,16] On the other hand, in our study, pain levels assessed using VAS-pain and disease activity levels assessed using DAS-28 were significantly higher in the subgroup of RA patients with NP in comparison to those in the subgroup of RA patients without NP. In a previous study that included RA patients with DAS-28 scores greater than 5.1, VAS-pain, PDQ, and quantitative sensory testing were applied respectively to evaluate potential nociceptive, neuropathic, and nociplastic pain components, and it was concluded that different pain components could be seen simultaneously in RA cases.[2] A prospective study of 567 patients revealed that although there was a decrease in the disease activity levels of patients measured based on DAS-28 scores following DMARD treatment, the VAS-pain scores of 22.6% of the patients were still high. It was argued that this result could be associated with the presence of an NP component. NP prevents patients from meeting remission criteria by affecting their global self-assessments.[18] The potential presence of NP should be considered in cases of high DAS-28 scores caused by the poor subjective global self-assessments of RA patients despite their low levels of inflammatory markers. Consistent with previous studies, our results showed no significant association between the NP component and SJC, whereas TJC, which is another subjective measure reported by the patient, was significantly higher in the NP subgroup.[6,19,20] This suggested that patients with NP may perceive joint tenderness more intensely. It is well known that TJC directly influences DAS-28 scores, similar to the global self-assessment of the patient. With treatments provided to alleviate or resolve NP, high disease activity can be prevented, and there may no longer be a need for a change or dose increase in the treatments of patients for nociceptive pain. Thus, the chance of treatment success can be increased by identifying the type of pain and providing the appropriate treatment accordingly in RA patients.[5]

In line with the literature, in our study, we observed significantly higher NP scores among RA patients with lower education levels.[19] This indicated that education levels may influence the experience of pain through factors such as pain perception and treatment adherence. Considering that psychosocial factors like stress and anxiety may be more prevalent in this group, these elements could further affect pain perception. Therefore, in managing NP among RA patients, it may be beneficial to develop individualized approaches and supportive educational programs that take education levels into account.

The quality of life of RA patients is lower compared to healthy individuals, and this situation is closely related to disease activity, pain intensity, and physical functions.[21] Noda et al.[6] evaluated health-related quality of life and physical functions in RA patients and reported that NP symptoms affected both variables negatively. In a recent study by Büyük et al.,[22] NP was assessed using multiple NP questionnaires, and high NP scores in RA patients were discovered to be associated with RA-QoL and HAQ scores. In our study, we evaluated the quality of life of the patients using RA-QoL and their functional status using HAQ. Similarly, we showed that the NP component of pain in RA patients affected both parameters negatively. Pain is the main complaint of RA patients. Patients may suffer from pain that is felt in the form of multiple attacks in a month or even a day. The inconsistence of pain in RA may affect the quality of life and daily activities of RA patients to a substantial extent.[5]

In our study, NP was more prevalent among patients receiving combined treatment with biologic agents and DMARDs. This may be due to chronic pain being misclassified as nociceptive pain, leading to escalation to combination therapy. Indeed, it was stated that even when disease activity is controlled in RA patients undergoing DMARD dose adjustments or combination therapy including biologics, chronic pain may persist.[19] NP has been identified as a potential cause of such treatment-resistant pain.[18,19,23] Treatments used in RA are costly and carry significant side effects. Moreover, these therapies are designed to suppress inflammation responsible for nociceptive pain and have no place in NP management.[20] A detailed evaluation of nociceptive and NP components, the development of patient-specific treatment strategies, and perhaps a combined approach targeting both types of pain may enhance patient satisfaction and quality of life while reducing functional limitations.[23,24]

Limitations

First, while no patients in the sample had a previous diagnosis of fibromyalgia or a history of antiepileptic or antidepressant drug use, no specific assessment of fibromyalgia was made in the study. Fibromyalgia symptoms may present similarly to NP symptoms. Second, because our study was cross-sectional, we could not make assessments of NP at the time of RA diagnosis or before the initiation of RA treatment. We believe that assessments to be made at the time of first diagnosis would help us identify the presence of NP in addition to nociceptive pain.

CONCLUSION

NP accompanies nociceptive pain in a significant proportion of RA patients. This neuropathic component is associated with high disease activity, low quality of life, and poor functional status. In our opinion, in cases where pain cannot be managed, evaluating the neuropathic component of pain before increasing or changing the treatment provided for nociceptive pain will be beneficial.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was carried out with the permission of the Ordu University Non-interventional Scientific Researches Ethics Committee (Date: 11.04.2025, Decision No: 2025/117).

Informed Consent

All patients signed and free and informed consent form.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

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

Author Contributions

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

REFERENCES

1. Koca TT, Göğebakan H, Çetin G. Central sensitization in rheumatoid arthritis: a cross-sectional analytic study. *Cent Asian J Med Hypotheses Ethics*. 2025:6(1):56-62. doi:10.47316/cajmhe.2025.6.1.06
2. Ahmed L, Biddle K, Blundell A, et al. Assessing the effects of distinct biologic therapies on rheumatoid arthritis pain by nociceptive, neuropathic and nociplastic pain components: a randomised feasibility study. *Pilot Feasibility Stud*. 2024;10(1):77. doi:10.1186/s40814-024-01505-4
3. Nieto FR, Clark AK, Grist J, Hathway GJ, Chapman V, Malcangio M. Neuron-immune mechanisms contribute to pain in early stages of arthritis. *J Neuroinflammation*. 2016;13(1):96.
4. Saitou M, Noda K, Matsushita T, Ukichi T, Kurosaka D. Central sensitisation features are associated with neuropathic pain-like symptoms in patients with longstanding rheumatoid arthritis: a cross-sectional study using the central sensitisation inventory. *Clin Exp Rheumatol*. 2022;40(5):980-987. doi:10.55563/clinexprheumatol/msy022
5. McWilliams DF, Walsh DA. Pain mechanisms in rheumatoid arthritis. *Clin Exp Rheumatol*. 2017;35 Suppl 107(5):94-101.
6. Noda K, Tajima M, Oto Y, et al. How do neuropathic pain-likesymptoms affect health-related quality of life among patients with rheumatoid arthritis? A comparison of multiple pain-related parameters. *Mod Rheumatol*. 2020;30(5)828-834. doi:10.1080/14397595.2019.1650462
7. Atik S, Sahin O, Atik I, Polat M. Neuropathic pain component in patients with ankylosing spondylitis and the relationship of neuropathic pain and disease activity parameters: a cross-sectional study. *J Musculoskelet Neuronal Interact*. 2024;24(3):284-290.
8. de Araújo Pereira F, de Almeida Lourenço M, de Assis MR. Evaluation of peripheral neuropathy in lower limbs of patients with rheumatoid arthritis and its relation to fall risk. *Adv Rheumatol*. 2022;62(1):9. doi:10.1186/s42358-022-00238-3
9. Mahajan A, Mruthyunjaya P, Padhee S, Ahmed S. People with osteoarthritis have a higher component of neuropathic pain as compared to those with rheumatoid arthritis: a cross-sectional study. *Rheumatol Int*. 2024;44(10):2179-2184. doi:10.1007/s00296-024-05657-7
10. Silva Santos Ribeiro P, Willemen HLDM, Eijkelkamp N. Mitochondria and sensory processing in inflammatory and neuropathic pain. *Front Pain Res (Lausanne)*. 2022;3:1013577. doi:10.3389/fpain.2022.1013577
11. Aletaha D, Neogi T, Silman AJ, et al. 2010 rheumatoid arthritis classification criteria: an American college of rheumatology/european league against rheumatism collaborative initiative. *Ann Rheum Dis*. 2010;69:1580-1588. doi:10.1136/ard.2010.138461
12. Smolen JS, Breedveld FC, Schiff MH, et al. A simplified Disease Activity Index for rheumatoid arthritis for use in clinical practice. *Rheumatol*. 2003;42(2):244-257. doi:10.1093/rheumatology/keg072
13. Alkan H, Ardic F, Erdogan C, Sahin F, Sarsan A, Findikoglu G. Turkish version of the painDETECT questionnaire in the assessment of neuropathic pain: a validity and reliability study. *Pain Med*. 2013;14(12):1933-1943. doi:10.1111/pme.12222
14. Küçükdeveci AA, Sahin H, Ataman S, Griffiths B, Tennant A. Issues in cross-cultural validity: example from the adaptation, reliability, and validity testing of a Turkish version of the Stanford Health Assessment Questionnaire. *Arthritis Rheum*. 2004;51(1):14-19. doi:10.1002/art.20091
15. Kutlay S, Küçükdeveci A, Gönül D, Tennant A. Adaptation and validation of the Turkish version of the Rheumatoid Arthritis Quality of Life Scale. *Rheumatol Int* 2003;23(1):21-26. doi:10.1007/s00296-002-0247-2
16. Ito S, Kobayashi D, Murasawa A, Narita I, Nakazono K. An analysis of the neuropathic pain components in rheumatoid arthritis patients. *Intern Med*. 2018;57(4):479-485. doi:10.2169/internalmedicine.9235-17
17. Murphy AE, Minhas D, Clauw DJ, Lee YC. Identifying and managing nociplastic pain in ındividuals with rheumatic diseases: a narrative review. *Arthritis Care Res (Hoboken)*. 2023;75(10):2215-2222. doi:10.1002/acr.25104
18. Baerwald C, Stemmler E, Gnüchtel S, et al. Predictors for severe persisting pain in rheumatoid arthritis are associated with pain origin and appraisal of pain. *Ann Rheum Dis*. 2024;83(10):1381-1388. doi:10.1136/ard-2023-225414
19. Kimyon G, Kara B, Akkan M, Ergin M. Do biological therapies have any effect on neuropathic pain in rheumatoid arthritis? What are the related factors? *Rheumatology Quarterly*. 2024;2:4. doi:10.4274/qrheumatol.galenos.2024.02886
20. Koop SM, ten Klooster PM, Vonkeman HE, et al. Neuropathic-like pain features and cross-sectional associations in rheumatoid arthritis. *Arthritis Res Ther*. 2015;17(1):237. doi:10.1186/s13075-015-0761-8
21. Hassen N, Moolooghy K, Kopec J, Xie H, Khan KM, Lacaille D. Determinants of health-related quality of life in adults living with rheumatoid arthritis: a systematic review. *Semin Arthritis Rheum*. 2025. doi:10.1016/j.semarthrit.2025.152717
22. Büyük F, Borman P, Bakırcı EŞ, et.al. The frequency of neuropathic pain and associated factors in patients with rheumatoid arthritis. *Aktuelle Rheumatologie*. 2025;50(03):177-183. doi:10.1055/a-2243-3584
23. Motyl G, Krupka WM, Maślińska M. The problem of residual pain in the assessment of rheumatoid arthritis activity. *Reumatologia*. 2024;62(3):176-186. doi:10.5114/reum/189779
24. Lassen J, Baron R. Rheumatic diseases and neuropathic pain: diagnosis and treatment. *Z Rheumatol*. 2021;80(3):226-233. doi:10.1007/s00393-020-00950-9

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1744007
J Health Sci Med. 2025;8(5):852-858

The effect of immersive virtual reality and music on anxiety, fear, and pain during circumcision surgery in children: a randomized controlled study

Hamide Nur Çevik Özdemir[1], Ahmet Ali Tuncer[2], İbrahim Kılıç[3]

[1]Department of Pediatric Nursing, Faculty of Health Sciences, Afyonkarahisar Health Sciences University, Afyonkarahisar, Turkiye
[2]Department of Pediatric Surgery, Faculty of Medicine, Afyonkarahisar Health Sciences University, Afyonkarahisar, Turkiye
[3]Department of Biostatistics, Faculty of Veterinary Medicine, Afyon Kocatepe University, Afyonkarahisar, Turkiye

Cite this article as: Çevik Özdemir HN, Tuncer AA, Kılıç İ. The effect of immersive virtual reality and music on anxiety, fear, and pain during circumcision surgery in children: a randomized controlled study. *J Health Sci Med.* 2025;8(5):852-858.

Received: 16.07.2025 ◆ **Accepted**: 14.08.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: Circumcision is a surgical procedure that causes pain, anxiety, and fear in children. This study aimed to examine the effectiveness of immersive virtual reality (IVR) and music on anxiety, fear, and pain levels of children during circumcision surgery.

Methods: This randomized controlled trial included boys aged 6-8 years who were circumcised at a university hospital between September 2022 and July 2023. Using stratified block randomization, the sample group's children were split into three groups: music (n=24), IVR (n=24), and control group (n=24). The Wong-Baker Faces Pain Rating Scale, the Children's Anxiety Meter Scale, the Children's Fear Scale, and a Participant Information Form were used to gather the data. Descriptive statistics, t tests, analysis of variance and Duncan and Bonferroni adjustments were used to examine the data.

Results: Children in the IVR and music groups experienced much less worry and dread during and after circumcision than those in the control group. The mean pain scores of the experimental groups during and after circumcision were significantly lower than those of the control group. Fear, anxiety, and pain levels were significantly different in the IVR group compared to the other groups ($p<0.05$).

Conclusion: Children's anxiety, fear, and pain levels were successfully decreased by the use of IVR and music interventions during the circumcision operation. Health professionals, nurses, can use effective nonpharmacological strategies, such as music and IVR, to manage fear, anxiety and pain associated with surgical procedures in children.

Keywords: Child, virtual reality, pain management, music, surgery

INTRODUCTION

Circumcision, one of the oldest and most common surgeries for children, is performed for medical, religious, and cultural reasons. Despite being seen as a straightforward medical surgery, children must be emotionally and cognitively ready for the circumcision process.[1,2]

Circumcision is painful for children even under general and local anesthesia.[3,4] Due to the risks of general anesthesia, the American Academy of Pediatrics recommends local anesthesia for newborns, infants, and compliant children.[5] Local anesthesia avoids complications like respiratory depression and urinary retention, which is why children over six are often circumcised under local anesthesia. During the procedure, children are aware and feel pain.[4,6] Circumcision is better tolerated when performed by experienced practitioners under sterile conditions with appropriate pain management.[5]

During the preoperative phase, children feel scared and anxious. Because they are not informed about the circumcision procedure, receive anesthesia, and cannot express their pain.[6,7] Previous studies have reported that children experience fear, anxiety, and pain during and after circumcision.[4,8] This has negative effects on the child, such as prolonged postoperative recovery time, decreased compliance with treatment and hospitalization, and posttraumatic stress.[7, 8]

Nonpharmacologic interventions have become more significant in lowering children's anxiety and fear of surgery. In addition to many nonpharmacological methods, such as therapeutic play, puppet shows, video games, cartoons, and music, patient education and operating room visits are also used to reduce preoperative anxiety.[9-13] Distraction is a nursing intervention that increases pain tolerance in children by distracting them from painful stimuli.[12,13]

Corresponding Author: Hamide Nur Çevik Özdemir, hamidenur43@hotmail.com

Immersive virtual reality (IVR) has become a popular distraction method in recent years. It involves the user actively engaging in a virtual environment through a head-mounted screen [14] and can be used anytime in clinical settings. Evidence supports IVR's effectiveness in reducing anxiety, stress, and pain in children patients receiving needle procedures, burn care, or cancer treatment.[14,15] Studies have shown that preoperative use of virtual reality (VR) positively impacts anxiety and postoperative pain.[6,7,16,17]

Musical interventions help children cope with pain and stress. Studies in the literature reveal that musical interventions during surgical procedures positively affect children's levels of pain, fear, and anxiety.[18-20]

Even though the consequences of distraction techniques or VR on fear and pain before circumcision have been examined in the literature, studies on their use during circumcision surgeries are limited.[4,6] There is a need to investigate the benefits of innovative distraction techniques such as IVR and music that actively engage the child, which may help the child cope more effectively with the pain and anxiety associated with the circumcision procedure, as these techniques can significantly impact the circumcision experience and recovery process.

This study's objective was to evaluate how music and an IVR intervention affected the children's pain, anxiety, and terror during circumcision surgery. Within the scope of this main objective, the hypotheses of the study were as follows:

- **H1:** IVR and music interventions have a significant effect on lowering the anxiety levels of children undergoing the circumcision procedure.
- **H2:** IVR and music interventions have a significant effect on lowering the fear levels of children undergoing the circumcision procedure.
- **H3:** IVR and music interventions have a major impact on decreasing pain in children undergoing circumcision surgery.

METHODS

Ethical Considerations

The study received ethical approval from the Afyonkarahisar Health Sciences University Non-interventional Clinical Researches Ethics Committee (Date: 05.08.2022, Decision No: 2022/9), and the implementation permission was obtained from the university research hospital. Verbal and written consent was obtained from participating children and their families. This study was retrospectively registered at ClinicalTrials.gov. All procedures were carried out in accordance with the ethical rules and the principles of the Declaration of Helsinki.

Study Design and Setting

This randomized controlled study was conducted in the pediatric surgery clinic of a university research hospital in western Turkey between September 2022 and July 2023. The study was divided into three groups: the music (n=24), the IVR (n=24), and the control (n=24).

While the control group received only routine clinical care, the music group listened to music of their own choosing, and the IVR group played an interactive video game.

Sample Size

The study's sample size was calculated using the G-power 3.1.9.2 tool. A power of 0.80 required a minimum of 21 participants in each group, an effect size of 0.80, and a significance threshold of α=0.05. Consequently, a total of 72 children-24 in each group-were included in the study, taking into account the possibility that some children might be excluded.

Randomization

An independent researcher used Random Sequence Generator (www.randomizer.org), a computer software, to divide the children into groups. Each volunteer participant who met the criteria was given a number between 1 and 72. This list was only given to the coresearcher during the application. Using stratified block randomization (by age), the children in the sample group were split into three groups: the IVR, the music, and the control group. There were 24 children in each group. The control and experimental groups were created to be homogeneous in terms of age and hospitalization history (p>0.05). Interventions were initiated after local anesthesia. The treatment of the children did not occur during the data collection phase of the study. The study's CONSORT flow diagram is displayed in the **Figure**.

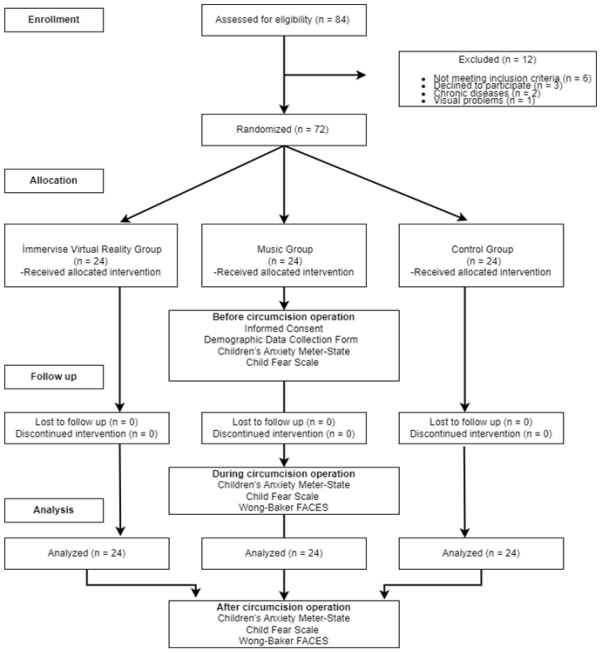

Figure. CONSORT diagram

Participants

A history of circumcision in a pediatric surgery clinic, consent to participate in the study, age between 6 and 8 years, no history of surgical intervention, no persistent discomfort, and no mental health issues were the requirements for inclusion. The following were the exclusion criteria: using any kind of pain medication 24 hours prior to the procedure,

having communication issues, or having any of the following contraindications to the procedure: hypopadias, epispadias, bleeding disorders, or anatomical diseases.

Data Collection Tools

Participant Information Form: The questionnaire asked about the children's and their families' sociodemographic details, such as age, prior hospitalization, education, decision to have circumcised, etc.

Children's Anxiety Meter Scale: Children's anxiety levels during invasive procedures are measured using this scale in healthcare settings. The scale yields scores ranging from 0 to 10.[21] An increase in the score indicates that children's concerns and anxiety increase. The Turkish validity of the scale was established in children aged between 4 and 10 years.[22] This study's Cronbach's alpha was 0.80.

Children's Fear Scale: This single-item self-report scale is used to measure pain-related fear in children. Five gender-neutral facial expressions make up the scale, which ranges from neutral on the far left to highly afraid on the far right. The range of points awarded for facial expressions is 0-4. The scale can be used by parents and observers before, during, and after medical procedures for children aged 5-10 years.[23] The Turkish validity of the scale was established.

Wong-Baker Faces Pain Rating Scale: This scale assesses pain in children aged 3 to 18. It was developed by Wong and Baker. The numerical rating scale has scores ranging from 0 to 10. A range of faces is displayed on the scale, from a crying face (10=great pain) to a smiling face (0=very happy/no pain).[24]

Research Instruments

Virtual reality glasses: In this investigation, the white "Meta Oculus Quest 2" virtual reality headset was utilized. Two Oculus touch controllers and a head-mounted display (HMD) are included with the Oculus Quest 2.The VR HMD is an independent device with built-in tracking and headphones.[25]

Music headphones: In this study, we used wireless music headphones with Bluetooth and stereo high-resolution sound quality. It is compatible with Android devices such as iPads and can cancel noise. It is a music headphone that complies with the World Health Organization's 75 dB sound level recommendation for children.[26]

Preintervention Phase

Pilot study: Nine children with the identical sample group characteristics participated in a pilot trial before the intervention. These children didn't have any bad to say about the IVR programs. The study sample did not include children from the pilot study group.

Intervention protocols: On the morning of the surgery, the children were admitted to the clinic early in the morning and prepared according to standard protocol. One of the researchers briefed the children and their families about the study when they arrived at the clinic right after being admitted to the hospital, and both verbal and written consent were acquired. The CAM-S, the CFS, and an introductory information form were then used to gather the data. Children meeting the inclusion criteria were randomly allocated to groups.

The researcher who would carry out the intervention told the children in the experimental groups after they had been brought to the operation room waiting area. The IVR group was informed about the use of VR glasses, and the music group was informed about the use of headphones. The children in the music group chose the music they wanted to listen to. Additionally, the children were told that if they had a headache, earache, eye pain, or motion sickness, the intervention would end. During the circumcision, children in the IVR group chose the VR game they wanted. For the control group, standard protocols were used.

Intervention phase: In this study, IVR was applied to one group and music intervention was applied to the other group to distract the child during the circumcision procedure. In addition to having a PhD in pediatric nursing, one of the researchers was a trained music therapy practitioner with expertise in virtual reality. This researcher conducted the interventions. In the operating room, the researcher adhered to the rules governing medical and surgical operations. The children in the IVR group were given the remote controls and VR glasses by the researcher shortly before the intervention. Music headphones were worn by the children in the group.

Circumcision surgical procedure: All circumcisions were conducted in the same operating room by the same pediatric surgeon and support team. Local anesthesia was applied to all children using the same technique. All children (in the experimental and control groups) were given bupivacaine HCl (Marcaine 0.5%; AstraZeneca, İstanbul, Turkiye) and lidocaine HCl+epinephrine (Jetokain; Adeka, İstanbul, Turkiye) local anesthesia with the penile block technique. The block was performed using markers. Interventions were initiated 20 minutes after local anesthesia. All children underwent circumcision using the thermocautery-assisted guillotine technique.[27]

The IVR group received both standard medical treatment and virtual reality gaming during the circumcision procedure. Children in the IVR group, who lay on the operating table during circumcision, played the interactive game of their choice for ten minutes.

Two Oculus Quest gallery games were made available to children. These activities have been employed in numerous studies in the literature and have been described as engaging, thrilling, and soothing for children.[7,22] Before each use, the Oculus Quest was put on a disposable silicone cushion. Before using the gadget, the patients put on a face mask and disposable headgear (personal protection equipment) to reduce direct contact. Before every intervention, the portions of the gadget that came into contact with the faces of the youngsters were cleansed.

The band listened to music at 60 dB through the same headphones and iPads. Children listened to the music they wanted with headphones for ten minutes. In previous studies, music was shown to be preferred by children.[28] Disposable ear pads were preferred, and the headphones were disinfected each time.

No intervention such as IVR or music was applied to the children in the control group during the circumcision process. Routine care was given.

An independent external observer applied the CAM-S, CFS, and WBS to the children in the experimental and control groups at the 12th minute of the circumcision procedure. The surgery was completed by the surgeon in approximately 15 minutes.

Postintervention phase: Two hours after circumcision, CAM-S, CFS, and WBS were applied again to all children in the experimental and control groups. No analgesic was given to the children between the times of scale application. The children in the control group were allowed to experience IVR and music by the researcher in the room in the clinic after the experiment, considering ethical concerns and to avoid affecting the research data.

Statistical Analysis

The IBM SPSS Statistics 23.0 package program was used for the data analysis. Descriptive statistics (frequency, percentage distribution, mean, etc.) were used for the analysis of sociodemographic data. Analysis of variance was used to compare the pain, anxiety, and fear levels of the children according to group. In addition, the differences between the procedure durations for each group were determined by repeated measures ANOVA and paired sample t tests. A post hoc evaluation was performed with Duncan and Bonferroni correction. Statistical analyses were accepted at a confidence interval of 95% and a significance level of $p<0.05$.

RESULTS

Characteristics of Children and Parents

The sociodemographic details of the children and their families are displayed in Table 1. When the participants were analyzed according to age, the mean age of the children was 7.17±0.91 years in the control group, 6.92±0.88 years in the IVR group, and 7.17±0.76 years in the music group. A total of 70.8% of the children in the control group, 45.8% of the children in the IVR group, and 66.7% of the children in the music group made the decision for circumcision together with their parents. A total of 98.8% of the parents were middle school graduates. To prevent bias in the study findings, it was ensured that there was no statistically significant difference between the groups in terms of children's age (p=0.508), parental education level (p=0.560), or decision for circumcision (p=0.171).

Table 1. Demographic characteristics of the children and parents

Variables	IVR group	Music group	Control group	Statistic test
	Mean (SD) or n (%)			
Age	6.92±0.88	7.17±0.76	7.17±0.91	F=0.683; p=0.508
Parents				
Mother	13 (54.2)	21 (87.5)	16 (66.7)	X^2=6.415; p=0.040*
Father	11 (45.8)	3 (12.5)	8 (33.3)	
Parent education				
Primary	5 (20.8)	5 (20.8)	7 (29.2)	X^2=4.878; p=0.560
Secondary	8 (33.3)	9 (37.5)	5 (20.8)	
High school	4 (16.7)	7 (29.2)	8 (33.3)	
≥University	7 (29.2)	3 (12.5)	4 (16.7)	
Circumcision decision				
Child	7 (29.2)	2 (8.3)	5 (20.8)	F=6409; p=0.171
Family	6 (25.0)	6 (25.0)	2 (8.3)	
Together	11 (45.8)	16 (66.7)	17 (70.8)	

*p<0.05, IVR: Immersive virtual reality, SD: Standard deviaiton

Anxiety Levels

Table 2 shows the test results regarding the comparison of the anxiety levels of circumcised children according to the group and procedure duration. According to the children's evaluations, there was no significant difference between the groups before the procedure (p>0.05). There were significant differences between the anxiety scores measured by both the children and the observer during and after the procedure (p<0.05). According to the mean scores, the anxiety level decreased significantly in both the IVR and music groups compared to the control group. However, although there was no statistically significant difference between the IVR group and the music group after the procedure, there was a greater decrease in the IVR group. Significant differences were detected between the procedure durations in each group (p<0.05). According to the evaluations of the children and the observer, the anxiety level was high in the IVR and music groups before the procedure, and there was a significant

Table 2. Comparison of anxiety scores according to group and procedure duration

Evaluation	Groups	Pre-procedure M±SD	During procedure M±SD	Post-procedure M±SD	p
Child reported	IVR	$3.33±0.48^{A}$	$2.91±1.50^{bB}$	$0.37±0.71^{bC}$	<0.001
	Music	$2.83±1.80^{A}$	$2.62±1.88^{bB}$	$0.76±1.16^{bC}$	
	Control	$3.20±2.02^{B}$	$4.79±2.12^{aA}$	$3.62±1.88^{aB}$	
	p	0.529	<0.001	<0.001	
Observer reported	IVR	$3.62±0.71^{aA}$	$3.00±1.53^{bB}$	$0.37±0.71^{bC}$	<0.001
	Music	$3.20±1.74^{bA}$	$2.79±1.97^{bB}$	$0.62±1.13^{bC}$	
	Control	$3.30±2.02^{bB}$	$4.79±2.12^{aA}$	$3.52±1.88^{aB}$	
	p	<0.001	<0.001	<0.001	

[a,b]There were significant differences between groups for each procedure duration; [A,B,C]There were significant differences between groups for each procedure duration. IVR: Immersive virtual reality, SD: Standard deviaiton

gradual decrease during and after the procedure. In the control group, a significant increase was observed during the procedure.

Fear Levels

The test results regarding the comparison of the fear levels of the children participating in the study according to the groups and procedure durations are given in Table 3. According to the evaluations made by the children and the observer, there was no significant difference between the groups before the procedure ($p>0.05$). In the children's and observers' evaluations, the fear levels during and after the procedure differed significantly between the groups ($p<0.05$). Both the IVR and music groups had lower scores than did the control group. In both evaluations, the fear levels significantly decreased during and after the procedure compared to before the procedure.

Pain Levels

Table 4 presents the results regarding the comparison of children's and observers' evaluations of pain according to group and procedure duration. According to both the children and observer evaluations, the pain level was much lower in the IVR and music groups than in the control group. However, while there was no significant difference between the IVR and music groups after the procedure ($p>0.05$), the IVR group had significantly lower pain levels than did the music group during the procedure.

Table 4. Comparison of pain scores according to group and procedure duration

Evaluation	Groups	During procedure M±SD	Post-procedure M±SD	p
Child reported	IVR	0.33±0.76^{c}	0.41±0.82^{b}	0.714
	Music	2.33±2.18^{b}	0.66±1.12^{b}	
	Control	3.91±1.50^{a}	2.75±1.42^{a}	<0.001
	p	<0.001	<0.001	
Observer reported	IVR	0.33±0.76^{c}	0.16±0.56^{b}	0.426
	Music	2.50±2.58^{b}	0.41±1.01^{b}	
	Control	3.91±1.50^{a}	2.75±1.42^{a}	<0.001
	p	<0.001	<0.001	

^{a,b,c}There were significant differences between groups for each procedure duration. IVR: Immersive virtual reality, SD: Standard deviaiton

DISCUSSION

This randomized controlled trial examined the effects of IVR and music on fear, anxiety, and pain during circumcision. Pain and anxiety levels were lower in the IVR and music groups than in the control group, with the IVR group showing the lowest pain scores.

IVR uniquely integrates multiple sensory experiences, creating an interactive virtual environment.[29] In this study, IVR significantly reduced children's anxiety and fear during circumcision. Consistent with previous research, IVR effectively served as a distraction to lower anxiety. Meta-analyses confirm VR's effectiveness in reducing anxiety and pain during surgery.[30,31] One study found VR animation reduced anxiety at all circumcision stages ($p<0.001$), while another reported increased anxiety at the start.[11] Biophilic VR was also shown to reduce intraoperative pain and anxiety in circumcised children.[6]

Based on the WBS findings in this study, it was determined that both IVR and music successfully reduced the level of intraoperative pain. There was a significant difference in anxiety and pain levels between the groups of circumcised children. This difference originated from the IVR group. There are very few studies in the literature evaluating VR therapies in circumcised children. In these studies, usually, the effect of classical VR on anxiety and pain before circumcision has been evaluated. In a study, it was determined that VR intervention reduced preoperative anxiety in circumcised children.[7] In another study, it was found that watching cartoons with VR glasses during circumcision reduced children's fear level.[32] In another study, it was reported that VR distraction through a guided imagery mechanism reduced children's anxiety levels during circumcision.[33] These studies were designed to be nonblinded, and bias could not be excluded. IVR is a combination of visual, auditory, and kinesthetic sensory methods. Therefore, IVR outperforms classical A/V distraction methods.[32,33] Our results align with those reported in the literature

Music provides relaxation and calmness to individuals. Studies in the literature have shown that listening to music before a surgical procedure positively affects the physiological parameters and psychological health of individuals.[19,20] In this study, listening to music during circumcision

Table 3. Comparison of fear scores according to group and procedure duration

Evaluation	Groups	Pre-procedure M±SD	During procedure M±SD	Post-procedure M±SD	p
Child reported	IVR	1.91±1.97^{A}	0.41±0.65^{bB}	0.12±0.00^{bB}	<0.001
	Music	1.50±1.14^{A}	1.12±1.15^{bA}	0.25±0.60^{bB}	
	Control	1.50±0.78^{B}	2.66±0.63^{aA}	1.83±0.86^{aB}	
	p	0.240	<0.001	<0.001	
Observer reported	IVR	1.95±0.95^{A}	0.41±0.65^{bB}	0.12±0.00^{bB}	<0.001
	Music	1.75±0.89^{A}	1.29±1.26^{bA}	0.25±0.60^{bB}	
	Control	1.50±0.78^{A}	2.66±0.63^{aB}	1.83±0.86^{A}	
	p	0.203	<0.001	<0.001	

^{a,b}There were significant differences between groups for each procedure duration; ^{A,B}There were significant differences between groups for each procedure duration. IVR: Immersive virtual reality, SD: Standard deviation

significantly reduced children's anxiety, fear, and pain levels. Although some studies have shown that the pain and anxiety responses of pediatric children exposed to different surgical interventions decrease after listening to music,[34,35] very few studies have evaluated the effect of music on pain and anxiety in circumcised children. One study reported that music reduced children's pain and anxiety levels after circumcision.[18] In another study, it was reported that intraoperative music intervention in children undergoing circumcision or inguinal hernia repair reduced the incidence of postoperative maladaptive behavior in children.[36] However, unlike our study findings, intraoperative music application did not affect postoperative pain or patient comfort. There is a need for further studies to determine the usefulness of music in reducing anxiety and pain in children during the intraoperative period.

In this study, intraoperative and postoperative pain scores were lower in the IVR and music groups than in the control group. Literature has shown that patient immersion in interactive IVRs can divert attention away from painful stimuli and reduce pain perception.[13,29] The study results are consistent with studies that have shown that IVR reduces pain symptoms.

Limitations

This study provides strong evidence of causal relationships thanks to its randomized controlled trial design. Blinding of the statistician during data analysis increased the reliability of the data. However, blinding was not implemented during the data collection phase. Data collection by an independent observer may increase the risk of observer bias in the measurements. Additionally, the study was conducted in a single center; this may limit the generalizability of the findings to different institutions and populations. Many factors, including personality traits, developmental levels, and family attitudes, influence children's perception of pain. These factors may also be a limitation of the study.

CONCLUSION

The results revealed that IVR and music were effective in reducing anxiety, fear, and pain during the intraoperative period in children. IVR can be safely used to reduce surgical procedure-related pain and anxiety in children. In future studies, it may be recommended that combinations of IVR and music be applied in larger samples. Pediatric surgeons, nurses can use IVR and music as nonpharmacological methods to increase children's comfort during surgical procedures such as circumcision.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was carried out with the permission of the Afyonkarahisar Health Sciences University Non-interventional Clinical Researches Ethics Committee (Date: 05.08.2022, Decision No: 2022/9).

Informed Consent

All patients signed and free and informed consent form.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

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

Author Contributions

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

Acknowledgments

The authors thank all the children and parents involved in this study. We would like to thank the hospital management for providing the research environment and the Scientific Research Projects Coordination Unit of Afyonkarahisar Health Sciences University for providing research tools and support.

REFERENCES

1. Many BT, Rizeq YK, Vacek J, et al. A contemporary snapshot of circumcision in US children's hospitals. *J Pediatr Surg.* 2020;55(6):1134-1138. doi:10.1016/j.jpedsurg.2020.02.031
2. Rızalar S, Büyük ET, Yıldırım N. Children's perspectives on the medical and cultural aspects of circumcision. *Iranian J Pediatr.* 2017;27:2. doi:10.5812/ijp.7561
3. Morris BJ, Krieger JN, Klausner JD. CDC's male circumcision recommendations represent a key public health measure. *Glob Health Sci Pract.* 2017;5(1):15-27. doi:10.9745/GHSP-D-16-00390
4. Suzan ÖK, Şahin ÖÖ, Baran Ö. Effect of puppet show on children's anxiety and pain levels during the circumcision operation: a randomized controlled trial. *J Pediatr Urol.* 2020;16(4):490-e1. doi:10.1016/j.jpurol.2020.06.016
5. American Academy of Pediatrics Task Force on Circumcision. Circumcision policy statement. *Pediatrics.* 2012;130(3):585-586. doi:10.1542/peds.2012-1989
6. Luo W, Chen C, Zhou W, et al. Biophilic virtual reality on children's anxiety and pain during circumcision: a randomized controlled study. *J Pediatr Urol.* 2023;19(2):201-210. doi:10.1016/j.jpurol.2022.10.023
7. Büyük ET, Odabaşoğlu E, Uzsen H, Koyun M. The effect of virtual reality on children's anxiety, fear, and pain levels before circumcision. *J Pediatr Urol.* 2021;17(4):567.e1. doi:10.1016/j.jpurol.2021.04.008
8. Zavras N, Tsamoudaki S, Ntomi V, Yiannopoulos I, Christianakis E, Pikoulis E. Predictive factors of postoperative pain and postoperative anxiety in children undergoing elective circumcision: a prospective cohort study. *Korean J Pain.* 2015;28(4):244-253. doi:10.3344/kjp.2015.28.4.244
9. Ayan G, Şahin ÖÖ. Effect of therapeutic play-based training program on pre- and postoperative anxiety and fear: a study on circumcision surgery in Turkish Muslim children. *J Pediatr Urol.* 2023;19(4):431-e1. doi:10.1016/j.jpurol.2023.04.032
10. Chow CH, Van Lieshout RJ, Schmidt LA, Buckley N. Tablet-based intervention for reducing children's preoperative anxiety: a pilot study. *J Dev Behav Pediatr.* 2017;38(6):409-416. doi:10.1097/DBP.0000000000000454
11. Giordano F, Zanchi B, De Leonardis F, Rutigliano C, Esposito F, Brienza N, Santoro N. The influence of music therapy on preoperative anxiety in pediatric oncology patients undergoing invasive procedures. *Arts Psychother.* 2020;68:101649. doi:10.1016/j.aip.2020.101649
12. Tuncay S, Tüfekci FG. The effect of nursing interventions with therapeutic play and video animations prepared with psychodrama technique in reducing fear, anxiety, and pain of children at male circumcision: a randomized controlled study. *Int J Urol.* 2023;30(7):592-599. doi:10.1111/iju.15184

13. Wang X, Liu H, Tang G, Sun F, Wu G, Wu J. The effect of distraction techniques on pain, fear, and anxiety in children undergoing circumcision: a meta-analysis of randomized controlled trials. *Am J Men's Health*. 2024;18(1):15579883241230166.
14. Chirico A, Lucidi F, De Laurentiis M, Milanese C, Napoli A, Giordano A. Virtual reality in health system: beyond entertainment. A mini-review on the efficacy of VR during cancer treatment. *J Cell Physiol*. 2016;231(2): 275-287. doi:10.1002/jcp.25117
15. Güler S, Şahan S, Öztürk S. The effect of video streaming with virtual reality glasses on pain, anxiety, and satisfaction applied in peripheral intravenous catheter process. *Pain Manag Nurs*. 2024;25(6):e472-e478. doi:10.1016/j.pmn.2024.05.009
16. Hassannia R, Sajjadi M, Shareinia H, Elmimehr R, Moravejjifar M. Effect of virtual reality on relieving pain and anxiety of circumcision in children. *Evid Based Care*. 2021;10(4):67-74.
17. Suparda DDF, Kurniawati ND, Wahyudi AS. Virtual reality games as a pain and anxiety reduction in circumcision children: a literature review. *Pediomaternal Nurs J*. 2023;9(2):85-94. doi:10.20473/pmnj.v9i2.43086
18. Bulut M, Alemdar DK, Bulut A, Şalcı G. The effect of music therapy, hand massage, and kaleidoscope usage on postoperative nausea and vomiting, pain, fear, and stress in children: a randomized controlled trial. *J Perianesth Nurs*. 2020;35(6):649-657. doi:10.1016/j.jopan.2020.03.013
19. Kühlmann AR, van Rosmalen J, Staals LM, et al. Music interventions in pediatric surgery (the music under surgery in children study): a randomized clinical trial. *Anesth Analg*. 2020;130(4):991-1001. doi:10.1213/ANE.0000000000003983
20. Yeniay D, Ozdemir AY. Effects of videos and therapeutic music on preoperative anxiety and postoperative anxiety and pain levels in boys who undergo circumcision. *J Perianesthesia Nurs*. 2023;38(6):918-924.
21. Ersig AL, Kleiber C, McCarthy AM, Hanrahan K. Validation of a clinically useful measure of children's state anxiety before medical procedures. *J Spec Pediatr Nurs*. 2013;18(4):311-319. doi:10.1111/jspn.12042
22. Gerçeker GÖ, Ayar D, Özdemir Z, Bektaş M. Gaining of children's state anxiety and children's fear scale to Turkish language. *Dokuz Eylul Univ Nurs Fac Electron J*. 2018;11(1):9-13.
23. McMurtry CM, Noel M, Chambers CT, McGrath PJ. Children's fear during procedural pain: Preliminary investigation of the Children's Fear Scale. *Health Psychol*. 2011;30(6):780. doi:10.1037/a0024817
24. Wong DL, Baker CM. Wong-Baker faces pain rating scale. *Pain Manag Nurs. 2012*.
25. Carnevale A, Mannocchi I, Sassi MSH, et al. Virtual reality for shoulder rehabilitation: accuracy evaluation of Oculus Quest 2. *Sensors (Basel)*. 2022;22(15):5511. doi:10.3390/s22155511
26. World Health Organization. Safe listening devices and systems: a WHO-ITU standard. 2019.
27. Tuncer AA, Deger M. Incidence of complications following thermocautery-assisted circumcisions. *Urol J*. 2018;15(6):359-364. doi:10.22037/uj.v0i0.4296
28. Ghasemi M, Hoseinialiabadi P, Yazdanpanah F, et al. Comparison of music and vapocoolant spray in reducing the pain of venous cannulation in children age 6-12: a randomized clinical trial. *BMC Pediatr*. 2022;22(1):1-8. doi:10.1186/s12887-022-03271-9
29. Alshatrat SM, Sabarini JM, Hammouri HM, Al-Bakri IA, Al-Omari WM. Effect of immersive virtual reality on pain in different dental procedures in children: a pilot study. *Int J Paediatr Dent*. 2022;32(2):264-272. doi:10.1111/ipd.12851
30. Eijlers R, Utens EM, Staals LM, et al. Meta-analysis: systematic review and meta-analysis of virtual reality in pediatrics: effects on pain and anxiety. *Anesth Analg*. 2019;129(5):1344. doi:10.1213/ANE.0000000000004165
31. Nordgard R, Lag T. The effects of virtual reality on procedural pain and anxiety in pediatrics: a systematic review and meta-analysis. *Front Virtual Real*. 2021;2:699383. doi:10.3389/frvir.2021.699383
32. Türkmen AS, Şahiner NC, Beşirik SA, Uysal M. The effect of virtual reality glasses against the fear of circumcision: a randomized controlled trial. *J Surg Med*. 2022;6(6):610-614. doi:10.28982/josam.993190
33. Firmansyah A, Setiawan H, Wibowo DA, Rohita T, Umami A. Virtual reality (VR) media distraction relieve anxiety level of the children during circumcision. In: 1st Paris Van Java International Seminar on Health, Economics, Social Science and Humanities (PVJ-ISHESSH 2020); 2021 Dec; Atlantis Press. p. 611-614.
34. Shahrbabaki RM, Nourian M, Farahani AS, Nasiri M, Heidari A. Effectiveness of listening to music and playing with lego on children's postoperative pain. *J Pediatr Nurs*. 2023;69:e7-e12. doi:10.1016/j.pedn.2022.11.023
35. Zhang TT, Fan Z, Xu SZ, et al. The effects of music therapy on peripherally inserted central catheter in hospitalized children with leukemia. *J Psychosoc Oncol*. 2023;41(1):76-86. doi:10.1080/07347332.2022.2044967
36. Buehler PK, Spielmann N, Buehrer S, Schmidt AR, Weiss M, Schmitz A. Intraoperative music application in children and adolescents: a pilot study. *Acta Anaesthesiol Scand*. 2017;61(8):895-903. doi:10.1111/aas.12935

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1681670
J Health Sci Med. 2025;8(5):859-864

Quantitative evaluation of cranial MRI findings in idiopathic intracranial hypertension

Fatma Kökcü
Department of Radiology, Faculty of Medicine, Tokat Gaziosmanpaşa University, Tokat, Turkiye

Cite this article as: Kökcü F. Quantitative evaluation of cranial MRI findings in idiopathic intracranial hypertension. *J Health Sci Med.* 2025;8(5):859-864.

Received: 22.04.2025 ◆ **Accepted**: 17.08.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: Idiopathic intracranial hypertension (IIH) is a syndrome characterized by intracranial pressure. The purpose of this study was to evaluate the accuracy of diagnosis by giving the quantitative values of imaging findings of IIH on cranial MRI.

Methods: This study included 37 patients who were diagnosed with IIH and 22 healthy controls were included by using a match-to-pair technique regarding their sex and age. Optic nerve sheath diameter (ONSD) and optic nerve diameter (OND), transverse diameter of Meckel cave (MCD), superior sagittal sinus area (SSSA), right and left transverse sinus area (RTSA, LTSA), transverse diameter of right lateral ventricle frontal horn (LVFHD), vertical length and A-P diameter of sella cavity, vertical length of pituitary gland (VLPG) were measured. ROC analysis was performed for the cut off values.

Results: There was statistically significant difference between patient group and control group in terms of ONSD, MCD, SSSA, RTSA, LTSA sella cavity A-P diameter, VLPG vertical length of sella and hypoplasia of dural venous sinus. On the other hand, there was no statistically significant difference between the patient and the control groups regarding LVFHD, OND. ONSD showed an AUC of 0.996 with a cutoff value of 6.4 mm >=, it was found to be highest reliable marker in the differential diagnosis of IIH patients from the controls.

Conclusion: We believe that this study can an important contribution to the diagnosis IIH and follow-up period after its treatment by measuring ONSD, MCD, Sella A-P and VLPG, vertical length of sella and SSSA.

Keywords: Idiopathic intracranial hypertension, magnetic resonance imaging, quantitative evaluation

INTRODUCTION

Idiopathic intracranial hypertension (IIH) is a syndrome characterized by intracranial pressure increase without any space occupying lesion. There are other definitions like meningitis serosa or pseudotumor cerebri as well. It usually presents by headache. It can be seen more in obese women who are in fertility period, but it can be observed in each age and gender.[1] The underlying pathophysiological mechanism is not known certainly. The incidence of the disease changes. However, its frequency has raised with respect to the increasing incidence of obesity.[2,3] Diagnosis of the disease is usually achieved through clinical findings, imaging findings and lumbar puncture. It is characterized by symptoms of IIH are severe headache in bilateral frontal and retro-orbital areas, temporary visual disorders, dizziness, diplopia, neck and back pain, nausea and vomiting. As to imaging findings,[4] distention of the optic nerve sheath, distention of the Meckel cave, slit like ventricles, flattened posterior optic globe and empty sella turcica, deformated pituitary gland are the main findings.[5]

The purpose of this study was to evaluate the accuracy of diagnosis by giving the quantitative values of imaging findings of IIH on cranial magnetic resonance imaging (MRI).

METHODS

Patients Subjects

This study included 37 patients who were diagnosed with IIH by using Modified Dandy Criteria (including headache, persistent or progressive visual dysfunction, papilledema, pulsatile tinnitus, greater than 250 mmH_2O CSF opening pressure without ventricular enlargement or intracranial mass on imaging, and normal CSF constituents) and examined for their headaches between January 2016 and May 2021. Images of the patients were scanned retrospectively through PACS system. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its lateramendments or comparable ethical standards. This retrospective study was approved by the institutional review board of the Tokat Gaziosmanpaşa University Faculty of Medicine Dean's Office Clinical Researches Ethics Committee (Date: 31.03.2022, Decision No: 22-KAEK-076).

Patients who were younger than 18, pregnant and who had intracranial mass, hydrocephalus or vascular pathology in brain MRI, who had treatment or medication which can

Corresponding Author: Fatma Kökcü, fatmakokcu79@hotmail.com

cause IIH or have systemic diseases, who had current or prior sinus vein thrombosis was not included in the study. The ones with pregnancy and postpartum status, and who had surgical operation which can change CSF circulation were also excluded from the participants group.

Control Subjects

Twenty-two healthy controls were included by using a match-to-pair technique regarding their sex and age. The ones who have any kind of nervous system disease and have primary headache syndromes like migraine and cluster headache were not included in the study.

MRI Protocol

All patients were scanned with 1.5 Tesla MRI device (1.5-Tesla magnet, GE Signa Excite HD; GE Medical Systems, Milwaukee, WI, USA). The patients were examined using 16 channel neurovascular head coil in the routine axial plane, using the sagittal T2W, transvers propeller T2W (TR: 9000 ms, TE: 90 ms, NEX: 1.0, slice thickness: 5.5 mm, slice spacing: 1.5 mm) coronal FLAIR T2W (TR: 4803 ms, TE: 107ms, NEX: 2.0, slice thickness: 7 mm, slice spacing: 1.5 mm), axial 3D BRAVO (TR: 9.25 ms, TE: 3.58 ms, NEX: 1.0, slice thickness: 1 mm, slice spacing: 0.5 mm) DWI (TR6992 ms, TE: 84.4 ms, NEX: 2.0, slice thickness: 5 mm, slice spacing: 5.5).

Image Analysis

All measurements were done by one neuroradiologist, blinded to patient and control groups and to the medical history, with 15 years of neuroradiology experience. Optic nerve sheath diameter (ONSD) (from widest portion) and optic nerve (ON) thickness (4 mm posterior to the globe) were measured on axial T2W series (Figure 1). Transverse diameter of Meckel cave (MCD) was measured on axial T2W series as well from middle part of cave (Figure 2). Furthermore, superior sagittal sinus area (SSSA) was measured from 1 cm superior of torcular herophili on the same series. Area measurements were done with manuel ROI drawing external contour of sinuses (Figure 3). Right transverse sinus area (RTSA) and left transverse sinus area (LTSA) were also measured in a similar way from sagittal T2W images from the middle part of sinuses.

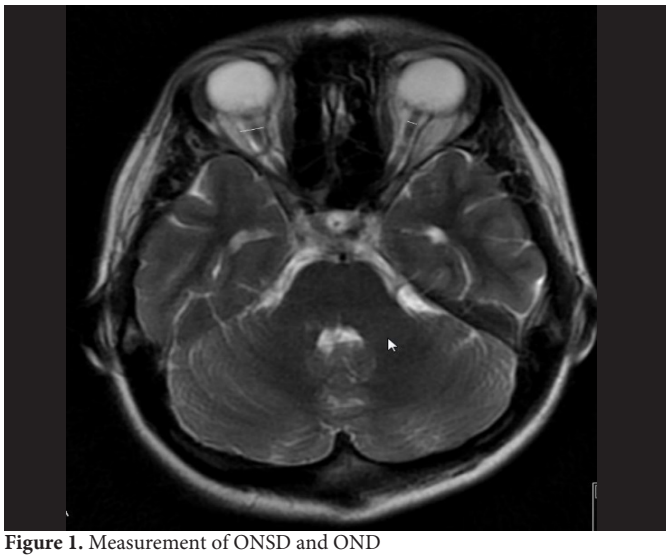

Figure 1. Measurement of ONSD and OND

ONSD: Optic nerve sheath diameter, OND: Optic nerve diameter

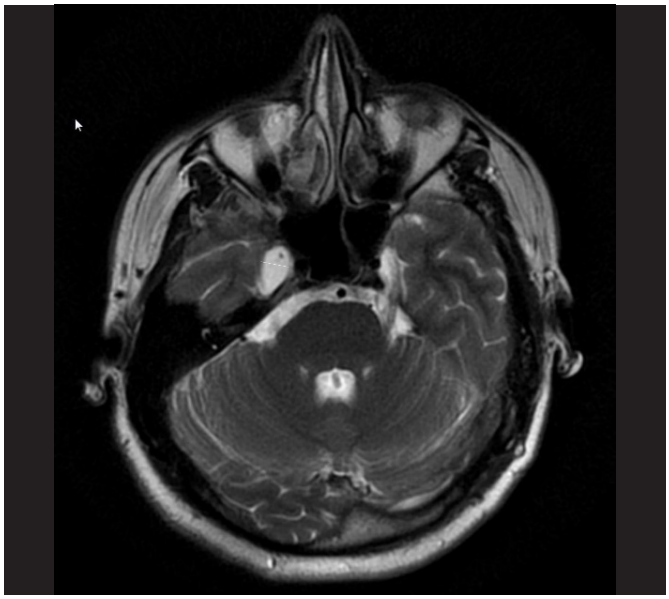

Figure 2. Measurement of MCD

MCD: Meckel cave diameter

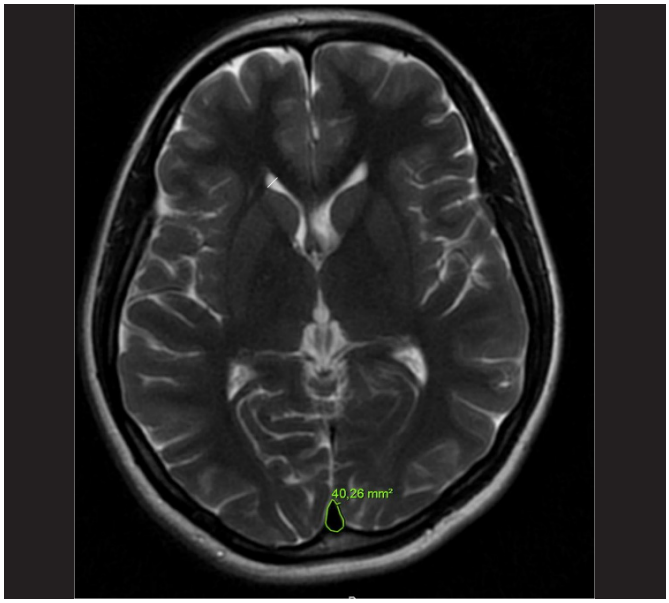

Figure 3. Measurement of LVFHD, area of SSSA

LVFHD: Lateral ventricle frontal horne diameter, SSSA: Superior sagittal sinus area

Transverse sinuses were examined in terms of hypoplasia from MPR images of axial 3D bravo series. Regarding ventricular compression, transverse diameter of right lateral ventricle frontal horn were measured on basal ganglions plane of axial T2W series (Figure 3). The vertical length of sella cavity (VLSC) (from the middle point of baseline to middle point of diaphragma sella) on and A-P diameter of sella cavity (from the tip of anterior clinoid to tip of posterior clinoid) were measured on midsagittal plane of T2W sagittal series (Figure 4). Vertical length of pituitary gland (VLPG) was measured on midsagittal plane.

Statistical Analysis

Statistical analysis was conducted by using SPSS 20 program. As the data did not show normal distribution, a nonparametric Mann-Whitney U test was run to compare the groups. p-value

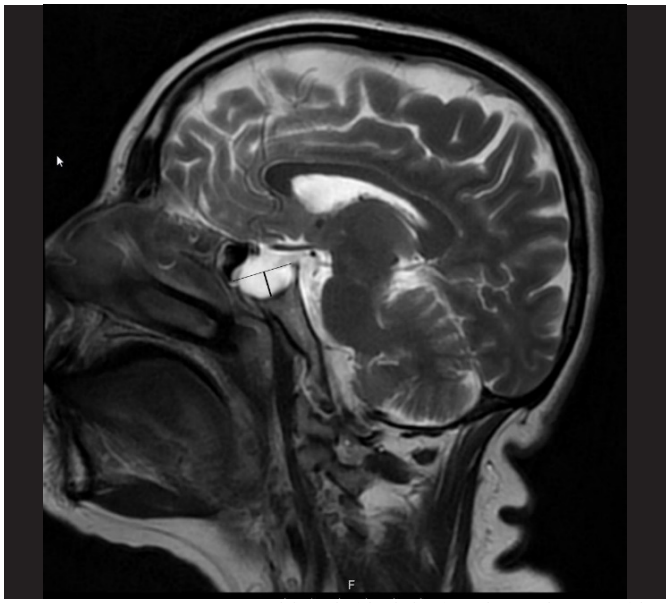

Figure 4. Measurement vertical length (thick line) and A-P diameter (thin line) of sella cavity

was lower than 0.05 and this was interpreted as significant. ROC analysis was performed for the cut off values. ROC curves were used for the comparison of sensitivity and specificity.

RESULTS

In this study total 59 patients consisting of 37 patient group and 22 control group were evaluated. Average age was 39 (18-71) in the patient group and 41 (18-69) in the control group and there was no statistically significant difference. The patient group included 24 female and 13 male patients. As to the control group, there were 13 female and 9 male subjects. Averages and statistical values of the conducted measurements were summarized in Table 1. Here 33 patients reported headaches (89.1%), 24 patients reported visual impairment (64.8%), 18 patients reported dizziness (48.6%), 16 patients (43.2%) reported back and neck pain, 7 patients (18.9%) reported pulsatile tinnitus. Average opening pressure CSF on LP was 380 mmH_2O (280-430). In all patients the CSF constituents was found to be normal.

Table 1. Comparison of measurements in IIH and control group

	IIH (37)		Control (22)		
	Mean	SD	Mean	SD	p
ONSD (mm)	8.0	1.0	5.0	0.6	<0.001
OND (mm)	3.0	0.5	2.9	0.4	0.225
MCD (mm)	7.0	1.7	4.9	1.2	<0.001
LVFHD (mm)	3.5	1.6	4.6	2.5	0.096
Sella A-P diameter (mm)	12.3	2.6	10.7	1.3	0.011
VLSC (mm)	9.2	2.1	7.2	0.9	<0.001
VLPG (mm)	2.7	1.1	5.1	1.9	<0.001
SSSA (mm²)	23.3	11.3	32.1	9.7	0.001
RTSA (mm²)	18.9	14.2	29.1	13.2	0.003
LTSA (mm²)	14.3	15.9	16.0	12.0	0.323

IIH: Idiopathic intracranial hypertension, ONSD: Optic nerve sheath diameter, OND: Optic nerve diameter, MCD: Meckel cave diameter, LVFHD: Lateral ventricle frontal horne diameter, VLSC: Vertical length of sella cavity, VLPG: Vertical length of pituitary gland, SSSA: Superior sagittal sinus area, RTSA: Right transves sinus area, LTSA: Left transves sinus area

It was found that there was statistically significant (p-value <0.005) difference between patient group and control group in terms of ONSD, MCD, SSSA, RTSA, LTSA, A-P diameter of sella cavity, VLSC, VLPG, and hypoplasia of dural venous sinus. On the other hand, there was no statistically significant difference between the patient and the control groups regarding transverse diameter of lateral ventricular frontal horn, transverse diameter of optic nerve. The value of ONSD, MCD, SSSA, RTSA, A-P diameter of sella cavity, VLS and VLPG in the differential diagnosis of IIH positive patients from control subjects was evaluated by the ROC curve. Since ONSD showed an AUC of 0.996 with a cutoff value of 6.4 mm >= (p<0.001), it was found to be highest reliable marker in the differential diagnosis of IIH patients from the controls. Also MCD showed an AUC of 0.850 with a cutoff value of 5.5 mm >= (p<0.001). VLS showed an AUC of 0.826 with a cutoff value of 8,8 mm >= (p<0.001)., A-P diameter of sella showed an AUC of 0.697 with a cutoff value of 11.8 mm >= (p<0.001). VLPG with a cutoff value of <=4mm (p<0.001 AUC:0.888), SSSA with a cutoff value of <=26 mm (p<0.001, AUC:0.750) and RTSA with a cutoff value of <=21mm (p<0.001, AUC:0.729) were reliable markers in the differential diagnosis of IIH patients from the controls too. This relationship was displayed in Figure 5, and Table 2.

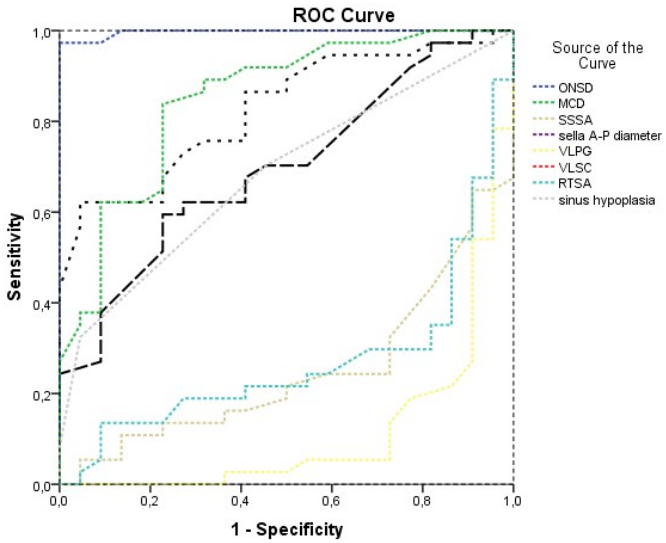

Figure 5. ROC curve of parameters in differential diagnosis of IIH patients for control

ROC: Receiver operating characteristic, IIH: Idiopathic intracranial hypertension

Table 2. ROC of p-value significant parameters in differential diagnosis of IIH patients for control

	Cutoff value	AUC/SD	Specificity	Sensitivity
ONSD (mm)	>=6.4	0.996/0.003	1	0.973
MCD (mm)	>=5.5	0.849/0.052	0.772	0.837
VLSC (mm)	>=8.8	0.826/0.052	0.9545	0.621
VLPG (mm)	<=4	0.887/0.047	0.727	0.945
SSSA (mm²)	<=26	0.750/0.063	0.727	0.756
RTSA (mm²)	<=21	0.729/0.068	0.818	0.702
Sella A-P diameter (mm)	>=11.8	0.697/0.068	0.594	0.772

90-100: Excellent, 80-90: Good, 70-80: Fair, 60-70: Poor, 50-60: Fail, ROC: Receiver operating characteristic, AUC: Area under curve, SD: Standard deviaition, ONSD: Optic nerve sheath diameter, MCD: Meckel cave diameter, VLSC: Vertical length of sella cavity, VLPG: Vertical length of pituitary gland, SSSA: Superior sagittal sinus area, RTSA: Right transves sinus area

When examined in relation to sinus hypoplasia, there was 59% (n: 26) sinus hypoplasia in the patient group and 37% (n: 14) of them were in LTS, 24% (n: 9) were in RTS and 8% (n: 3) were bilateral. 45% (n: 10) rate of sinus hypoplasia was detected in the control group. 40% (n: 9) of them were in LTS 4% (n: 1) were in RTS. There were no patients with bilateral hypoplasia in the control group. Sinus hypoplasia was significantly more in the patient group than it was in the control group (p: 0.012). Rate of LTS hypoplasia was similar in both groups, however RTS hypoplasia was more in the patient group.

DISCUSSION

One of the most common findings related to IIH is the distention of the ONS and tortuosity.[5-7] There has been multiple research investigating this subject and it has been claimed by some studies that distention of the ONS and tortuosity has not changed statistically significantly.[8] In some studies, patients with IIH and healthy people were compared through volume measurement. According to the last study conducted based on volume measurement,[1] when ONS volume is higher than 201 mm^3, it is interpreted as significant for IIH diagnosis. Furthermore, it was emphasized in some studies, which were based on measurement of the ONSD, that increase in the diameter of the ONS is significant for the diagnosis of the IIH and sensitivity and specificity were lower. However, to conduct a volume measurement there should be either an automatic volume measurement program or it should be carried out by conducting semi-automatic volume measurement by drawing ROI through external contour of the ONS on each section. Semi-automatic volume calculation programs are used in the centers with no automatic volume calculation program and this is time-consuming. Therefore, diameters of the ONS and ON were calculated in this study as it was fast and practical. It was claimed in some studies that ON volume does not change in patients with IIH while some assert that it increases in these patients.[1] In this study, there was no difference between two groups in terms of optic nerve diameter. As to the ONSD it was statistically significantly higher in the patient group and it was due to the CSF flow through intraorbital subarachnoid distance as a result of intracranial pressure increase in the patients with IIH. Furthermore, flattening in glob posterior was observed due to the same mechanism.[9] Also, ONSD was the highest reliable marker in the differential diagnosis of IIH patients from the controls with a high AUC. In a patient with suspected intracranial hypertension based on clinical and examination findings when ONSD is higher than 6.4 mm, it can be reported IIH with high sensitivity and specificity.

Compression and decrease in diameter of dural venous sinuses resulting from pressure which is secondary to intracranial pressure increase.[10] Superior sagittal sinus area was measured from 1 cm superior of torcular herophili. When it is compared to the control group, average of superior sagittal sinus area was 23.3 mm^2 while it was 32.1 mm^2 in the control group and there was statistically significant difference. In a similar study Rohr et al.[10] also ran cross-sectional area measurement from superior sagittal sinus. When they compared the situation before and after IIH treatment, they found that sinus area increased after the treatment. In the present study, when SSSA is less than 26 mm^2, it can be reported highly probable IIH according to ROC analysis.

Transverse sinus stenosis is another IIH finding and it has been seen in 30-93% of patients.[11] There is no certain information regarding whether stenosis is a cause or result of IIH. Statistically significant rate of transverse sinus stenosis was found in the study by Horev et al.[1] They measured diameters of transverse sinus before and after lumbar puncture (L-P) and compared the results. It was seen that diameters of transverse sinus significantly increased after L-P. As to the ONS diameter and pituitary gland height obtained after L-P, there was no difference. However our study was based on routine MRI and there was no examination of MR venography of the patients. On the other hand, transverse sinus hypoplasia was significantly more in the patient group (p: 0.012) when an evaluation was conducted in terms of hypoplasia. The kind of hypoplasia which was seen in the patient group was RTS hypoplasia. To get a more optimal examination we conducted measurement of the area instead of diameter, and when we compared the results to the control group's area of RTS was found to be narrower in the patient group. There was no significant difference regarding LTS area. The cause of average area difference in transverse sinuses was actually the fact that RTS hypoplasia was more in the patient group decreased average of area. As LTS hypoplasia was not different in the patient and control groups, average area measurements were not also statistically different.

Distention of the Meckel's cave is another IIH finding.[12,13] The distention is caused by chronic intracranial pressure increase and pouch, which is like meningocele, formed in spaces filled with CSF. On the other hand, there is also literature claiming the narrowing in Meckel's cave in IIH.[14] There is also publications showing that there is no change in MCD.[15] To get a quantitative evaluation, we measured transverse diameters of Meckel's cave at axial section. In this study, it was seen that MCD was statistically significantly wider in the patient group than the control group. In ROC analysis, MCD showed high AUC. If MCD is higher than 5.5 mm in a patient with suspected intracranial hypertension it can be reported highly probable IIH with high sensitivity and specificity

Empty or partial empty sella appearance is a radiological finding supporting IIH. Chronic intracranial pressure increase and increase in CSF pressure cause compression and height decrease in pituitary gland, and remodeling and enlargement in sella bone structure.[16-19] As a result, empty/partial empty sella appearance is formed. When the relation between vertical lengths of pituitary gland and sella were examined in some morphological studies, it was found that statistically significant decrease was seen in the patients with IIH.[18-20] However, Beier et al.[21] claimed in their study that empty sella was not related to IIH. Furthermore, Hoffmann et al.[1] compared normal patients to the patients with IIH by conducting volume measurement of pituitary gland. In this study, we measured vertical length of pituitary gland. Moreover, we measured vertical length of sella cavity and A-P diameter of sella to distinguish it from empty/partial empty sella appearance seen in elderly ages without any relation to IIH. In the patients with IIH, vertical length of pituitary gland

was seen to have decreased, and vertical length of sella cavity and A-P diameter of sella was found to have increased. Both VLSC and VLPG showed good AUC in differentiation of IIH from to control subjects. A-P diameter of sella showed poor AUC.

It has been stated by some studies in the literature that slit like ventricles by ventricular compression is a finding of IIH.[22] On the other hand, some studies claim that ventricular configuration in IIH patients is normal.[23,24] Zheng et al.[25] compared changes of ventricular volume in normal subjects and CVT related IIH. They did not find significant difference in ventricular system volumes between these groups. In this study diameter measurement was conducted at ventricular system to obtain a quantitative evaluation. As in hydrocephalus measurement of diameter was run from lateral ventricle frontal horn. When the patient and control groups were compared, there was no statistically significant difference in terms of diameter of ventricles.

Limitations

Main limitation of this study was that the number of patients was not high and there was no access to detailed clinical information patients such as symptoms duration period. To our knowledge, while there are many studies in the literature on imaging findings of idiopathic intracranial hypertension, the number of studies providing cut-off values is quite limited. However, our study is preliminary, and we believe that cut-off values can be correlated with larger patient groups in larger centers. Furthermore, there was no optimal evaluation in terms of venous stenosis due to the absence of MRI venography.

CONCLUSION

As a result, final diagnosis of IIH is achieved through measurement of CSF pressure after LP although clinical findings and imaging may help the process. We believe that this study can make an important contribution the diagnosis IIH and follow-up period after its treatment by measuring ONSD, MCD, Sella A-P and vertical diameter, vertical length of pituitary gland and SSSA. Sometimes it can be confusing whether it is within normal limits or pathological when reporting IIH cranial MR imaging findings such as increase of ONSD, MCD or decrease of VLPG. According to present study it can be reported highly probable IIH when the measurements are higher than 6.4 mm for optic nerve sheath diameter, 5.5 mm for Meckel cave diameter, 8.8 mm for vertical length of sella and less than 4 mm for vertical length of pituitary gland, 26 mm^2 for superior sagittal sinus area and 21 mm^2 for right transverse sinus area in a patient with suspected intracranial hypertension based on clinical and examination findings.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was carried out with the permission of the Tokat Gaziosmanpaşa University Faculty of Medicine Dean's Office Clinical Researches Ethics Committee (Date: 31.03.2022, Decision No: 22-KAEK-076).

Informed Consent

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

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

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

Author Contributions

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

Acknowledgement

Associated Prof. Osman Demir did the statistical analysis.

REFERENCES

1. Hoffmann J, Schmidt C, Kunte H, et al. Volumetric assessment of optic nerve sheath and hypophysis in idiopathic intracranial hypertension. *AJNR Am J Neuroradiol*. 2014;35(3):513-518. doi:10.3174/ajnr.A3694
2. Bénard-Séguin É, Costello F. Idiopathic intracranial hypertension with papilledema. *Neurosurg Clin N Am*. 2025;36(2):281-298. doi:10.1016/j.nec.2024.12.001
3. Zhou C, Zhou Y, Liu L, et al. Progress and recognition of idiopathic intracranial hypertension: a narrative review. *CNS Neurosci Ther*. 2024;30(8):e14895. doi:10.1111/cns.14895
4. Mathew NT, Ravishankar K, Sanin LC. Coexistence of migraine and idiopathic intracranial hypertension without papilledema. *Neurology*. 1996;46(5):1226-1230. doi:10.1212/wnl.46.5.1226
5. Agid R, Farb RI, Willinsky RA, Mikulis DJ, Tomlinson G. Idiopathic intracranial hypertension: the validity of cross-sectional neuroimaging signs. *Neuroradiology*. 2006;48(8):521-527. doi:10.1007/s00234-006-0095-y
6. Brodsky MC, Vaphiades M. Magnetic resonance imaging in pseudotumor cerebri. *Ophthalmology*. 1998;105(9):1686-1693. doi:10.1016/S0161-6420(98)99039-X
7. Manfré L, Lagalla R, Mangiameli A, et al. Idiopathic intracranial hypertension: orbital MRI. *Neuroradiology*. 1995;37(6):459-461. doi:10.1007/BF00600093
8. Gibby WA, Cohen MS, Goldberg HI, Sergott RC. Pseudotumor cerebri: CT findings and correlation with vision loss. *AJR Am J Roentgenol*. 1993;160(1):143-146. doi:10.2214/ajr.160.1.8416612
9. Acheson JF. Idiopathic intracranial hypertension and visual function. *Br Med Bull*. 2006;79-80:233-244. doi:10.1093/bmb/ldl019
10. Rohr A, Bindeballe J, Riedel C, et al. The entire dural sinus tree is compressed in patients with idiopathic intracranial hypertension: a longitudinal, volumetric magnetic resonance imaging study. *Neuroradiology*. 2012;54(1):25-33. doi:10.1007/s00234-011-0850-6
11. Ball AK, Clarke CE. Idiopathic intracranial hypertension. *Lancet Neurol*. 2006;5(5):433-442. doi:10.1016/S1474-4422(06)70442-2
12. Bialer OY, Rueda MP, Bruce BB, Newman NJ, Biousse V, Saindane AM. Meningoceles in idiopathic intracranial hypertension. *AJR Am J Roentgenol*. 2014;202(3):608-613. doi:10.2214/AJR.13.10874
13. San Millán D, Kohler R. Enlarged CSF spaces in pseudotumor cerebri. *AJR Am J Roentgenol*. 2014;203(4):W457-W458. doi:10.2214/AJR.14.12787
14. Degnan AJ, Levy LM. Narrowing of Meckel's cave and cavernous sinus and enlargement of the optic nerve sheath in Pseudotumor Cerebri. *J Comput Assist Tomogr*. 2011;35(2):308-312. doi:10.1097/RCT.0b013e31820d7a70

15. Balık AÖ, Akıncı O, Yıldız S, Hasırcı Bayır BR, Ulutaş C. Role of neuroimaging markers on predicting of idiopathic intracranial hypertension. *Acta Radiol*. 2024;65(8):999-1006. doi:10.1177/02841851241256008
16. Kyung SE, Botelho JV, Horton JC. Enlargement of the sella turcica in pseudotumor cerebri. *J Neurosurg*. 2014;120(2):538-542. doi:10.3171/2013.10.JNS131265
17. Ranganathan S, Lee SH, Checkver A, et al. Magnetic resonance imaging finding of empty sella in obesity related idiopathic intracranial hypertension is associated with enlarged sella turcica. *Neuroradiology*. 2013;55(8):955-961. doi:10.1007/s00234-013-1207-0
18. Yuh WT, Zhu M, Taoka T, et al. MR imaging of pituitary morphology in idiopathic intracranial hypertension. *J Magn Reson Imaging*. 2000;12(6):808-813. doi:10.1002/1522-2586(200012)12:6<808::aid-jmri3>3.0.co;2-n
19. Kesler A, Yaffe D, Shapira M, Kott E. Optic nerve sheath enlargement and reversal of optic nerve head in pseudotumor cerebri. *Harefuah*. 1996;130(7):457-503.
20. Korkmazer B, Karaman AK, Kızılkılıç EK, et al. Efficacy of dural sinus quantitative measurements in idiopathic intracranial hypertension: a practical diagnostic feature. *Clin Neuroradiol*. 2023;33(2):545-554. doi:10.1007/s00062-022-01244-0
21. Beier D, Korsbæk JJ, Bsteh G, et al. Magnetic resonance imaging signs of idiopathic intracranial hypertension. *JAMA Netw Open*. 2024;7(7):e2420138. doi:10.1001/jamanetworkopen.2024.20138
22. Weisberg LA. Computed tomography in benign intracranial hypertension. *Neurology*. 1985;35(7):1075-1078. doi:10.1212/wnl.35.7.1075
23. Agid R, Farb RI, Willinsky RA, Mikulis DJ, Tomlinson G. Idiopathic intracranial hypertension: the validity of cross-sectional neuroimaging signs. *Neuroradiology*. 2006;48(8):521-527. doi:10.1007/s00234-006-0095-y
24. Arkoudis NA, Davoutis E, Siderakis M, et al. Idiopathic intracranial hypertension: imaging and clinical fundamentals. *World J Radiol*. 2024;16(12):722-748. doi:10.4329/wjr.v16.i12.722
25. Zheng YM, Chen J, Yuan MG, Wu ZJ, Dong C. Does a change in ventricular size predict a diagnosis of cerebral venous thrombosis-related acute intracranial hypertension? Results of a retrospective imaging study. *Acta Radiol*. 2019;60(10):1308-1313. doi:10.1177/0284185118823346

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1754179

J Health Sci Med. 2025;8(5):865-870

Effects of cigarette smoking on handgrip strength and muscular endurance in young adults

Fırat Koç

Department of Head of Physical Anthropology, Faculty of Humanities and Social Sciences, Hitit University, Çorum, Turkiye

Cite this article as: Koç F. Effects of cigarette smoking on handgrip strength and muscular endurance in young adults. *J Health Sci Med.* 2025;8(5):865-870.

Received: 30.07.2025 ◆ **Accepted**: 20.08.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: The aim of this study is to evaluate the effects of cigarette use on handgrip strength (HGS) in young adult individuals in the context of gender and hand dominance, and to explore the physiological and behavioral mechanisms of this relationship.

Methods: In this observational study involving a total of 140 university students, participants were divided into three groups based on their smoking status: never smokers, former smokers, and current smokers. Three repeated HGS measurements were performed on both the dominant and non-dominant hands of each individual using a Jamar brand hand dynamometer. Data were analyzed using one-way ANOVA and Bonferroni-corrected post-hoc tests.

Results: Although smokers exhibited higher HGS values in the initial measurements, a significant performance decline was observed by the third measurement. This decline was more pronounced in the non-dominant hand (up to 4.3 kg), indicating impairments in muscular endurance and fatigue tolerance. The effect was observed across all groups regardless of gender and hand dominance. The findings suggest that smoking adversely affects not only maximal muscle strength but also neuromuscular coordination and recovery capacity.

Conclusion: Cigarette use has significantly negative effects on muscle performance and fatigue tolerance. Handgrip strength is an effective indicator not only of peripheral muscle strength but also of systemic health status and lifestyle factors. It is recommended that HGS measurements be used more widely in public health screenings and smoking cessation programs.

Keywords: Handgrip strength, cigarette use, muscle fatigue, neuromuscular function, physical performance

INTRODUCTION

Handgrip strength (HGS) is a widely used indicator of upper extremity muscle function and an emerging biomarker of general health, with established links to morbidity, mortality, and functional decline.[1,2] HGS is not only indicative of upper extremity strength but also reflects general muscle strength and physiological reserve, making it closely associated with several health indicators such as sarcopenia, reduced physical performance, frailty syndrome, and mortality.[3-5]

The literature has shown that HGS is inversely associated with a wide range of chronic diseases, including cardiovascular diseases,[6] metabolic syndrome,[7] cancer,[8] chronic obstructive pulmonary disease (COPD),[9] and neurological disorders.[10] Particularly, low handgrip strength has been linked to reduced lifespan and increased risk of dependency in older age.[11] In the UK Biobank study, which tracked 500,000 individuals, low HGS was significantly associated with cardiovascular mortality, respiratory diseases, and cancer-related deaths.[12] Similarly, in a study conducted by Gubelmann and colleagues in the Swiss population, lower HGS values were significantly inversely associated with cardiovascular risk markers such as body fat percentage, systolic blood pressure, glucose levels, and hs-CRP.[13]

One of the most significant environmental factors contributing to reduced muscle strength is cigarette smoking. Smoking induces inflammatory processes that increase oxidative stress, disrupt mitochondrial functions, and lead to the degradation of muscle proteins, thereby exerting harmful effects on muscle tissue.[14,15] Additionally, physiological changes such as chronic hypoxemia and reduced peripheral oxygen distribution can contribute to decreased muscle strength.[16] These effects are particularly observable in localized muscle performance measures such as handgrip strength. Kopiczko and colleagues reported significantly lower HGS levels in young female smokers compared to their non-smoking peers.[17] Similar findings were reported by Jiang and colleagues in a study focused on hand and finger muscle strength.[8]

Moreover, cigarette smoking is not only linked to biological mechanisms but is also associated with behavioral and social factors. Smokers are generally known to engage in lower levels of physical activity, exhibit unhealthy dietary habits, and have lower overall health awareness.[19] These lifestyle factors can also directly or indirectly affect HGS levels. Therefore, when evaluating the impact of smoking on HGS, it is essential to

Corresponding Author: Fırat Koç, kocfiratkoc@gmail.com

control for variables such as sex, age, body-mass index (BMI), physical activity level, nutrition, and history of chronic diseases.[20]

Handgrip strength, as an indicator of both muscle function and general health, holds an important position as a biomarker in the field of public health. However, the effects of cigarette use on this parameter-especially in young adults, middle-aged individuals, and at-risk populations-have not been fully elucidated. The aim of this study is to evaluate the effects of smoking on handgrip strength, discuss the potential physiological and behavioral mechanisms underlying this relationship, and provide a comprehensive contribution to the literature on the subject.

Therefore, the aim of this study is to investigate the acute effects of cigarette smoking on HGS in young adults by examining performance differences across smoking status, sex, and hand dominance. By controlling for potential confounding variables, this study seeks to clarify whether smoking impairs muscular endurance and recovery capacity during repeated effort, as measured by HGS.

The study was designed to test the hypothesis that cigarette smoking negatively affects muscular endurance and recovery capacity, as measured through repeated handgrip trials.

METHODS

Materials

This study received ethical approval from the Hitit University Non-interventional Researches Ethics Committee (Date: 06.01.2025, Decision No: 2024-27). All procedures were carried out in accordance with the ethical rules and the principles of the Declaration of Helsinki. A total of 140 voluntary undergraduate students, aged between 19 and 25 years, participated in the study. Of the participants, 52.1% were male (n=73) and 47.9% were female (n=67). The mean age was 21.22±1.6 years for males and 21.10±1.7 years for females. Written informed consent was obtained from all individuals prior to data collection, in accordance with the ethical principles of human research.

To assess the statistical adequacy of the sample, a post-hoc power analysis was conducted using G*power 3.1 software. Based on the observed effect size in repeated-measures ANOVA (partial eta squared: 0.08), with an alpha level of 0.05 and three comparison groups, the calculated statistical power was 0.92. This result indicates that the sample size was sufficiently robust to detect significant differences with high reliability. Although the sample was formed through voluntary participation, the age and sex distribution was largely consistent with that of the overall student population within the faculty, thereby supporting the representativeness of the sample.

To ensure sample homogeneity, individuals with any diagnosed chronic illness (e.g., diabetes, hypertension, cardiovascular disorders), physical disabilities, or mobility limitations were excluded from the study. Furthermore, all participants were actively enrolled undergraduate students studying at the same academic level within the same faculty. This criterion contributed to the sample's homogeneity not only in terms of educational background but also in terms of socioeconomic status, as the participants were likely to share similar socioeconomic conditions.

Methods

This study was conducted using a descriptive and comparative research design. Participants were categorized into three groups based on their smoking status; group I consisted of individuals who had never smoked (n=44); group II included former smokers who had quit smoking (n=48); and group III comprised individuals who were current active smokers at the time of the study (n=48).

Hand dominance was determined by participants' responses to the question; "Which hand do you use when writing, eating, or throwing a ball?" Responses were classified as "right" or "left." Participants who reported using both hands were evaluated based on the hand they preferred for most tasks. This simple and widely used method is considered a reliable and practical approach in studies involving functional measurements.

Smoking status was assessed via self-report. Individuals who stated that they smoked at least one pack of cigarettes per day (approximately 20 cigarettes) were classified as "active smokers." Those who smoked less than this amount were excluded from the study to ensure clearer statistical differentiation based on usage frequency and dose. This threshold was selected based on established classifications in epidemiological research, where smoking ≥20 cigarettes per day is commonly categorized as heavy or active smoking. This cutoff has been used to represent high-dose nicotine exposure and its physiological impacts in studies by the CDC and the Surgeon General.[21,22] Moreover, Hecht[23] describes this level of consumption as a benchmark for increased biomarker expression and carcinogenic exposure in smokers.[23] This approach allowed for a more distinct evaluation of the impact of cigarette use on handgrip strength.

Handgrip strength measurements were conducted using a Jamar brand hydraulic hand dynamometer (USA), which has international validity. Three repeated measurements were taken for both the right and left hands of each participant. During the measurements, participants were seated with their elbow flexed at 90°, and the arm close to the torso, and were instructed to squeeze the device with maximum force. A one-minute rest interval was provided between each trial to allow for the assessment of muscle fatigue effects.

Statistical Analysis

All statistical analyses were conducted using IBM® SPSS® Statistics for Windows, Version 26.0 (IBM Corp., Armonk, NY, USA). Descriptive statistics (mean±standard deviation) were calculated for each group. Between-group comparisons were performed using one-way ANOVA, followed by Bonferroni-adjusted post-hoc tests where applicable. To compare repeated measurements, repeated-measures ANOVA was employed. The internal consistency of the measurements was evaluated

using Cronbach's alpha coefficient, which indicated good reliability (α=0.81). A p-value of less than 0.05 was considered statistically significant in all analyses.

RESULTS

BMI data were collected using standardized anthropometric measurements, including height and weight. To further characterize the physical profile of the sample, a BMI classification table is presented in Table 1.

Table 1. BMI classification of participants

BMI categoryx	BMI range (kg/m²)	n	(%)
Underweight	<18.5	12	8.6%
Normal weight	18.5-24.9	87	62.1%
Overweight	25.0-29.9	31	22.1%
Obese (class I)	30.0-34.9	8	5.7%
Obese (class II-III)	≥35.0	2	1.5%
Total		140	100%

BMI: Body-mass index

Analyses conducted to determine the effects of variables such as smoking status, sex, and hand dominance on handgrip strength revealed significant differences between the groups. The findings are reported in detail below.

Among right-hand dominant male participants, active smokers exhibited the highest HGS during the initial trial (45.2±7.0 kg), with a statistically significant difference compared to non-smokers (p=0.04, r=0.274). However, by the third trial, this advantage diminished notably, showing a marked decline in performance (41.6±6.5 kg, p<0.001, r=0.415). The observed 3.6 kg drop in dominant hand strength highlights the detrimental impact of smoking on muscular endurance and recovery capacity. A similar trend was observed in the non-dominant (left) hand. Active smokers again showed higher initial grip values (41.9±6.4 kg) in the non-dominant hand, but experienced a significant 3.5 kg decrease (from 41.9 kg to 38.4 kg) by the third measurement (p<0.001, r=0.380). This decline, even in the non-dominant hand, indicates impaired fatigue tolerance. Overall, the data suggest that smoking negatively affects muscle performance, regardless of hand dominance (Table 2).

Left-hand dominant male participants also demonstrated a consistent pattern. Active smokers recorded the highest grip strength in the first trial (40.4±6.4 kg), yet experienced a 4.0 kg decrease in the third trial (36.4±6.3 kg), a statistically significant decline (p=0.04, r=0.362). This supports the notion that smoking compromises muscular endurance over short-term repetitive effort. In the non-dominant (right) hand, the decline was also evident-from 42.1±6.7 kg to 38.4±6.4 kg-resulting in a 3.7 kg difference. Although the p-value approached significance (p=0.06), the moderate correlation coefficient (r=0.341) still suggests a meaningful association between smoking and reduced grip endurance (Table 3).

Among right-hand dominant female participants, active smokers initially demonstrated higher grip strength compared to never-smokers and former smokers in both hands. However, this advantage significantly diminished over repeated trials, with a marked performance drop observed during the third measurement. A fatigue difference of 3.3 kg in the right (dominant) hand and 4.3 kg in the left (non-dominant) hand was detected, both statistically significant (p<0.01).

The more pronounced decline in the non-dominant hand suggests that smoking impairs muscular endurance and recovery capacity more clearly in less-utilized extremities. Effect size analysis showed a moderate clinical impact (r=0.285 for the right hand, r=0.314 for the left hand) (Table 4).

These findings indicate that, similar to male participants, smoking adversely affects neuromuscular endurance in females as well.

Among left-hand dominant female participants, active smokers demonstrated the highest initial grip strength in the dominant hand (36.0±5.9 kg). However, by the third trial, their performance declined to 32.2±5.7 kg-a fatigue drop of 3.8 kg, statistically significant (p=0.05, r=0.272). This moderate effect

Table 2. Right-hand dominant male participants-handgrip strength (kg) by measurement order and hand dominance

Hand/trial	Never smoked (mean±SD)	Former smokers (mean±SD)	Current smokers (mean±SD)	Fatigue difference	p-value	r-value
Right (dominant)-1st	43.1±6.2	42.3±6.7	45.2±7.0		0.04**	0.274
Right (dominant)-3rd	42.1±6.0	40.6±6.4	41.6±6.5	3.6 kg	<0.001**	0.415
Left (non-dominant)-1st	40.1±5.8	39.6±6.1	41.9±6.4		0.05	0.241
Left (non-dominant)-3rd	39.3±5.5	38.2±5.8	38.4±6.1	3.5 kg	<0.001**	0.380

p<0.05 statistically significant (*), p<0.001 highly statistically significant (**), SD: Standard deviation

Table 3. Left-hand dominant male participants-handgrip strength (kg) by measurement order and hand dominance

Hand/trial	Never smoked (mean±SD)	Former smokers (mean±SD)	Current smokers (mean±SD)	Fatigue difference	p-value	r-value
Left (dominant)-1st	38.1±5.9	37.3±6.1	40.4±6.4		0.06	0.232
Left (dominant)-3rd	36.9±5.6	36.0±6.0	36.4±6.3	4.0 kg	0.04*	0.362
Right (non-dominant)-1st	40.0±6.2	39.5±6.5	42.1±6.7		0.08	0.210
Right (non-dominant)-3rd	38.9±6.1	38.2±6.3	38.4±6.4	3.7 kg	0.06	0.341

p<0.05 statistically significant (*), p<0.001 highly statistically significant (**), SD: Standard deviation

Table 4. Right-hand dominant female participants-handgrip strength (kg) by measurement order and hand dominance						
Hand/trial	**Never smoked (mean±SD)**	**Former smokers (mean±SD)**	**Current smokers (mean±SD)**	**Fatigue difference**	**p-value**	**r-value**
Right (dominant)-1st trial	32.5±5.5	33.2±5.8	34.8±6.0		0.07	0.122
Right (dominant)-3rd trial	31.8±5.3	32.0±5.6	31.5±5.7	3.3 kg	<0.01*	0.285
Left (non-dominant)-1st trial	33.4±5.2	32.8±5.7	35.4±6.0		0.06	0.144
Left (non-dominant)-3rd trial	32.5±5.0	31.4±5.4	31.1±5.8	4.3 kg	<0.01**	0.314
p<0.05 statistically significant (*), p<0.001 highly statistically significant (**), SD: Standard deviation						

indicates that repeated exertion disproportionately impacts muscular endurance in smokers, even when starting strength is relatively high.

In the non-dominant (right) hand, a similar trend was observed with a 3.0 kg reduction from first to third trial, though the change did not reach statistical significance (p=0.07, r=0.221). While the p value approaches significance, the r value suggests a small to moderate effect that may still be clinically relevant (**Table 5**).

Positive and statistically significant correlations were found between BMI and HGS in both dominant and non-dominant hands. The strongest correlation was observed between HGS scores of the dominant and non-dominant hands (r=0.814, p<0.001) (**Table 6**).

Table 6. Pearson correlation between BMI and handgrip strength measurements		
Variables	**p-value**	**r-value**
BMI and dominant hand HGS	0.001**	0.320
BMI and non-dominant hand HGS	0.003**	0.276
Dominant hand HGS and non-dominant hand HGS	<0.001**	0.814
p<0.05 Statistically significant (*), p<0.001 highly statistically significant (**), BMI: Body-mass index, HGS: Handgrip strength		

DISCUSSION

In this study, the effects of smoking status on HGS in young adults were thoroughly examined, taking into account measurement sequence, hand dominance, and biological sex. The results revealed that current smokers exhibited higher grip strength in the initial trials compared to other groups; however, this advantage markedly declined in subsequent measurements. The performance drop observed in the third trial was consistent across both sexes and in both dominant and non-dominant hands, clearly underscoring the adverse impact of smoking on muscular endurance and fatigue resistance.

Kopiczko et al.[17] reported significantly lower handgrip strength among young female smokers. Our study supports this observation, showing a 4.3 kg decline in the non-dominant hand of female smokers across repeated trials. This similarity suggests that female smokers may experience faster muscular fatigue due to smoking-induced metabolic stress. Similarly, Jiang et al.[8] found reduced hand and finger strength in smokers. Our male participants demonstrated a comparable pattern, with active smokers experiencing up to 4.0 kg reductions in repeated grip trials, particularly in the non-dominant hand. These results reinforce the idea that smoking has a generalized detrimental effect on localized muscle endurance.

Importantly, the methodological design of the study was structured to minimize potential confounding factors by ensuring sample homogeneity. Participants were of the same educational level, free from chronic diseases or mobility limitations, and, based on supporting socioeconomic data, generally came from similar financial and social backgrounds. These criteria enhance the internal validity of the study by reducing the influence of extraneous variables such as BMI, comorbidities, or unequal access to resources. In line with this, positive and statistically significant correlations were found between BMI and handgrip strength in both dominant and non-dominant hands, which supports the physiological relevance of body composition in interpreting muscle performance outcomes. Therefore, the observed differences in handgrip strength can be more confidently attributed to smoking status rather than to external confounders.

The pronounced fatigue difference observed specifically in non-dominant (left) hand measurements suggests that smoking may negatively affect not only maximal force production but also neuromuscular coordination and motor control mechanisms. These findings are consistent with existing literature that emphasizes smoking's detrimental impact on peripheral muscle function and circulatory efficiency. Additionally, results from left-hand dominant participants indicate that regardless of motor dominance, smoking significantly impairs performance in short-term, repetitive physical tasks, affecting not only peak strength but also fatigue resistance and recovery capacity.

Table 5. Left-hand dominant female participants-handgrip strength (kg) by measurement order and hand dominance						
Hand/trial	**Never smoked (mean±SD)**	**Former smokers (mean±SD)**	**Current smokers (mean±SD)**	**Fatigue difference**	**p-value**	**r-value**
Left (dominant)-1st trial	34.2±5.1	33.7±5.5	36.0±5.9		0.07	0.131
Left (dominant)-3rd trial	32.8±5.0	32.2±5.3	32.2±5.7	3.8 kg	0.05*	0.272
Right (non-dominant)-1st	32.7±4.9	32.4±5.4	34.3±5.6		0.09	0.114
Right (non-dominant)-3rd	31.5±4.8	30.8±5.2	30.0±5.5	3.0 kg	0.07	0.221
p<0.05 statistically significant (*), p<0.001 highly statistically significant (**), SD: Standard deviation						

Nie et al.[2] and Hendriks et al.[21] identified a link between reduced grip strength and cognitive decline. The marked fatigue observed in our study, particularly in less-used limbs, may reflect early impairments in neuromuscular coordination, which are known to be associated with neurocognitive decline. These parallels suggest that HGS might act as a behavioral biomarker not only for muscular health but also for neurophysiological integrity in smokers.

Previous studies have shown that HGS is inversely correlated with numerous chronic diseases such as cardiovascular disease,[12] metabolic syndrome,[26,27] COPD,[9] neurological disorders,[2] and cancer.[8] Our study concretizes these relationships by demonstrating that smoking's pathophysiological effects can be captured through HGS. Furthermore, low HGS is associated with sarcopenia, loss of functional independence, and increased mortality risk.[5,11]

Smoking induces harmful effects on skeletal muscle through oxidative stress, chronic inflammation, and mitochondrial dysfunction.[23] Moreover, tissue hypoxia caused by carbon monoxide and other toxins accelerates muscle fatigue.[16] The decline in HGS observed across both hands in smokers aligns with these physiological mechanisms. Peterson et al.[1] demonstrated that lower grip strength correlates with accelerated biological aging, as measured by DNA methylation. The acute fatigue observed in our smoker group, despite higher initial strength, may reflect a diminished cellular recovery capacity indicative of early epigenetic aging processes. This aligns with the concept that smoking accelerates biological wear and tear.

Gurel et al.[26] and Wang et al.[27] emphasized the relationship between low HGS and cardiometabolic disorders such as hypertension and insulin resistance. Our findings of diminished grip performance in smokers, even among young adults, may indicate early manifestations of impaired metabolic health. This supports the utility of HGS as an early, non-invasive screening tool for at-risk individuals.

Bikbov et al.[28] found that HGS correlates with retinal nerve fiber thickness and ocular perfusion. The observed performance loss in our study may not only indicate muscular fatigue but also suggest systemic vascular impacts of smoking that affect peripheral and possibly ocular microcirculation.

Moreover, HGS has been correlated with critical outcomes such as fall risk, osteoporosis, and quality of life,[17,20] underscoring its value in community health screening. Reuter et al.[10] confirmed the reliability of repeated HGS measurements. In our study, the consistent decline in HGS across three trials provides further evidence for its sensitivity in detecting fatigue-related impairments in muscle function. This validates the use of HGS in behavioral and physiological research involving smoking.

At the societal level, socioeconomic disadvantage has been associated with both higher smoking prevalence and lower HGS.[16,19] White et al.[16] highlighted the interplay between smoking and socioeconomic status. Since our sample was homogenous in age, education, and socioeconomic background, the observed differences in HGS can more confidently be attributed to smoking rather than social disparities. This strengthens the internal validity of our findings.

Celis-Morales et al.[12] reported that HGS is a strong predictor of all-cause mortality. The significant grip strength decline observed in our smoker group may not only reflect short-term muscular fatigue but also forecast longer-term systemic health risks. This underscores the public health importance of incorporating HGS screening in smoking-related interventions.

In conclusion, our findings show that smoking has acute and significant adverse effects on HGS, impacting both immediate force generation and post-effort recovery capacity. HGS should be considered not only as a physical performance measure but as a multifaceted indicator of systemic health. Incorporating HGS measurement into public health strategies could serve as a powerful tool for monitoring behavioral risk factors like smoking and promoting earlier interventions.

Limitations

This study has several limitations. Firstly, due to its cross-sectional design, causal relationships cannot be strongly established. Smoking status was assessed through self-reported data and was not biologically verified. Since the sample consisted solely of university students from the same faculty, the generalizability of the findings is limited. Additionally, certain variables that may affect muscle performance-such as nutritional status and sleep patterns-were not controlled.

Nevertheless, the study demonstrates the negative effects of smoking on muscular endurance and recovery capacity, highlighting the potential value of handgrip strength as a biomarker of systemic health. Future longitudinal and controlled studies involving different age groups and multiple centers are recommended to enhance the validity of these findings.

CONCLUSION

This study provides important insights into the relationship between smoking, HGS, muscular endurance, and overall physical capacity. The findings indicate that individuals who smoke exhibit significantly lower HGS levels in both dominant and non-dominant hands, and that this disparity becomes more pronounced following short rest periods. This suggests that smoking negatively affects not only instantaneous muscle strength but also the recovery capacity of the muscles. In the existing literature, HGS is recognized not only as a measure of muscle function but also as a biomarker for a broad spectrum of health indicators, including cardiovascular risk, cognitive decline, epigenetic aging, and mortality. The multifaceted impact of smoking on HGS reinforces its destructive consequences on systemic health. Accordingly, HGS should not merely be viewed as a clinical measurement but as a holistic assessment tool for lifestyle factors and chronic disease risk. The observed effects of smoking on muscular strength underscore the necessity of incorporating HGS screening into public health policies and supporting smoking cessation programs with such physiological biomarkers. Early screening, particularly among young adults and middle-aged populations, can facilitate timely interventions and support

strategies aimed at preserving quality of life. In conclusion, this study reveals that smoking leads to substantial physiological, neurological, and functional impairments. Handgrip strength emerges as a practical and effective tool for identifying these adverse effects. Future longitudinal research may further clarify the causal pathways of this relationship and aid in the development of targeted intervention strategies.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was carried out with the permission of the Hitit University Non-interventional Researches Ethics Committee (Date: 06.01.2025, Decision No: 2024-27).

Informed Consent

Written informed consent was obtained from all participants prior to data collection.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

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

Author Contributions

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

REFERENCES

1. Peterson MD, McGrath R, Zhang P, Wang Y. Grip strength is inversely associated with DNA methylation age acceleration. *J Cachexia Sarcopenia Muscle.* 2022;13(2):955-963. https://doi.org/10.1002/jcsm.12935
2. Nie J, Xu Y, Xue J, Yu D, Zhao Y, Hu C. The relationship between grip strength and cognitive impairment: evidence from NHANES. *Brain Behav.* 2025;15(1):e3365. doi:10.1002/brb3.3365
3. Ramlagan S, Peltzer K, Phaswana-Mafuya N. Hand grip strength and associated factors in non-institutionalised men and women 50 years and older in South Africa. *BMC Res Notes.* 2014;7:8. doi:10.1186/1756-0500-7-8
4. Rantanen T, Guralnik JM, Foley D, et al. Midlife hand grip strength as a predictor of old age disability. *JAMA,* 1999;281(6):558-560. doi:10.1001/jama.281.6.558
5. Rantanen T, Volpato S, Ferrucci L, Heikkinen E, Fried LP, Guralnik JM. Handgrip strength and cause-specific and total mortality in older disabled women: exploring the mechanism. *J Am Geriatr Soc.* 2003;51(5):636-641. doi:10.1034/j.1600-0579.2003.00207.x
6. Gubelmann C, Vollenweider P, Marques-Vidal P. Association of grip strength with cardiovascular risk markers. *Eur J Prevent Cardiol.* 2016;24(5):496-504. doi:10.1177/2047487316674140
7. Hu X, Jiang Y, Shen Y, Ren H, Tang W. Association of handgrip strength with all-cause and cause-specific mortality: a prospective cohort study. *Am J Prevent Med.* 2019;57(2):182-190. doi:10.1016/j.amepre.2019.03.004
8. Jiang L, Zhang L, Lin T. Body composition and weakness of hand grip strength and pinch strength in patients with upper extremity fracture. *J Human Nutrit Dietet.* 2020;33(5):708-715. doi:10.1111/jhn.12785
9. Johansson G, Ahmadi Z, Sundh J, Janson C. Hand grip strength is associated with fatigue among men with COPD: epidemiological data from northern Sweden. *J Chronic Obstruct Pulmon Dis.* 2022;19(1):60-67. doi:10.1080/15412555.2022.2037249
10. Reuter CP, Da Silva MC, Renner JDP, Mello ED. Reliability and validity of indices of hand-grip strength and endurance. *Austr Occupat Ther J.* 2011;58(1):36-43. doi:10.1111/j.1440-1630.2010.00899.x
11. Guralnik JM, Ferrucci L, Simonsick EM, Salive ME, Wallace RB. Lower-extremity function in persons over the age of 70 years as a predictor of subsequent disability. *N Engl J Med.* 2000;332(9):556-562. doi:10.1056/nejm199503023320902
12. Celis-Morales CA, Welsh P, Lyall DM, et al. Associations of grip strength with cardiovascular, respiratory, and cancer outcomes and all-cause mortality: prospective cohort study of half a million UK Biobank participants. *BMJ.* 2018;361:k1651. doi:10.1136/bmj.k1651
13. Al-Obaidi SM, Anthony J, Al-Shuwai N, Dean E. Differences in functional performance among smokers and nonsmokers. *Physiotherapy.* 2014;100(2):156-160. doi:10.1016/j.physio.2013.09.004
14. Reimers CD, Knapp G, Reimers AK. Does physical activity increase life expectancy? A review of the literature. *Nutrients.* 2016;8(5):295. doi:10.3390/nu8050295
15. Tseng LA, Delmonico MJ, Visser M, Kritchevsky SB, Newman AB, Harris TB. Body composition and physical performance in old age: the health ABC study. *PloS One.* 2018;13(7):e0197909. doi:10.1371/journal.pone.0197909
16. White A, Amos A, Hawkins J. Smoking and socioeconomic status in England: the rise of the never smoker and the disadvantaged smoker. *BMJ Open.* 2016;6(11):e010155. doi:10.1136/bmjopen-2015-010155
17. Kopiczko A, Łopuszańska-Dawid M, Gryko K. Bone mineral density and handgrip strength as predictors of osteoporosis in elderly women. *Scient Rep.* 2018;8:13259. doi:10.1038/s41598-018-31306-1
18. Singh GK, Daus GP, Allender M, et al. Social determinants of health and tobacco use: findings from the 2021 national health interview survey. *Int J Equity Health.* 2022;21:85. doi:10.1186/s12939-022-01673-9
19. Kim YS, Park SH. Grip strength asymmetry and risk of falls in older adults: a population-based cohort study. *Scient Rep.* 2018;8:13259. doi:10.1038/s41598-018-31306-1
20. Kaur J, Jain S, Dhamija S, Rana N. Investigation of hand muscle fatigue and its influential factors in manual tasks. *Ergonomics.* 2021;64(12):1565-1574. doi:10.1080/00140139.2021.1943155
21. Jamal A, Phillips E, Gentzke AS, et al. Current Cigarette Smoking Among Adults - United States, 2016. *Morbidity Mortality Weekly Rep.* 2018;67(2):53-59. doi:10.15585/mmwr.mm6702a1
22. United States Public Health Service Office of the Surgeon General, & National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health. (2020). Smoking Cessation: A Report of the Surgeon General. US Department of Health and Human Services.
23. Hecht SS. Tobacco carcinogens, their biomarkers and tobacco-induced cancer. Nature reviews. *Cancer.* 2003;3(10):733-744. doi:10.1038/nrc1190
24. Hendriks S, Ranson JM, Peetoom K, et al. Risk factors for young-onset dementia in the UK Biobank. *JAMA Neurol.* 2024;81(2):134-142. doi:10.1001/jamaneurol.2023.4441
25. Rossi PD, Zammit AR, Schwartz BS, Gross AL. Grip strength and brain volume: the role of cognition and physical function in aging. *J Alzheimer's Dis.* 2018;65(3):1037-1047. doi: 10.3233/JAD-180011
26. Gurel F, Turan S, Aydin E, Kaya H, Çalışkan S. Handgrip strength and cardiometabolic health in middle-aged adults. *Nutrients.* 2023;15(4):918. doi:10.3390/nu15040918
27. Wang D, Liu Y, Zhang X, Hu G. Low handgrip strength and metabolic syndrome in Chinese adults: a nationwide study. *Nutrients.* 2024;16(10):1951. doi:10.3390/nu16101951
28. Bikbov MM, Zainullin RM, Gilmanshin TR, et al. Hand grip strength and ocular associations: the Ural eye and medical study. *Br J Ophthalmol.* 2023;107(10):1567-1574. doi:10.1136/bjo-2022-321404

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1746967
J Health Sci Med. 2025;8(5):871-878

Impact of refugee status and maternal age on pregnancy outcomes: a comparative study from Turkiye

Gülay Balkaş, Şevki Çelen
Department of Perinatology, Etlik Zübeyde Hanım Women's Health Training and Research Hospital, Ankara, Turkiye

Cite this article as: Balkaş G, Çelen Ş. Impact of refugee status and maternal age on pregnancy outcomes: a comparative study from Turkiye. *J Health Sci Med.* 2025;8(5):871-878.

Received: 20.07.2025 ◆ **Accepted**: 23.08.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: This study aimed to elucidate sociodemographic, obstetric, and neonatal outcomes among Syrian refugee and native Turkish adolescents compared with adult women, to determine whether adverse maternal and neonatal outcomes are primarily attributable to young maternal age or exacerbated by Syrian refugee status.

Methods: This retrospective cohort study analyzed 23,832 deliveries at a tertiary care center in Turkiye's state-funded healthcare system from January 2015 to December 2018, with 5,720 singleton pregnancies meeting inclusion criteria. The study population comprised 2,235 adolescent pregnant women aged 15-19 years (985 Syrian refugees and 1,250 Turkish natives) and 3,485 adult pregnant women aged 20-35 years. Sociodemographic characteristics, obstetric and neonatal outcomes were compared between the groups.

Results: Refugee adolescents were younger, had a lower body mass index, insufficient weight gain during pregnancy and had inadequate antenatal care compared to other groups ($p<0.001$). Smoking and being unmarried were more prevalent among native adolescents ($p<0.001$), however the illiteracy rate was significantly higher among refugee adolescents ($p<0.001$). First- and second-trimester prenatal screening tests uptake was lower among refugee adolescents, but gestational diabetes mellitus (GDM) screening rates were higher ($p<0.001$). The cesarean delivery rate was higher among adult pregnant women, while vaginal deliveries were more common among refugee adolescents ($p<0.001$). Refugee adolescents experienced higher rates of pre-eclampsia and preterm birth, whereas GDM was more prevalent among adults ($p<0.001$).

Conclusion: Adolescent pregnancy is associated with significant maternal and neonatal health risks, exacerbated among Syrian refugee adolescents due to socioeconomic disparities, language barriers, limited education, and limited access to healthcare facilities. Targeted interventions are urgently needed to improve antenatal care and provide comprehensive education on fertility and reproductive health, support family planning, and prevent child marriage.

Keywords: Adolescent pregnancy, adverse pregnancy outcomes, antenatal care, health disparities, native adolescents, refugee adolescents

INTRODUCTION

Adolescence, a critical transition from childhood to adulthood marked by significant psychological and social changes, encompasses approximately 17% of the global population.[1] The World Health Organization (WHO) defines adolescent pregnancy as occurring in females aged 10-19 years, with those aged 10-14 classified as younger adolescents.2 According to the WHO, as of 2019, approximately 21 million adolescent pregnancies occurred annually among women aged 15-19 in developing countries, with 50% being unintended and resulting in an estimated 12 million live births.2 Despite the steady decline in the adolescent fertility rate over the past two decades, it remains at 11.6%, with significant regional disparities, particularly in low- and middle-income countries.[2,3]

According to the Turkish Demographic and Health Survey, 3.5% of adolescents in Turkiye had given birth, 1.5% had been married before the age of 15, and 0.2% had delivered a child before reaching that age.[4] However, a study conducted in seven hospitals across Turkiye between 2015 and 2017 reported that 7.9% of all births occurred during adolescence.[5] Although pregnancy- and childbirth-related complications remain the leading causes of death among adolescents in the region, adolescent pregnancy has not been directly identified as a cause of maternal mortality in Turkiye.[6,7]

Adolescent pregnancy is associated with significant physical and psychological challenges that adversely affect maternal and neonatal health.[8] Research indicates that pregnant adolescents face an increased risk of complications, including anemia, nutritional deficiencies, inadequate weight gain, spontaneous abortion, preterm birth, pre-eclampsia, birth trauma, and various labor interventions.[9] Moreover, adolescent mothers exhibit higher rates of postpartum complications, such as

Corresponding Author: Gülay Balkaş, dr.gulaybalkas@gmail.com

hemorrhage, postpartum depression, and challenges related to urinary, sexual, and breastfeeding health.[10] Neonates born to adolescent mothers are also at an increased risk of adverse outcomes, including prematurity, low birth weight, meconium aspiration, respiratory distress syndrome, hypoglycemia, jaundice, and mortality.[11]

According to the United Nations High Commissioner for Refugees, approximately 27.1 million refugees were recorded worldwide as of mid-2021, with 83% hosted in low- and middle-income countries.[12] Syria constitutes the largest source of refugees, representing 27% of the global refugee population, while Turkiye hosts the highest number, with 3.8 million individuals.[12] Data from the Turkish Presidency for Migration Management indicate that 47.6% of Syrian refugees in Turkiye are women, and 18.8% are adolescent girls aged 10-18 years.[13]

In addition to biological immaturity, pregnant adolescents particularly those from vulnerable populations such as refugees, are at increased risk of adverse maternal and neonatal outcomes.[14] These outcomes are exacerbated by distinct social challenges, including linguistic barriers, low educational attainment or illiteracy, legal constraints, inadequate nutrition and limited access to healthcare services.[15,16] These factors influence pregnancy outcomes to varying degrees, prompting ongoing discourse regarding the relative contributions of sociodemographic and economic determinants compared to intrinsic age-related risks.

This study aimed to elucidate maternal, obstetric, and neonatal outcomes among Syrian refugee and native Turkish adolescents (15-19 years) compared with adult women (20-35 years). It addressed the following research questions: 1. How does adolescent pregnancy influence maternal health outcomes? 2. How does adolescent pregnancy affect neonatal health outcomes? 3. How does refugee status among adolescents impact maternal and neonatal outcomes? 4. Are adverse outcomes in refugee adolescents primarily attributable to young maternal age or exacerbated by socio-environmental factors?

METHODS

Ethics

This study was conducted in accordance with the Declaration of Helsinki and was approved by the Etlik Zübeyde Hanım Women's Health Training and Research Hospital Clinical Researches Ethics Committee (Date: 25.08.2021, Decision No: 2012/88). Due to the retrospective nature of the study, the local ethics committee waived the requirement for informed consent.

Study Design and Eligibility Criteria

This retrospective cohort study analyzed 5,720 singleton pregnancies among women aged 15-35 years who underwent antenatal assessments and delivered at a tertiary care center within Turkiye's state-funded healthcare system between January 2015 and December 2018. Participants were categorized into two groups based on maternal age at delivery; the adolescent group (15-19 years) and the adult group (20-35 years). The adolescent group was then subdivided into two subgroups: refugee adolescents and native adolescents. Patients with incomplete data, multiple gestations, or pre-existing systemic diseases were excluded from the analysis.

Data Collection and Assessment Criteria

All data were obtained from the hospital's electronic database and patient medical records. Information was collected on sociodemographic and maternal antenatal characteristics, including age, educational level, marital status, type of marriage (formal or religious), pre-pregnancy body-mass index (BMI), parity, gestational weight gain, frequency of antenatal follow-up visits, and rates of iron, folic acid, and vitamin D supplementation. Hemoglobin (Hb) and hematocrit (Hct) levels were assessed at the time of admission for delivery. Gestational age was determined based on the last menstrual period and confirmed by fetal crown-rump length measured via ultrasonography between 11 and 14 weeks of gestation.[17] The first-trimester combined screening test was performed between 11 and 14 weeks, while the second-trimester triple screening test was conducted between 16 and 20 weeks of gestation.[18] Hyperemesis gravidarum was defined as persistent nausea and vomiting accompanied by positive urinary ketones on urinalysis and a weight loss exceeding 5% of pre-pregnancy body weight.

Obstetric outcomes, including type of delivery, indications for cesarean delivery (CD), oligohydramnios and polyhydramnios, were evaluated.[19,20] Pre-eclampsia and eclampsia were diagnosed according to the guidelines of the American College of Obstetricians and Gynecologists.[21] All participants underwent routine screening for gestational diabetes mellitus (GDM) between 24 and 28 weeks of gestation, with diagnoses made in accordance with the criteria of the American Diabetes Association.[22] Preterm delivery was defined as delivery before 37 completed weeks of gestation.[23] Premature rupture of membranes (PROM) was defined as the rupture of the membranes prior to the onset of uterine contractions, whereas preterm premature rupture of membranes (PPROM) referred to PROM occurring before 37 weeks of gestation.[24,25] Neonatal outcomes, including gestational age at birth, birth weight, neonatal intensive care unit (NICU) admission rates, and Apgar scores at 1 and 5 minutes, were recorded.

Statistical Analysis

Data were analyzed using the Statistical Package for the Social Sciences (SPSS), version 28.0 (IBM Corporation, Armonk, NY, USA). Continuous variables were reported as mean±standard deviation (SD) for normally distributed data or as median (minimum-maximum) for non-normally distributed data. Categorical variables were presented as frequencies and percentages. Continuous variables that were normally distributed were assessed using one-way analysis of variance (ANOVA) and the independent-sample T test, while those that were not normally distributed were evaluated using the Kruskal-Wallis and Mann-Whitney U tests. Categorical variables were compared across three groups (native Turkish adolescents, Syrian refugee adolescents and adult women) using the Pearson Chi-squared test. Post-hoc pairwise comparisons were performed when significant differences were detected. Statistical significance was defined as a p-value <0.05.

RESULTS

The study analyzed 23,832 deliveries at the study hospital during the specified period, of which 5,720 met the inclusion criteria for live singleton pregnancies. Of these, 2,235 were pregnant adolescent women aged 15-19 (985 refugees and 1,250 natives), and 3,485 were pregnant adult women aged 20-35. Sociodemographic and antenatal characteristics of the study population are presented in Table 1. Refugee adolescents represented the youngest group, with a median age of 17 years (15-19), compared to 18 years (15-19) among native adolescents and 29 years (20-35) among adult women (p<0.001). Refugee adolescents exhibited lower pre-pregnancy BMI, inadequate gestational weight gain, significantly lower Hb and Hct levels, and attended fewer than four antenatal care visits (p<0.001). Active smoking and unmarried status were more common among native adolescents compared to refugee adolescents and adult women (p<0.001). Screening for GDM was more frequently performed among refugee adolescents (p<0.001), whereas participation in first-trimester combined screening and second-trimester triple screening tests was significantly lower in this group (p<0.001).

Obstetric characteristics and outcomes are presented in Table 2. Refugee adolescents had significantly higher rates of vaginal delivery (79.2%), while CD was more common among adult women (54.1%) (p<0.001). Emergency CD due to fetal distress was the most frequent indication among adolescents, while elective CD due to a history of previous CD was more common among adults (p<0.001). Pre-eclampsia was more prevalent among refugee adolescents than among native adolescents and adults (p<0.001). No statistically significant differences were observed between the groups in terms of hyperemesis gravidarum, eclampsia, postpartum hemorrhage, and polyhydramnios.

Neonatal outcomes are shown in Table 3. Refugee adolescents had a significantly lower gestational age at birth (38±4.5 weeks), lower birth weight (3089±246 grams), and a higher preterm birth rate compared with native adolescents and adults (21%, 17.7%, and 13.9%, respectively) (all p<0.001).

Table 1. Sociodemographic and antenatal characteristics of the study population

Variables	Refugee adolescents[1] (n=985)	Native adolescents[2] (n=1250)	Adults[3] (n=3485)	p-value	p-value (1 vs. 2)	p-value (1 vs. 3)	p-value (2 vs. 3)
Age (years) (min-max)	17 (15-19)	18 (15-19)	29 (20-40)	<0.001[c]	<0.001[a]	<0.001[a]	<0.001[a]
BMI (kg/m²) (mean±SD)	20.8±3.1	22.3±2.9	24.5±3.9	<0.001[d]	<0.001[b]	<0.001[b]	<0.001[b]
Education level (n, %)							
Uneducated	595 (60.4)	106 (8.5%)	110 (3.2%)				
Primary education	382 (38.8)	1125 (90%)	3275 (93.9%)	<0.0011	<0.0011	<0.0011	<0.0011
High school	8(0.8%)	19 (1.5%)	100 (2.9%)				
Marital status (n, %)							
Married	601 (61%)	655 (52.4%)	3242 (93%)	<0.0011	<0.0011	<0.0011	<0.0011
Single	384 (39%)	595 (47.6%)	243 (7%)				
Type of marriage (n, %)							
Formal	60 (10%)	104 (15.9%)	3005 (92.7%)	<0.0011	<0.0011	<0.0011	<0.0011
Religious	541 (90%)	551 (84.15)	237 (7.3%)				
Active smoking (n, %)	48 (4.9%)	225 (18%)	419 (12%)	<0.0011	<0.0011	<0.0011	<0.0011
Parity rate (mean±SD)	1.8±0.9	1.2±0.6	1.5±0.7	<0.001[d]	<0.001[d]	<0.001[d]	<0.001[d]
Parity (n, %)							
Nulliparous	325 (33.1%)	510 (40.8%)	974 (27.9%)	<0.0011	<0.0011	<0.0011	<0.0011
Multiparous	660 (66.9%)	740 (59.2%)	2511 (72.1%)				
Antenatal follow-up times (n, %)							
<4	455 (46.2%)	400 (32)	585 (16.8)	<0.0011	<0.0011	<0.0011	<0.0011
≥4	530 (53.8%)	850(68%)	2900 (83.2%)				
Weight gain during pregnancy (mean±SD)	8.45±4.50	10.35±3.45	12.97±3.69	<0.001[d]	<0.001[d]	<0.001[d]	<0.001[d]
Hemoglobin before birth (g/dl) (min-max)	10.48 (7.8-14.8)	11.05 (6.8-16.5)	11.85 (5.7-16.6)	<0.0011	<0.001[a]	<0.001[a]	0.062[a]
Hematocrit (mean±SD)	33.65±2.3	34.85±3.2	35.68±2.6	0.001[d]	0.025[d]	<0.001[d]	0.002[d]
Folic acid intake (n, %)	304 (30.9%)	765 (61.2%)	2750 (78.9%)	<0.0011	<0.0011	< 0.0011	<0.0011
Iron replacement therapy (n, %)	376 (38.2%)	1025 (82%)	2958 (84.9%)	<0.0011	<0.0011	<0.0011	<0.0531
Vitamin D supplementation (n, %)	245 (27.9%)	950 (74.4%)	2987 (85.7%)	0.0011	<0.0011	<0.0011	<0.0011
First-trimester combined test (n, %)	180 (18.3%)	325 (26%)	1680 (48.2%)	<0.0011	<0.0011	<0.0011	<0.0011
Second-trimester triple test (n, %)	122 (12.4%)	248 (19.8%)	848 (24.3%)	<0.0011	<0.0011	<0.0011	<0.0011
GDM screening test (<37 weeks) (n, %)	178 (18.1%)	156 (12.5%)	530 (15.2%)	<0.0011	<0.0011	<0.0011	<0.0011

min: Minimum, max: Maximum, BMI: Body-mass index, SD: Standard deviation, GDM: Gestational diabetes mellitus, [1]Pearson's Chi-square test statistics, [a]Mann-Whitney test, bindependent sample T test, [c]Kruskal-Wallis test, [d]One-way ANOVA test and a significant p-value is <0.05

Table 2. Maternal characteristics and obstetric outcomes

Variables	Refugee adolescents[1] (n=985)	Native adolescents[2] (n=1250)	Adults[3] (n=3485)	p-value	p-value (1 vs. 2)	p-value (1 vs. 3)	p-value (2 vs. 3)
Mode of delivery (n, %)							
Spontaneous vaginal	780 (79.2%)	761 (60.9%)	1601 (45.9%)	<0.001	<0.001	<0.001	<0.001
Cesarean delivery	205 (20.8%)	489 (39.1%)	1884 (54.1%)				
CD indications (n, %)							
Previous CD	46 (4.7%)	71 (5.7%)	324 (17.2%)	<0.001	<0.001	<0.001	<0.001
Fetal distress	69 (7%)	106 (8.5%)	180 (9.6%)				
CPD	53 (5.4%)	80 (6.4%)	95 (5%)				
Breech presentation	3 (1.5%)	7 (1.4%)	89 (4.7%)				
Others	34 (2.2%)	225 (17.1%)	1196(17.6%)				
Hyperemesis gravidarum (n, %)	30 (3%)	27 (2.7%)	109 (3.1%)	0.154	0.128	0.254	0.185
Pre-eclampsia (n, %)	97 (9.8%)	56 (4.5%)	226 (6.5%)	<0.001	<0.001	<0.001	<0.001
Eclampsia (n, %)	1 (0.1%)	1 (0.08%)	3 (0.09%)	0.775	0.257	0.458	0.653
GDM (n, %)	11 (1%)	14 (1.1%)	195 (5.6%)	<0.001	0.062	<0.001	<0.001
PPROM (n, %)	34 (3.5%)	40 (3.2%)	73 (2.1%)	<0.001	0.563	<0.001	<0.001
PROM (n, %)	20 (2%)	26 (2.1%)	80 (2.3%)	0.622	0.452	0.832	0.654
Postpartum hemorrhage (n, %)	3 (0.3%)	4 (0.32%)	10 (0.29%)	0.093	0.123	0.158	0.232
Oligohydramnios (n, %)	69 (7%)	86 (6.9%)	150 (3.5%)	<0.001	0.854	<0.001	<0.001
Polyhydramnios (n, %)	22 (2.2%)	35 (2.8%)	101 (2.9%)	0.675	0.753	0.563	0.853

CD: cesarean delivery, CPD: Cephalopelvic disproportion, GDM: gestational diabetes mellitus, PPROM: preterm premature rupture of membranes, PROM: premature rupture of membranes, Pearson's chi-square test was used for all comparisons and a significant p-value is <0.05.

Table 3. Neonatal outcomes of the study population

Variables	Refugee adolescents[1] (n=985)	Native adolescents[2] (n=1250)	Adults[3] (n=3485)	p-value	p-value (1 vs. 2)	p-value (1 vs. 3)	p-value (2 vs. 3)
Gestational age at birth (weeks) (mean±SD)	38±4.50	38.8±3.2	39.4±3.2	<0.001a	<0.001a	<0.001a	<0.001a
Preterm delivery (<37 weeks) (n, %)	207 (21%)	221 (17.7%)	485 (13.9%)	<0.0011	<0.0011	<0.0011	<0.0011
Birth weight (grams) (mean±SD)	3089±246	3260±631	3465±435	<0.001a	<0.001a	<0.001a	<0.001a
Low birth weight (<2500 grams) (n, %)	147 (14.9%)	123 (9.8%)	157 (4.5%)	<0.0011	<0.0011	<0.0011	<0.0011
Apgar score (<7 at 1st minute) (n, %)	71 (7.2%)	69 (5.5%)	181 (5.2%)	<0.0011	<0.0011	<0.0011	0.7541
Apgar score (<7 at 5th minute) (n, %)	29 (2.9%)	21 (1.7%)	66 (1.9%)	<0.0011	<0.0011	<0.0011	0.5421
Neonatal intensive care unit admission (n, %)	88 (8.9%)	101 (8.1%)	261 (7.5%)	<0.0011	<0.0011	<0.0011	<0.0011

SD: Standard deviation, 1Pearson's Chi-square test statistics, aKruskal-Wallis test and a significant p-value is <0.05.

DISCUSSION

According to the WHO, adolescent pregnancies constitute approximately 11% of global births, predominantly occurring in low- and middle-income countries.2 Approximately 50% of these pregnancies are unintended, often leading to unsafe abortions or neglected antenatal care.[26] Due to biological immaturity and insufficient antenatal care, adolescents face elevated risks of adverse obstetric outcomes, rendering adolescent pregnancy a critical global public health concern.[27] The Syrian civil war has exacerbated these challenges, with millions of refugee adolescents encountering economic, social, and linguistic barriers, as well as limited access to healthcare facilities in host countries.[28,29]

Adequate antenatal care is essential for reducing maternal and perinatal morbidity and mortality by directly detecting and managing complications and identifying women at risk during labor and delivery.[30] The Republic of Turkiye has established a state-funded healthcare system for Syrian refugees, providing access to health facilities beyond refugee camps. Within this framework, pregnant refugees receive free antenatal care, including vitamin and iron supplementation, in accordance with the Ministry of Health's routine antenatal care guidelines, which recommend a minimum of four follow-up visits for low-risk pregnancies.[31] Despite sustained efforts, significant gaps in antenatal care uptake have persisted over the past 15 years, particularly among Syrian refugee adolescents. In our study, 46.2% of refugee adolescents attended fewer than four antenatal follow-up visits, compared to 68.0% of native adolescents and 83.2% of adult women who completed four or more visits. Similarly, Erenel et al.[32] reported that 41.3% of pregnant Syrian refugees received no antenatal care.

Furthermore, illiteracy significantly exacerbated barriers to antenatal care access, with 60.4% of refugee adolescents illiterate compared to 8.5% of native adolescents and 3.2% of adult women (p<0.001), aligning with high illiteracy rates among Syrian refugees reported by Demirci et al.[33]

Our results revealed significant differences in prenatal screening rates. Prenatal aneuploidy screening uptake was significantly lower among refugee adolescents, with 18.3% completing the first-trimester combined test and 12.4% the second-trimester triple test (p<0.001), consistent with rates of 16.5% and 11.9%, respectively, reported by Golbasi et al.[29] In contrast, their participation in GDM screening was notably higher. This disparity may be attributed to cultural influences, as some Turkish women reportedly decline GDM testing due to misconceptions regarding the safety of glucose loading and its potential risks to infants. These findings indicate that, beyond structural barriers, cultural attitudes and levels of health literacy significantly shape screening behaviors.[34]

Child marriage, a prevalent issue among Syrian refugees, contributes to early childbearing and elevated adolescent pregnancy rates. UNICEF reported a rise in child marriage among Syrian refugees in Jordan from 18% in 2012 to 32% in 2014, compared to 13% pre-conflict in Syria.[35,36] While child marriage was an accepted cultural practice in Syria prior to the crisis, the subsequent economic collapse and the inability to afford education or secure employment have intensified this practice. It is often seen as a socially acceptable strategy for achieving economic stability in host countries. Furthermore, refugee adolescents frequently have limited knowledge of family planning and lack the health literacy necessary to use contraceptives effectively.[37] As a result, and consistent with previous research,[29,33] our findings demonstrated that refugee adolescents had a lower maternal age (17 years) and a higher parity rate (1.8±0.9) compared to native adolescents (18 years, 1.2±0.6) (p<0.001).

Preterm birth, affecting approximately 13 million infants annually, is a multifactorial complication extensively studied globally.[38] The current literature presents conflicting evidence regarding preterm birth risk among adolescents aged 15-19 years, with some studies reporting elevated risk compared to adult women,[39] while others find no statistically significant difference.[40] Biological factors, including low maternal age, immature uterine development, short cervical length (<25 mm), progesterone dysregulation and anemia associated with nutritional deficiencies, have been identified as contributors to this risk.[41] Moreover, beyond biological immaturity, sociodemographic determinants, including marital status, poverty, and limited access to antenatal care further exacerbate the likelihood of preterm birth among adolescents.[9,10] Consistent with prior literature, this study demonstrated a significantly higher rate of preterm birth among refugee adolescents (21%) compared to native adolescents (17%) (p<0.001).

Furthermore, newborns of refugee adolescents exhibited significantly lower birth weights, reduced 1st- and 5th-minute Apgar scores, and higher rates of NICU admission compared to those of native adolescents and adults (p<0.001). These adverse outcomes may be attributed to multiple factors, including preterm birth, inadequate supplementation with essential vitamins and iron, and maternal psychological distress such as post-traumatic stress disorder. Prior research has established associations between micronutrient deficiencies, maternal mental health conditions, and adverse neonatal outcomes.[42,43]

This study demonstrated pronounced disparities in CD rates by maternal age and refugee status, with significantly higher rates observed among adult women (54.1%) compared to native adolescents (39.1%) and refugee adolescents (20.8%) (p<0.001). Consistent with previous studies,[33,44] the Turkish Demographic and Health Survey reported CD rates of 51.2% among adults and 33.3% among adolescents.4 Several hypotheses have been proposed to explain the lower CD rates in adolescents. Jolly et al.[45] suggested that enhanced myometrial contractility and greater connective tissue elasticity in adolescents facilitate spontaneous vaginal delivery. The higher rates of preterm birth among refugee (21%) and native adolescents (17.7%) compared to adults (13.9%) (p<0.001), along with significantly lower birth weights (3089±450 g and 3260±420 g vs. 3465±430 g, respectively; p<0.001), may contribute to the increased likelihood of vaginal delivery in adolescent groups. However, Zeteroğlu et al.[46] found no significant influence of biological immaturity on CD rates in a study of 40,391 pregnancies.

Beyond biological factors, socio-environmental barriers significantly influence obstetric outcomes among refugee adolescents. Prior literature identifies financial limitations, transportation difficulties, fear of mistreatment, security concerns, cultural stigma, limited availability of female healthcare providers and linguistic barriers as impediments to accessing antenatal care.[47,48] In line with these findings, the present study found significantly higher vaginal delivery rates among refugee adolescents (79.2%) compared to native adolescents (60.9%) and adult women (45.9%) (p<0.001). This disparity may reflect restricted access to elective or emergency CD interventions. Previous research has demonstrated that limited antenatal care and systemic obstacles in conflict-affected or refugee settings often hinder timely obstetric interventions.[49,50]

Pre-eclampsia, a major contributor to maternal and perinatal morbidity, poses heightened risks for adolescents. Current literature presents conflicting evidence, with some studies reporting an elevated pre-eclampsia risk in adolescents,[5,51] while others find no significant difference.[10,41,52] The pathogenesis of pre-eclampsia in adolescents may differ from that in older women, possibly linked to biological factors such as uterine immaturity that may contribute to defective deep placentation.[53,54] Moreover, Martinez et al.[55] reported higher prevalence rates of pre-eclampsia in adolescent pregnancies in low- and middle-income countries (11.5% and 10.6%, respectively) compared to the global estimate of 6.7%, with lower rates in very high-income countries (5.1%). Consistent with existing literature, this study identified a significantly higher pre-eclampsia rate among refugee adolescents (9.8%) compared to native adolescents (4.5%) and adult women (6.5%) (p<0.001). This disparity may be partially attributed to maternal vitamin D and iron deficiencies, compounded by barriers to antenatal care and limited health literacy.[56,57]

Previous studies have established associations between micronutrient deficiencies and increased pre-eclampsia risk.[58]

This study has several critical strengths that substantially advance the current understanding of adolescent pregnancy outcomes in low- and middle-income contexts, particularly among refugee populations. First, the large, well-characterized cohort of 5,720 singleton pregnancies, drawn from a total of 23,832 deliveries, includes three analytically distinct subgroups: Syrian refugee adolescents, native Turkish adolescents, and adult women. This design provides sufficient statistical power to detect clinically meaningful differences in obstetric and neonatal outcomes. Second, the study's comparative framework enables the disentanglement of the effects of age and refugee status within Turkiye's universal, state-funded healthcare system-an important context given Turkiye's status as host to the world's largest refugee population. Third, the inclusion of comprehensive sociodemographic, obstetric, and neonatal outcomes addresses critical gaps in understanding the interplay of biological and socio-environmental factors. Finally, this study addresses a critical gap in global reproductive health literature by focusing on refugee adolescents, a high-risk, understudied population, thereby elucidating how socioeconomic vulnerabilities exacerbate biological risks, providing evidence-based insights for targeted maternal health interventions in humanitarian contexts.

Limitations

Several important limitations of this study should be acknowledged. Its retrospective design may have resulted in incomplete data and limits the ability to draw causal inferences. Using data from a single tertiary care center, which typically manages high-risk pregnancies, may have introduced selection bias, potentially inflating rates of adverse outcomes. Furthermore, access to comprehensive antenatal services within Turkiye's high-resource public healthcare system may have attenuated some adverse outcomes, thereby limiting the detection of disparities. Unmeasured confounders, such as psychosocial stressors, restrict the ability to fully quantify or explore the mechanisms underlying the observed outcomes. Additionally, the findings may have limited generalizability to other refugee populations or healthcare systems with differing capacities. Prospective, multicenter studies incorporating mixed-methods approaches are needed to identify optimal interventions for adolescent pregnancies, particularly among vulnerable refugee populations.

CONCLUSION

Adolescent pregnancy among refugee populations presents significant risks to maternal and neonatal health due to compounded biological and social vulnerabilities. Our findings highlight the critical need for integrated, culturally sensitive healthcare strategies that transcend clinical management to address these challenges. First, enhancing antenatal care accessibility through community-based outreach, mobile health units, and interpreter services is essential to minimizing delays in care. Second, comprehensive reproductive health education tailored to adolescents and their families must be prioritized to address misconceptions and promote informed decision-making regarding contraception and family planning. Third, policy frameworks should ensure equitable healthcare coverage for refugee adolescents by reducing financial and structural barriers through national insurance schemes and cross-sector collaborations. Additionally, policymakers should support prospective multicenter studies and systematic evaluations of interventions to inform evidence-based strategies. Addressing adolescent pregnancy in refugee populations requires a dual clinical and policy approach, ensuring that both health outcomes and broader social determinants are improved in tandem.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was carried out with the permission of the Etlik Zübeyde Hanım Women's Health Training and Research Hospital Clinical Researches Ethics Committee (Date: 25.08.2021, Decision No: 2012/88).

Informed Consent

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

Referee Evaluation Process

Externally peer reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

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

Author Contributions

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

Acknowledgment

The authors extend their gratitude to the pregnant women who participated in the study.

REFERENCES

1. Sawyer SM, Azzopardi PS, Wickremarathne D, Patton GC. The age of adolescence. *Lancet Child Adolesc Health*. 2018;2(3):223-228. doi:10.1016/S2352-4642(18)30022-1
2. World Health Organization (WHO). Adolescents Pregnancy. Updated April 10, 2024. Accessed January 16, 2023. https://www.who.int/news-room/fact-sheets/detail/adolescent-pregnancy.
3. Akombi-Inyang BJ, Woolley E, Iheanacho CO, Bayaraa K, Ghimire PR. Regional trends and socioeconomic predictors of adolescent pregnancy in Nigeria: a nationwide study. *Int J Environ Res Public Health*. 2022;19(13):8222. doi:10.3390/ijerph19138222
4. Hacettepe University Institute of Population Studies. Turkiye Demographic and Health Survey 2018. Accessed June 15, 2019. https://dhsprogram.com/publications/publication-FR372-DHS-Final-Reports.cfm.
5. Bas EK, Bulbul A, Uslu S, Bas V, Elitok GK, Zubarioglu U. Maternal Characteristics and Obstetric and Neonatal Outcomes of Singleton Pregnancies Among Adolescents. *Med Sci Monit Int Med J Exp Clin Res*. 2020;26:e919922-1. doi:10.12659/MSM.919922

6. Karacam Z, Kizilca Cakaloz D, Demir R. The impact of adolescent pregnancy on maternal and infant health in Turkiye: systematic review and meta-analysis. *J Gynecol Obstet Hum Reprod*. 2021;50(4):102093. doi:10.1016/j.jogoh.2021.102093
7. Isguder CK, Arslan O, Gunkaya OS, Kanat-Pektas M, Tug N. Adolescent pregnancies in Turkiye: a single center experience. *Ann Saudi Med*. 2024;44(1):11-17. doi:10.5144/0256-4947.2024.11
8. World Health Organization (WHO). Maternal, Newborn, Child and Adolescent Health: Stillbirth. Accessed April 25, 2018. https:// www.who.int/maternal_child_adolescent/epidemiology/stillbirth/en/
9. Zhang T, Wang H, Wang X, et al. The adverse maternal and perinatal outcomes of adolescent pregnancy: a cross sectional study in Hebei, China. *BMC Pregnancy Childbirth*. 2020;20(1):339. doi:10.1186/s12884-020-03022-7
10. Rexhepi M, Besimi F, Rufati N, Alili A, Bajrami S, Ismaili H. Hospital-based study of maternal, perinatal and neonatal outcomes in adolescent pregnancy compared to adult women pregnancy. *Maced J Med Sci*. 2019; 7(5):760-766. doi:10.3889/oamjms.2019.210
11. Marvin-Dowle K, Soltani H. A comparison of neonatal outcomes between adolescent and adult mothers in de-veloped countries: a systematic review and meta-analysis. *Eur J Obstet Gynecol Reprod Biol:X*. 2020;6:100109. doi:10.1016/j.eurox.2020.100109
12. UNHCR Refugee data finder. Accessed April 16, 2023. https://www.unhcr.org/refugee-statistics/
13. Republic of Turkiye Ministry of Interior, Presidency of Migration Management. Annual Migration Reports. Accessed May 10, 2023. https://www.goc.gov.tr/reports3
14. Urindwanayo D, Richter S. Teenage pregnancy in refugee camps: a narrative synthesis. *J Int Womens Stud*. 2020;21(1):255-270.
15. Gibson-Helm M, Boyle J, Cheng IH, East C, Knight M, Teede H. Maternal health and pregnancy outcomes among women of refugee background from Asian countries. *Int J Gynecol Obstet*. 2015;129(2):146-151. doi:10.1016/j.ijgo.2014.10.036
16. Winn A, Hetherington E, Tough S. Caring for pregnant refugee women in a turbulent policy landscape: perspectives of health care professionals in Calgary, Alberta. *Int J Equity Health*. 2018;17(1):91. doi:10.1186/s12939-018-0801-5
17. The Fetal Medicine Foundation. Online education: the 11-13 weeks scan. Accessed August 20, 2019. https://fetalmedicine.org/education/the-11-13-weeks-scan
18. Rose NC, Kaimal AJ, Dugoff L, Norton ME, American College of Obstetricians and Gynecologists. Screening for fetal chromosomal abnormalities: ACOG practice bulletin, number 226. *Obstet. Gynecol.* 2020;136(4):e48-e69. doi:10.1097/AOG.0000000000004084
19. Morris RK, Meller CH, Tamblyn J, et al. Association and prediction of amniotic fluid measurements for adverse pregnancy outcome: systematic review and meta-analysis. *BJOG Int J Obstet Gynaecol*. 2014; 121(6):686-699. doi:10.1111/1471-0528.12589
20. Hamza A, Herr D, Solomayer EF, Meyberg-Solomayer G. Polyhydramnios: causes, diagnosis and therapy. *Geburtshilfe Frauenheilkd*. 2013;73(12):1241–1246. doi:10.1055/s-0033-1360163
21. American College of Obstetricians and Gynecologists. Gestational hypertension and preeclampsia: ACOG practice bulletin, number 222. *Obstet. Gynecol*. 2020;135(6):e237-e260.
22. Care D. 5. facilitating positive health behaviors and well-being to improve health outcomes: standards of care in diabetes-2024. *Diabetes Care*. 2024;47(Suppl 1):S77-S110. doi:10.2337/dc24-S005
23. Giouleka SM, Tsakiridis I, Kostakis N, et al. Preterm labor: a comprehensive review of guidelines on diagnosis, management, prediction and prevention. *Obstet. Gynecol Surv*. 2022;77(5):302-317. doi:10.1097/OGX.0000000000001023
24. Pettker MD. Prelabor Rupture of Membranes. ACOG practice bulletin, number 217. *Obstet. Gynecol*. 2020;135(6): e237-e248. doi:10.1097/AOG.0000000000003700
25. Kumar D, Moore RM, Mercer BM, Mansour JM, Redline RW, Moore JJ. The physiology of fetal membrane weakening and rupture: Insights gained from the determination of physical properties revisited. *Placenta*. 2016;42:59-73. doi:10.1016/j.placenta.2016.03.015
26. Mohamed S, Chipeta MG, Kamninga T, et al. Interventions to prevent unintended pregnancies among adolescents: a rapid overview of systematic reviews. *Systematic Reviews*. 2023;12(1);198. doi:10.1186/s13643-023-02361-8
27. Early Childbearing and Teenage Pregnancy Rates by Country. UNICEF DATA. Accessed January 29, 2025. https://data.unicef.org/topic/child-health/adolescent-health/
28. DeJong J, Sbeity F, Schlecht J, et al. Young lives disrupted: gender and well-being among adolescent Syrian refugees in Lebanon. *Confl Health*. 2017;11(Suppl 1):23. doi:10.1186/s13031-017-0128-7
29. Golbasi C, Vural T, Bayraktar B, Golbasi H, Yildirim AGS. Maternal and neonatal outcomes of Syrian adolescent refugees and local adolescent Turkish citizens: a comparative study at a tertiary care maternity hospital in Turkiye. *GORM*. 2022;28(2):135-143. doi:10.21613/GORM.2021.1186
30. Alibhai KM, Ziegler BR, Meddings L, Batung E, Luginaah I. Factors impacting antenatal care utilization: a systematic review of 37 fragile and conflict-affected situations. *Confl Health*. 2022;16(1):33. doi:10.1186/s13031-022-00459-9
31. Turkish Ministry of Health. Prenatal Care Management Guide. Accessed April 13, 2018. https://sbu.saglik.gov.tr/ Ekutuphane/kitaplar/dogumonubakim.pdf
32. Erenel H, Aydogan Mathyk B, Sal V, Ayhan I, Karatas S, Koc Bebek A. Clinical characteristics and pregnancy outcomes of Syrian refugees: a case-control study in a tertiary care hospital in Istanbul, Turkiye. *Arch Gynecol* 2017;295(1):45-50. doi:10.1007/s00404-016-4188-5
33. Demirci H, Yildirim Topak N, Ocakoglu G, Karakulak Gomleksiz M, Ustunyurt E, Ulku Turker A. Birth characteristics of Syrian refugees and Turkish citizens in Turkiye in 2015. *Int J Gynaecol Obstet*. 2017;137(1): 63-66. doi:10.1002/ijgo.12088
34. Balkas G, Çelen S. Early prediction of gestational diabetes mellitus and insulin therapy requirement using first-trimester PAPP-A and free β-hCG MoMs levels: a retrospective case-control study. *J Clin Med*. 2024;13(24):7725. doi:10.3390/jcm13247725
35. UNICEF. A study on early marriage in Jordan 2014. Accessed August 19, 2018. https://www.unicef.org/media/files/UNICEFJordan_EarlyMarriageStudy2014-email.pdf
36. Syrian Central Bureau of Statistics, League of Arab States. Family Health Survey in Syrian Arab Republic-2009, 2011. https://edirc.repec.org/data/stagvsy.html
37. West L, Isotta-Day H, Ba-Break M, Morgan R. Factors in use of family planning services by Syrian women in a refugee camp in Jordan. *J Fam Plann Reprod Health Care*. 2017;43(2):96-102. doi:10.1136/jfprhc-2014-101026
38. Simmons LE, Rubens CE, Darmstadt GL, Gravett MG. Preventing preterm birth and neonatal mortality: exploring the epidemiology, causes, and interventions. *Semin Perinatol*. 2010;34(6):408-415. doi:10.1053/j.semperi.2010.09.005
39. Jahromi BN, Daneshvar A. Pregnancy outcome of parturients below 16 years of age. *Saudi Med J*. 2005;26(9):1417-1419.
40. Usynina AA, Postoev V, Odland JØ, Grjibovski AM. Adverse pregnancy outcomes among adolescents in Northwest Russia: a population registry-based study. *Int J Environ Res Public Health*. 2018;15(2):261. doi:10.3390/ijerph15020261
41. Korenčan S, Pinter B, Grebenc M, Verdenik I. The outcomes of pregnancy and childbirth in adolescents in Slovenia. *Zdr Varst*. 2017;56(4):268.
42. Alderdice F, McNeill J, Lynn F. A systematic review of systematic reviews of interventions to improve maternal mental health and well-being. *Midwifery*. 2013;29(4):389-399. doi:10.1016/j.midw.2012.05.010
43. Gernand AD, Schulze KJ, Stewart CP, West Jr KP, Christian P. Micronutrient deficiencies in pregnancy worldwide: health effects and prevention. *Nat Rev Endocrinol*. 2016;12(5):274-289. doi:10.1038/nrendo.2016.37
44. Buyuktiryaki M, Canpolat FE, Alyamac Dizdar E, Okur N, Kadıoglu Simsek G. Neonatal outcomes of Syrian refugees delivered in a tertiary hospital in Ankara, Turkiye. *Confl Health*. 2015;9(1):38. doi:10.1186/s13031-015-0066-1
45. Jolly MC, Sebire N, Harris J, Robinson S, Regan L. Obstetric risks of pregnancy in women less than 18 years old. *Obstet Gynecol*. 2000;96(6): 962-966. doi:10.1016/S0029-7844(00)01075-9
46. Zeteroglu S, Sahin I, Gol K. Cesarean delivery rate in adolescent pregnancy. *Eur Contracept Reprod Health Care*. 2005;10(2):119-122. doi:10.1080/13625180500131600
47. Ganchimeg T, Ota E, Morisaki N, et al. Pregnancy and childbirth outcomes among adolescent mothers: a World Health Organization multicountry study. *BJOG Int J Obstet Gynaecol*. 2014;121:40-48. doi:10.1111/1471-0528.12630

48. Loto OM, Ezechi OC, Kalu BKE, Loto AB, Ezechi LO, Ogunniyi SO. Poor obstetric performance of teenagers: is it age-or quality of care-related? *J Obstet Gynaecol.* 2004;24(4):395-398. doi:10.1080/01443610410001685529

49. Benova L, Tuncalp O, Moran AC, Campbell OMR. Not just a number: examining coverage and content of antenatal care in low-income and middle-income countries. *BMJ Glob Health*. 2018;3(2):e000779. doi:10.1136/bmjgh-2018-000779

50. Strong J, Varady C, Chahda N, Doocy S, Burnham G. Health status and health needs of older refugees from Syria in Lebanon. *Confl health*. 2015;9(1):12. doi:10.1186/s13031-014-0029-y

51. Rupakala BM, Shruthi AG, Nagarathnamma R. A study on teenage pregnancy and its maternal and fetal outcome. *Int J Sci Res*. 2013;5(5):2486-2489.

52. Karai A, Gyurkovits Z, Nyári TA, Sári T, Németh G, Orvos H. Adverse perinatal outcome in teenage pregnancies: an analysis of a 5-year period in Southeastern Hungary. *J Matern Fetal Neonatal Med.* 2019;32(14):2376-2379. doi:10.1080/14767058.2018.1438393

53. Brosens I, Muter J, Ewington L, et al. Adolescent preeclampsia: pathological drivers and clinical prevention. *Reprod Sci.* 2019;26(2):159-171. doi:10.1177/1933719118804412

54. Jeha D, Usta I, Ghulmiyyah L, Nassar A. A review of the risks and consequences of adolescent pregnancy. *J Neonatal-Perinat Med*. 2015;8(1):1-8. doi:10.3233/NPM-1581403

55. Zangiacomi Martinez E, da Roza DL. Ecological analysis of adolescent birth rates in Brazil: association with Human Development Index. *Women Birth*. 2020;33(2):e191-e198. doi:10.1016/j.wombi.2019.04.002

56. Oh C, Keats EC, Bhutta ZA. Vitamin and mineral supplementation during pregnancy on maternal, birth, child health and development outcomes in low- and middle-income countries: a systematic review and meta-analysis. *Nutrients*. 2020;12(2):491. doi:10.3390/nu12020491

57. Vašková J, Klepcová Z, Špaková I, et al. The importance of natural antioxidants in female reproduction. *Antioxidants*. 2023;12(4):907. doi:10.3390/antiox12040907

58. Palacios C, De-Regil LM, Lombardo LK, Peña-Rosas JP. Vitamin D supplementation during pregnancy: updated meta-analysis on maternal outcomes. *J Steroid Biochem Mol Biol*. 2016;164:148-155. doi:10.1016/j.jsbmb.2016.02.008

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1752483
J Health Sci Med. 2025;8(5):879-883

Evaluation of the performance of current Artificial Intelligence Chatbots regarding patient information after coronary artery bypass surgery

Gökhan Yüksel[1], Selami Gürkan[2]

[1]Department of Nursing, Faculty of Health Sciences, Tekirdağ Namık Kemal University, Tekirdağ, Turkiye
[2]Department of Cardiovascular Surgery, Tekirdag Namık Kemal University, Tekirdağ, Turkiye

Cite this article as: Yüksel G, Gürkan S. Evaluation of the performance of current Artificial Intelligence Chatbots regarding patient information after coronary artery bypass surgery. *J Health Sci Med.* 2025;8(5):879-883.

Received: 28.07.2025 ◆ **Accepted**: 25.08.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: This study aims to evaluate the performance of existing Artificial Intelligence (AI) driven Chatbots regarding patient information after coronary artery bypass (CABG) surgery.

Methods: On July 1, 2025, a standardized medical prompt concerning the recovery process after CABG was submitted to ten prominent AI Chatbots: GPT-4o, GPT 4.1, Grok-4, Claude Opus-4, DeepSeek R1, Gemini Pro, Microsoft Copilot, Llama 4, Mistral Large 2, and Perplexity Sonar. Each response was assessed using two validated scoring systems; the modified Ensuring Quality Information for Patients (mEQIP) and the newly developed Quality Analysis of Medical Artificial Intelligence (QAMAI). Readability was evaluated using the average reading level consensus (ARLC) calculator, which aggregates eight standard readability formulas.

Results: Among the tested Chatbots, Perplexity Sonar achieved the highest mEQIP score (91.7%) and the highest QAMAI score (29/30), while Gemini Pro received the lowest scores in both evaluations (72.2% mEQIP, 25/30 QAMAI). The average mEQIP score across all platforms was 80.43%, and the mean QAMAI score was 27/30, indicating generally high-quality responses. Readability assessment revealed that DeepSeek R1 provided the most comprehensible content (ARLC: 9.92, equivalent to a reading age of 15-16 years), while Llama 4 produced the most complex output (ARLC: 14.69, age 23+). The average ARLC across all Chatbots was 11.9, which corresponds to a college-level reading difficulty and exceeds the recommended sixth to eighth-grade readability level for patient education materials.

Conclusion: AI Chatbots show promising capabilities in delivering post-CABG patient information, often achieving high scores in quality assessments. However, inconsistencies remain in readability, completeness, and source transparency. Despite the increasing sophistication of AI-generated health information, the elevated reading levels and inconsistent citation practices may hinder accessibility for general patient populations. To enhance their role in patient education, future Chatbot iterations should prioritize user-centered design, medical guideline compliance, and content simplification.

Keywords: Coronary artery bypass grafting, Artificial Intelligence, Chatbots

INTRODUCTION

Coronary artery disease (CAD) continues to be one of the most important health problems of our time. In 2013, CAD was the predominant cause of death in Turkiye, accounting for 38.8% of mortalities.[1] This national burden is similar to trends in other high-income areas of the world: 2.3 million individuals in the United Kingdom (UK) have CAD. Coronary artery disease is the second leading cause of death in the UK.[2] Coronary Artery Bypass Grafting (CABG) is the primary intervention for severe CAD, including triple vessel disease and substantial left main stenosis.[3] While there has been a recent decline in all cardiac revascularization procedures, over 200,000 CABG surgeries are performed in the United States annually.[4] Given the serious consequences of incorrect or delayed treatment for CAD, ensuring the quality and accessibility of online information about CABG is crucial for helping many patients better understand the topic and for the postoperative care process for those undergoing this surgery.[5]

The internet has become a primary source of patient information, particularly during and after the COVID-19 pandemic, as people increasingly turn to it for timely health updates, and access to hospitals remains highly limited. More than 5% of internet searches are related to health. It is common for people to research symptoms or treatment plans online before consulting a doctor[6]. Although internet information is presented in several formats to accommodate individuals from varied backgrounds, the reliability of the sources remains uncertain. Misinformation may exacerbate

Corresponding Author: Gökhan Yüksel, gyuksel@nku.edu.tr

worry and uncertainty, impede patients from seeking prompt medical care, and erode trust in healthcare personnel.[7] Artificial Intelligence (AI)-driven Chatbots have emerged as a significant source of health information in the digital era.[8] These systems offer swift, anonymous assistance and practical guidance, making them particularly attractive to patients who are unable to consult a healthcare expert directly. Their accessibility and easy-to-use and accessible features have contributed to their increasing appeal in addressing health-related concerns. However, the accuracy, consistency, and quality of the information provided must be rigorously ensured to enable patients to use it effectively. AI-driven Chatbots may offer reliable disease-related information, evidence-based treatment options, and insights into potential issues.[9,10]

This study aims to evaluate and compare the performance, readability, and text quality of current AI Chatbots in patient information after CABG surgery.

METHODS

On July 1, 2025, the following prompt was entered one by one into 10 current Chatbots at Tekirdağ Namık Kemal University, Department of Cardiovascular Surgery, and the responses were recorded. Since no living or human data were used in this study, ethics committee approval was not obtained. All procedures were carried out in accordance with the ethical rules and the principles. Chatbots used: GPT-4o, GPT 4.1, Grok-4, Claude Opus-4, DeepSeek R1, Gemini Pro, Microsoft Copilot, Llama 4, Mistral Large 2, and Perplexity Sonar. New, unused accounts were opened, and access to licensed versions of the previously used ones was provided. A detailed prompt was established for CABG, directing each Chatbot to produce a medically precise and comprehensible text aimed at a public readership. The prompt was organized as follows: 'What should be done during the recovery process after cardiac bypass surgery (CABG)? Please provide information on the necessary steps to be taken during the recovery process. Kindly verify that the passage is medically precise and adheres to the most recent norms and recommendations in cardiovascular surgery. Can you use medical terminology while maintaining clarity for a general audience?'

The modified Ensuring Quality of Information for Patients (mEQIP) and Quality Analysis of Medical Artificial Intelligence (QAMAI) tools were used to assess the quality of the produced texts. The EQIP tool was developed in 2004 and comprises 36 items that assess specific characteristics of dependability, content quality, and structural readability.[11] It has been widely applied in recent years to evaluate information sources of different subspecialties in medicine, including surgery, dermatology, internal medicine, urology, and infectious diseases.[12-14] Only two yes-or-no alternatives were offered for each issue to eliminate assessor subjectivity in incomplete responses. The option "N/A" was offered for elements that were not relevant to the kind of source. The score is categorized as follows: 0-25%: critical quality issues; 26-50%: significant quality concerns; 51-75%: satisfactory quality with minor deficiencies; 76-100%: proficiently composed. This systematic scoring framework guarantees an impartial and uniform assessment of answer quality.

The QAMAI tool was created using the Modified DISCERN (mDISCERN) instrument. The mDISCERN is a well-established and extensively used instrument for evaluating the quality of health information disseminated via websites, social networks, YouTube, and other multimedia platforms. The use of mDISCERN for assessing information generated by AI is infeasible, as the tool considers specific human attributes, such as board certification and the credibility of the content provider, which do not apply to AI. In parallel with the mDISCERN, a panel of experts decided to develop a unidimensional construct of the instrument with six items, assessed by Likert Scales. The six domains of information quality were posited to be interrelated, like mDISCERN, reflecting a single dimension: the quality of the information's content. Each aspect was assessed using a Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The QAMAI included six items: accuracy, clarity, relevance, completeness, sources, and usefulness. The scores were aggregated into a comprehensive metric (QAMAI score) that assessed the quality of the information. The total QAMAI score was evaluated between 6 and 30. Quality score assessments were scored, and consensus was reached by an experienced cardiovascular surgeon (SG) and an experienced cardiovascular clinic nurse (GY). Kappa statistics were used to assess the degree of concordance among the two evaluators (Pearson correlation coefficient r=0.976).

After the quality consensus among cardiovascular specialists, each passage generated by the Chatbot was assessed for reading difficulty using the average reading level consensus (ARLC) calculator. This online calculator, available at https://readabilityformulas.com/calculator-arlic-formula.php, computes the average of eight established readability formulae. The formulae include the Automated Readability Index, Flesch-Kincaid grade level, Flesch reading ease, Gunning Fog Index, Coleman-Liau Readability Index, SMOG Index, FORCAST Readability Formula, and Linsear Write Readability Index. The calculator then produces a difficulty score based on educational levels, spanning from very simple (first grade, ages 6-7) to highly tough (college graduate, age 23 and above).[15] Since the number of words, syllables, and sentences was used in these formulas, these were also noted separately.

RESULTS

Among the 10 Chatbots compared, Perplexity Sonar had the highest mEQIP score (91.7%) and Gemini Pro had the lowest (72.2%). The distribution of mEQIP scores by category is shown in **Table 1**.

Among the 10 Chatbots compared, like mEQIP scores, Perplexity Sonar had the highest QAMAI score (29) and again Gemini Pro had the lowest (25) score. The distribution of mEQIP scores by category is shown in **Table 2**.

When we examine the readability evaluation, Deepseek produced the most easily understandable text (ARLC: 9.92), while Llama was the Chatbot with the most complex language to understand (ARLC: 14.69) (**Table 3**).

Table 1. Comparison of mEQIP scores of Chatbots with subgroups

	Content (out of 36)	Identification (out of 12)	Structure (out of 24)	Total	%
GPT-4o	34	2	23	59	73.6
GPT 4.1	35	5	23	63	87.5
Grok-4	34	4	23	60	83.3
Claude Opus-4	31	2	22	55	76.4
DeepSeek R1	34	2	22	58	80.6
Gemini Pro	30	2	20	52	72.2
Microsoft Copilot	33	2	22	57	79.2
Llama 4	33	2	23	58	80.6
Mistral Large 2	32	2	23	57	79.2
Perplexity Sonar	34	10	22	66	91.7

mEQIP: Modified Ensuring Quality Information for Patients

Table 3. Comparison of language complexity criteria of Chatbots

	ARLC	Reading level	Age range
DeepSeek R1	9.92	Somewhat difficult	15-16
GPT 4.1	10.56	Fairly difficult	16-17
GPT-4o	10.94	Fairly difficult	16-17
Perplexity Sonar	11.15	Fairly difficult	16-17
Claude Opus-4	11.20	Fairly difficult	17-18
Grok-4	11.77	Difficult	17-18
Microsoft Copilot	12.14	Difficult	17-18
Mistral Large 2	12.66	Very difficult	18-20
Gemini Pro	13.80	Professional	21-22
Llama 4	14.69	Extremely difficult	23+

ARLC: Average reading level consensus

DISCUSSION

This study is among the first investigations in the literature to assess the quality of replies generated by various AI Chatbot models in the postoperative CABG surgery era. Our study's primary finding is that the Chatbots exhibit a wide range of reading and quality levels. These levels pertain to works comprehensible and readable by persons with a university-level education. The second primary result is that the Chatbots exhibit significant variations in the mEQIP and QAMAI scores. The Perplexity Sonar achieved a substantially superior mEQIP and QAMAI score compared to the other Chatbots, ChatGPT and Gemini. The mEQIP scores revealed that all Chatbots attained a mean score of 80.43%, and a QAMAI score of 27 (out of 30). When evaluating the quality of the texts, we revealed that current Chatbots produce more qualified content than their predecessors, and many of them exceed the upper quality threshold of 75%.[15,16] However, when we examine the language simplicity, the average ARLC score of 11.9 suggests that the texts are too academic and difficult to understand for the majority of the patients. If the text is written for patient education or public information purposes, further simplification is recommended. The American Medical Association (AMA) and the National Institutes of Health (NIH) recommend that health materials should be at a 6-8 grade level.[17] This is because many adults may struggle to understand complex terms or long sentence structures. As such, current Chatbots still use language structures that are too complex for patient information. Earlier research indicates that health literacy plays a crucial role in adherence to treatment and overall health outcomes.[18] This implies that the AI-generated materials currently available might not be beneficial for everyone, especially older adults or individuals with lower levels of education.

We assessed the quality, precision, and readability of the information offered by ten of the leading Chatbots in response to a standardized medical inquiry concerning postoperative care after CABG surgery. The results indicated that performance differed significantly among the platforms. Perplexity Sonar received the highest ratings in both the mEQIP and QAMAI assessments, while Gemini Pro consistently ranked last. These findings illustrate that the medical content produced by Chatbots is not always uniform, highlighting the necessity for caution when using AI-driven tools for patient education.

The mEQIP tool indicated that the majority of Chatbots delivered information ranging from good to excellent quality. Nevertheless, there were significant inconsistencies in the identification and structural aspects that impacted the overall usability. For instance, GPT 4.1 excelled in maintaining structure, while Gemini Pro struggled with clarity and completeness. It is crucial to recognize that if the information

Table 2. Comparison of QAMAI scores of Chatbots with subcategories

	Accuracy	Clarity	Relevance	Completeness	Sources	Usefulness	Total
GPT-4o	5	5	5	4	2	5	26
GPT 4.1	5	5	5	5	2	5	27
Grok-4	5	5	5	5	2	5	27
Claude Opus-4	5	4	5	5	1	5	25
DeepSeek R1	5	5	5	5	2	5	28
Gemini Pro	5	4	5	4	2	5	25
Microsoft Copilot	5	5	5	5	3	5	28
Llama 4	5	5	5	5	3	5	28
Mistral Large 2	5	5	5	4	3	5	27
Perplexity Sonar	5	4	5	5	5	5	29

QAMAI: Quality Analysis of Medical Artificial Intelligence

provided is not well-structured or sufficiently comprehensive, patients may struggle to understand or retain important recovery guidance.

Additionally, while some Chatbots provided well-structured and research-backed information, most responses lacked citations for their sources. Various studies examining AI Chatbots across different medical fields have also highlighted this issue.[19] This suggests that although AI-generated information might appear valuable, it is often difficult to authenticate. Only Perplexity Sonar consistently provided sources, with only a handful of others attempting to support their statements with references. Consequently, users had no clear way to verify the accuracy or recency of the information presented. This research builds on earlier findings that indicate AI-driven health communication tools are evolving rapidly, yet remain far from perfect. Chatbots such as GPT-4o and Grok-4 demonstrated strong performance in terms of accuracy and clarity; however, variations among different platforms highlight the necessity of ongoing monitoring.[5]

To improve current Chatbots' designs, various actions can be taken. Adaptive readability algorithms that adjust the text's difficulty according to the patient's reading level, ensuring medical accuracy is not compromised, should be implemented. Clear citation methods are required so that users can check the accuracy and timeliness of the material. Organized checklists or bullet-point summaries for important postoperative directives should be added to make the information more thorough. The ability to adapt to diverse cultures and languages, facilitating patient education that is sensitive to these differences, should be included. Lastly, user feedback loops should be utilized to enable Chatbots to continually improve based on patients' understanding and satisfaction with them.

Our findings indicate that while current Chatbots offer valuable information about post-CABG surgery, their reliability is compromised by several factors; they are often difficult to understand, fail to cite sources consistently, and lack consistent material. Targeted enhancements, such as adaptive readability restrictions, verified sources, and organized, patient-centered material, might make them an important part of postoperative treatment. Following health literacy standards and evidence-based criteria for outputs is a clear way to help patients understand, follow, and get better results. As digital health becomes more popular, improving the design of Chatbots is not simply a technological improvement; it is also an important step toward safe and fair patient education.

Limitations

Firstly, it assessed a singular standardized prompt so that outcomes may vary with alternative inquiries or clinical scenarios. Secondly, despite the enhanced interrater reliability, the scoring method retains an element of subjectivity. Third, readability formulae may inadequately reflect the clarity or cultural suitability of the content. Moreover, Chatbot replies may fluctuate over time, and the inclusion of only ten prevalent English-language models constrains generalizability. Subsequent research should examine various prompts, incorporate additional Chatbot systems, and analyze outcomes in diverse languages.

Although text quality measures like BLEU, ROUGE-L, BERTScore, and METEOR, along with factual accuracy and hallucination rates, are significant in NLP research, their use in patient-facing medical information remains unstandardized and unvalidated. The fact that these scores were not evaluated in our study is another limitation. Likewise, measurements of technical efficiency, such as response length, generating time, and repetition rate, were excluded, as these factors are not within the scope of our emphasis on the quality and readability of medical material. Finally, differences in model architecture, training data, and techniques such as reinforcement learning from human feedback (RLHF) or retrieval-augmented generation were not analyzed due to limited transparency from developers, and thus, performance differences could not be correlated with these technical specifications. Reproducibility is also limited because Chatbot APIs (Application Programming Interface) are frequently updated, and developers do not consistently disclose technical details such as API version or parameter settings.

CONCLUSION

AI Chatbots demonstrate potential in delivering postoperative information following CABG surgery; yet, their responses differ in quality, comprehensiveness, and readability. Although the majority provide generally dependable content, the reading level exceeds the suggested standards for patient education, and the citation of sources is variable. Future Chatbot development should prioritize enhanced clarity of language, consistent reference, adherence to therapeutic guidelines, and human oversight to augment their use.

ETHICAL DECLARATIONS

Ethics Committee Approval

No ethical approval was needed because this is not a human study, but only online information was used.

Informed Consent

No informed consent was obtained because no patient or their information was used.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

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

Author Contributions

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

REFERENCES

1. Turkey Statistical Agency 2013 Statistical Analysis, 1 April 2014 issue: 16162.
2. Bhatnagar P, Wickramasinghe K, Wilkins E, Townsend N. Trends in the epidemiology of cardiovascular disease in the UK. *Heart*. 2016;102(24): 1945-1952. doi:10.1136/heartjnl-2016-309573
3. Melly L, Torregrossa G, Lee T, Jansens JL, Puskas JD. Fifty years of coronary artery bypass grafting. *J Thorac Dis*. 2018;10(3):1960-1967. doi: 10.21037/jtd.2018.02.43
4. Montrief T, Koyfman A, Long B. Coronary artery bypass graft surgery complications: a review for emergency clinicians. *Am J Emerg Med*. 2018;36(12):2289-2297. doi:10.1016/j.ajem.2018.09.014
5. Wai Tai C, Wing Kwan L, Chan J, Angelini GD. Ensuring quality information for patients tool to assess patient information on CABG websites: systemic search and evaluation. *Perfusion*. 2025;40(6):1468-1476. doi:10.1177/02676591241303842
6. Hesse BW, Nelson DE, Kreps GL, et al. Trust and sources of health information: the impact of the Internet and its implications for health care providers: findings from the first Health Information National Trends Survey. *Arch Intern Med*. 2005;165(22):2618-2624. doi:10.1001/archinte.165.22.2618
7. Silver MP. Patient perspectives on online health information and communication with doctors: a qualitative study of patients 50 years old and over. *J Med Internet Res*. 2015;17(1):e19. doi:10.2196/jmir.3588
8. King MR. The future of AI in medicine: a perspective from a Chatbot. *Ann Biomed Eng*. 2023;51(2):291-295. doi:10.1007/s10439-022-03121-w
9. Şahin MF, Doğan Ç, Topkaç EC, et al. Which current Chatbot is more competent in urological theoretical knowledge? A comparative analysis by the European board of urology in-service assessment. *World J Urol*. 2025;43(1):116. doi:10.1007/s00345-025-05499-3
10. Şahin MF, Topkaç EC, Doğan Ç, et al. Still using only ChatGPT? The comparison of five different Artificial Intelligence Chatbots' answers to the most common questions about kidney stones. *J Endourol*. 2024; 38(11):1172-1177. doi:10.1089/end.2024.0474
11. Moult B, Franck LS, Brady H. Ensuring quality information for patients: development and preliminary validation of a new instrument to improve the quality of written health care information. *Health Expect*. 2004;7(2):165-175. doi:10.1111/j.1369-7625.2004.00273.x
12. Anıl H, Kayra MV. The digital dialogue on premature ejaculation: evaluating the efficacy of Artificial Intelligence-driven responses. *Int Urol Nephrol*. 2025;57(9):2829-2836. doi:10.1007/s11255-025-04461-x
13. Melloul E, Raptis DA, Oberkofler CE, Dutkowski P, Lesurtel M, Clavien PA. Donor information for living donor liver transplantation: where can comprehensive information be found? *Liver Transpl*. 2012;18(8):892-900. doi:10.1002/lt.23442
14. McCool ME, Wahl J, Schlecht I, Apfelbacher C. Evaluating written patient information for eczema in German: comparing the reliability of two instruments, DISCERN and EQIP. *PLoS One*. 2015;10(10):e0139895. doi:10.1371/journal.pone.0139895
15. Malak A, Şahin MF. How Useful are current Chatbots regarding urology patient information? Comparison of the ten most popular Chatbots' responses about female urinary incontinence. *J Med Syst*. 2024;48(1): 102. doi:10.1007/s10916-024-02125-4
16. Goodman RS, Patrinely JR, Stone CA Jr, et al. Accuracy and reliability of Chatbot responses to physician questions. *JAMA Netw Open*. 2023;6(10): e2336483. doi:10.1001/jamanetworkopen.2023.36483
17. Wilson EA, Makoul G, Bojarski EA, et al. Comparative analysis of print and multimedia health materials: a review of the literature. *Patient Educ Couns*. 2012;89(1):7-14. doi:10.1016/j.pec.2012.06.007
18. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. *Ann Intern Med*. 2011;155(2):97-107. doi:10.7326/0003-4819-155-2-201107190-00005
19. Hua HU, Kaakour AH, Rachitskaya A, Srivastava S, Sharma S, Mammo DA. Evaluation and comparison of ophthalmic scientific abstracts and references by current Artificial Intelligence Chatbots. *JAMA Ophthalmol*. 2023;141(9):819-824. doi:10.1001/jamaophthalmol.2023.3119

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1751139
J Health Sci Med. 2025;8(5):884-888

Therapeutic drug monitoring of carbamazepine and valproic acid: interindividual variability in epilepsy patients

Tuğçe Uskur[1], Oya Güven[2]

[1]Department of Medical Pharmacology, Faculty of Medicine, Kırklareli University, Kırklareli, Turkiye
[2]Department of Emergency Medicine, Faculty of Medicine, Kırklareli University, Kırklareli, Turkiye

Cite this article as: Uskur T, Güven O. Therapeutic drug monitoring of carbamazepine and valproic acid: interindividual variability in epilepsy patients. *J Health Sci Med.* 2025;8(5):884-888.

Received: 26.07.2025 ◆ **Accepted**: 26.08.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: Epilepsy is a chronic neurological disorder affecting 1-3% of the global population, characterized by recurrent seizures. Monitoring blood levels of valproic acid and carbamazepine, two commonly used antiepileptic drugs, is essential for effective seizure control. This study aims to analyze the serum concentrations, demographic characteristics and concomitant drug use of patients treated with valproic acid or carbamazepine.

Methods: This retrospective study was conducted at Kırklareli Training and Research Hospital, analyzing the medical records of epilepsy patients treated with valproic acid or carbamazepine. Medical records of 440 epilepsy patients were retrospectively analyzed. Demographic data, concurrent medication use and blood drug levels were recorded, and rates of achieving therapeutic levels were compared.

Results: Of the 440 patients, 378 were treated with valproic acid, and 62 with carbamazepine. In the valproic acid group, 62.2% reached therapeutic levels, with an average concentration of 68.15 µg/ml. In the carbamazepine group, only 27.4% reached therapeutic levels, with an average of 10.07 µg/ml. A statistically significant difference was found between the two groups. A significant association was found between drug interaction counts and non-therapeutic levels in the carbamazepine group ($p<0.05$).

Conclusion: Valproic acid was more likely to be within the therapeutic serum range. Lower therapeutic levels in carbamazepine-treated patients, indicate a need for careful monitoring, dose adjustments and consideration of potential drug interactions. Future studies with larger populations are needed to confirm these findings and support personalized treatment approaches.

Keywords: Epilepsy, carbamazepine, valproic acid, blood-drug level, therapeutic drug monitoring

INTRODUCTION

Epilepsy is one of the most long-standing and prevalent neurological disorders, affecting an estimated 50 million people worldwide. According to the latest definition by the International League Against Epilepsy (ILAE) in 2014, epilepsy is characterized by either two or more unprovoked or reflex seizures occurring more than 24 hours apart or a single unprovoked or reflex seizure with a high probability (at least 60%) of further seizures over the next ten years.[1]

The treatment of epilepsy requires a long-term and continuous approach. Therapeutic options include medical therapy, surgical interventions, and ketogenic diets.[2] Among antiepileptic drugs (AEDs), phenobarbital, phenytoin, ethosuximide, carbamazepine, valproic acid, benzodiazepines, vigabatrin, lamotrigine, topiramate, felbamate, gabapentin, pregabalin, tiagabine, zonisamide, levetiracetam, and oxcarbazepine are used. Carbamazepine and valproic acid are the most used AEDs globally.[3-5] Beyond epilepsy, these two drugs have various other indications. Valproic acid (VPA) is FDA-approved for treating complex partial seizures in pediatric and adult patients aged ten years and older, as well as for monotherapy and adjunctive therapy of various seizure types, including simple and complex absence seizures. It is also used off-label for bipolar disorder-associated manic episodes, migraine prophylaxis, status epilepticus, diabetic peripheral neuropathy, and postherpetic neuralgia.[6] Carbamazepine is used for epilepsy, trigeminal neuralgia, and acute manic and mixed episodes of bipolar disorder; its epilepsy indications include partial seizures with complex symptomatology (psychomotor, temporal lobe), generalized tonic seizures (grand mal), and mixed seizure patterns.[7]

The mechanisms of action of these drugs vary. Carbamazepine blocks voltage-dependent sodium channels, while valproic acid inhibits GABA metabolism, reduces excitatory aspartate and increases inhibitory glycine. It also affects blocking voltage-dependent sodium channels and activating calcium-dependent potassium channels.[6,7] These complex mechanisms necessitate careful monitoring of blood drug levels to ensure therapeutic efficacy and safety.

Corresponding Author: Tuğçe Uskur, tugceuskur@gmail.com

The therapeutic efficacy of these drugs can vary depending on individual differences. Many factors, such as age, sex, liver and kidney function, genetic variations, pregnancy, comorbidities, and concomitant medication use, can affect the pharmacokinetics and serum levels of drugs.[8,9] Therefore, monitoring blood levels is not only essential for enhancing treatment efficacy but also critical for preventing adverse effects and enabling personalized dose adjustments.[8]

Extraordinary situations, such as the COVID-19 pandemic, have further complicated access to healthcare and adherence to treatment, thereby increasing the importance of therapeutic drug monitoring.[10] In this context, understanding the relationship between serum levels of antiepileptic drugs and clinical outcomes plays a key role in treatment management.

This study aims to investigate the serum drug levels of epilepsy patients treated with valproic acid or carbamazepine in the Emergency Department of Kırklareli Training and Research Hospital between January 1 and December 31, 2023, and to examine the relationship of these levels with demographic characteristics and concomitant drug use.

METHODS

Data Source

Ethical approval for this retrospective study was obtained from the Kırklareli University Non-interventional Clinical Researches Ethics Committee (Date: 05.07.2024, Decision No: 04). All procedures were carried out in accordance with the ethical rules and the principles of the Declaration of Helsinki.

In this study, data were retrospectively obtained from the files of 440 patients diagnosed with epilepsy who were treated with valproic acid or carbamazepine and admitted to Kırklareli Training and Research Hospital's Emergency Department within the period from January 1 to December 31, 2023. The dataset included demographic characteristics such as gender, age, concurrent medication use and the patients' drug blood levels. Eligibility for participation in the study was determined based on the following criteria: (1) patients clinically diagnosed with epilepsy, (2) patients treated with valproic acid or carbamazepine with drug-level monitoring, and (3) patients over the age of 18 and (4) patients who attended the emergency department.

Since this was a retrospective study using anonymized patient data, the requirement for informed consent was waived by the ethics committee in accordance with national and institutional ethical guidelines.

No specific consent was obtained from participants as data analysis was conducted anonymously. The need for informed consent was disclaimed by the ethics committee due to the retrospective nature of the study, which involved analysis of deidentified patient data from medical records.

Valproic Acid and Carbamazepine Determination

The analysis of Valproic Acid and Carbamazepine was performed using the immunological method on the ROCHE COBAS C503 device.

The therapeutic blood levels for antiepileptic drugs were considered to be 8-12 µg/ml for carbamazepine[11] and 50-100 µg/ml for valproic acid.[12]

Based on these threshold values, patients were categorized into two groups according to their drug levels: therapeutic and non-therapeutic levels. The non-therapeutic group included patients with subtherapeutic and supratherapeutic drug levels.

Drug Interaction Analysis

Drug interaction analyses was conducted using the Drugs.com database. Drugs used by patients with carbamazepine or valproic acid in the database were investigated, and a list containing drug interactions was obtained; these interactions were classified as minor, moderate, major, or no interaction based on their severity.

Statistical Analysis

The analysis of the data obtained from the study was performed using the SPSS version 24 statistical program. The normality of the distribution of continuous variables was tested using the Shapiro-Wilk test. Descriptive statistics were used to describe contin-uous variables [mean (M), standard deviation (SD), minimum (min), median (med), maximum (max)]. The comparison of two independent continuous variables that did not follow a normal distribution was performed using the Mann-Whitney U test. The relationship between two continuous variables that did not follow a normal distribution was analyzed using Spearman's rho correlation analysis. Associations between categorical variables were evaluated using the Chi-square test or, when appropriate, Fisher's exact test. A p-value less than 0.05 was considered to indicate statistical significance.

RESULTS

Among the 440 epilepsy patients treated with carbamazepine and valproic acid included in our study, 378 were prescribed valproic acid, and 62 were prescribed carbamazepine. Regarding gender distribution, 52% (229 patients) were male, and 48% (211 patients) were female. When analyzing the gender distribution by medication, it was found that of the 62 patients treated with carbamazepine, 61.3% (38 patients) were male, and 38.7% (24 patients) were female. Among the 378 patients treated with valproic acid, 50.5% (191 patients) were male, and 49.5% (187 patients) were female (**Table 1**).

Table 1. Demographic profile of the participants

Profile	Total	n (valproic acid)	n (carbamazepine)
Male, n (%)	229	191 (50.5%)	38 (61.3%)
Female, n (%)	211	187 (49.5%)	24 (38.7%)
Mean±SD (years)	43.1	45.5	42.7

SD: Standard deviation

The average age of the 440 epilepsy patients included in the study is 43.1, with those treated with carbamazepine having an average age of 45.5 and those treated with valproic acid having an average age of 42.7. No statistically significant

difference was found in the age distribution between the medication groups and blood drug levels.

Among the 378 patients treated with valproic acid, 235 had therapeutic drug blood levels, whereas only 17 out of the 62 patients treated with carbamazepine had therapeutic drug blood levels. This finding indicates a statistically significant difference in the rates at which valproic acid and carbamazepine achieve therapeutic levels (Table 2).

Table 2. Therapeutic blood drug levels in patients using valproic acid and carbamazepine

Category	Valproic acid	Carbamazepine	p
Therapeutic level	235 (62.16%)	17 (27.42%)	<0.001
Non-therapeutic level	143 (37.83%)	45 (72.58%)	

There is a statistically significant difference in the rates of therapeutic levels between the drugs valproic acid and carbamazepine (p<0.001, Fisher's Exact test).

When considering a therapeutic blood level range for valproic acid of 50-100 µg/ml, it was found that 235 of the 378 patients treated with valproic acid had therapeutic drug levels. Among these patients, 116 were male and 119 were female, with an average drug blood level of 68.15 µg/ml. Additionally, 132 patients had valproic acid levels below 50 µg/ml, with an average drug blood level of 35.51 µg/ml. Of these patients, 67 were male and 65 were female. Furthermore, 11 patients had drug levels above the therapeutic range, with 8 males and 3 females showing an average drug blood level of 152.9 µg/ml (Table 3).

When considering a therapeutic blood level for carbamazepine as 8-12 ug/ml, it was found that 17 out of 62 patients treated with carbamazepine had therapeutic drug blood levels. Among these patients, 13 were male and 4 were female, with an average blood drug level of 10.07 mmol/L. It was observed that 38 patients had carbamazepine levels below 8 ug/ml, with an average drug level of 5.17 ug/ml. Among the patients whose carbamazepine blood levels were below the therapeutic range, 21 were male and 17 were female. For 7 patients, including 3 women and 4 men, the carbamazepine blood levels were above the therapeutic range, with an average drug level of 15.75 ug/ml (Table 3).

In the carbamazepine group, a statistically significant relationship was found between the number of drug interactions and therapeutic drug levels (p=0.0393, Fisher's exact test). Patients who did not reach therapeutic levels had a higher number of total drug interactions compared to those within the therapeutic range. In contrast, no statistically significant relationship was observed between drug interaction counts and therapeutic levels in the valproic acid group (p=0.1354, Fisher's exact test).

Additionally, drug-drug interactions with the potential to alter the serum levels of valproic acid or carbamazepine were recorded. Among the 62 patients in the carbamazepine group, 16 had potential drug-drug interactions that could influence serum carbamazepine levels, of whom 11 were female. In the valproic acid group, 44 out of 378 patients (35 males and 9 females) had potential interactions that could affect valproic acid levels. The frequency of such interactions was significantly higher in the carbamazepine group compared to the valproic acid group (p=0.0048, Fisher's exact test) (Table 4).

DISCUSSION

Therapeutic drug monitoring (TDM) enhances treatment efficacy and reduces drug side effects by considering individual differences. It helps create personalized treatment plans using modern analytical techniques and pharmacokinetic principles.[11-13] Monitoring blood levels of antiepileptic drugs (AEDs) is a crucial method in clinical evaluation, particularly for ensuring effective seizure control, and it has been standard practice since the introduction of AEDs to the market.[14] Factors such as pregnancy, liver and kidney diseases, drug interactions due to polypharmacy, and ageing can affect the efficacy of AEDs. Therefore, therapeutic drug monitoring can assist in dosage adjustments to achieve a seizure-free quality of life.[15]

Valproic acid and carbamazepine are among the most prescribed AEDs globally due to their effectiveness and safety, and monitoring their blood levels is highly important during treatment. In this retrospective study, serum concentrations of valproic acid or carbamazepine and their relationship with patients' demographic characteristics and drug-drug interaction was analyzed in epilepsy patients.

Table 3. Distributions of therapeutic levels of valproic acid and carbamazepine

Drug	Valproic acid			Carbamazepine		
Gender	Therapeutic level	Below therapeutic level	Above therapeutic level	Therapeutic level	Below therapeutic level	Above therapeutic level
Male	116 (60.7%)	67 (35%)	8 (4.3%)	13 (34.2%)	21 (55.3%)	4 (10.5%)
Female	119 (63.6%)	65 (34.7%)	3 (1.7%)	4 (16.7%)	17 (70.8%)	3 (12.5%)
p-value	0.317			0.319		

There is no statistically significant difference in the distribution of therapeutic carbamazepine levels between men and women (p>0.05, Fisher's exact test).

Table 4. Association between therapeutic drug levels and potential drug interactions in the carbamazepine and valproic acid groups

Antiepileptic drug	Total patients (n)	Patients with interactions potentially affecting drug level (n)	Sex distribution of interaction group (M/F)	p-value (Fisher's exact test)	Significant association between interaction count and therapeutic drug level
Carbamazepine	62	16	5/11	0.0048	Yes (p=0.0393)
Valproic acid	378	44	35/9		No (p=0.1354)

In the carbamazepine group, a higher number of drug interactions was associated with non-therapeutic levels (p=0.0393, Fisher's Exact test). No significant association was observed in the valproic acid group (p=0.1354, Fisher's Exact test). Potential interactions affecting serum levels were more frequent in the carbamazepine group (p=0.0048, Fisher's Exact test). M/F: Male/female

The study found that more patients were prescribed valproic acid than carbamazepine. However, there was no statistically significant difference in gender and age distribution between the two drug groups (p>0.05). According to the study results, there was a statistically significant difference in the rates of achieving therapeutic levels between valproic acid (62.16%) and carbamazepine (27.42%), with valproic acid being more likely to reach therapeutic levels. Contrary to our findings, a study by Sharma et al.[16] indicated that 63% of patients treated with carbamazepine achieved therapeutic drug levels, compared to 45.99% of those treated with valproic acid. Another study published in 2020 found that 75.5% of carbamazepine samples and 54.87% of valproic acid samples were within the therapeutic range.[17,18] Previous reports have suggested valproic acid is poorly controlled.[16,19] Our findings align more closely with previously published studies for valproic acid, in which the majority of patients had drug levels within the therapeutic range. This consistency may be partly attributed to the pharmacokinetic characteristics of valproic acid, which are generally more predictable due to its lower potential for hepatic enzyme induction.[20] In contrast, carbamazepine results showed considerable variability, with a significant proportion of patients presenting with subtherapeutic or supratherapeutic levels. This discrepancy may be due to several factors, demographic characteristics of the patients, including potential patient non-compliance, sample size differences, and particularly the high rate of drug-drug interactions associated with carbamazepine, as observed in our dataset.

One of the key findings of this study was the significant impact of drug-drug interactions on therapeutic drug levels, particularly in the carbamazepine group. As reported in the literature, carbamazepine is known to interact with approximately 760 drugs, including 240 major, 487 moderate, and 33 minor interactions,[21] whereas valproic acid interacts with significantly fewer compounds. In our study, patients using carbamazepine who experienced higher rates of drug interactions were less likely to reach therapeutic levels (p=0.0393), suggesting that polypharmacy and drug interactions substantially affect carbamazepine metabolism and clearance.

Additionally, the frequency of potentially interacting drugs that could alter blood drug levels was significantly higher in the carbamazepine group than in the valproic acid group (p=0.0048). This may further explain the lower proportion of patients with therapeutic carbamazepine levels. Moreover, the observation that most of the patients with potential carbamazepine interactions were female may be related to the higher prevalence of comorbidities and consequently increased polypharmacy in women, which could have contributed to the observed variability in carbamazepine serum levels (both below and above the therapeutic range) in this subgroup. In contrast, no statistically significant association was found between drug interaction count and valproic acid serum levels (p=0.1354), indicating that valproic acid may have a more stable pharmacokinetic profile even in the presence of polypharmacy.[22]

These findings underscore the importance of routine therapeutic drug monitoring, especially in patients prescribed carbamazepine. Regular monitoring of serum carbamazepine levels allows timely dose adjustments, ensuring patients reach and maintain therapeutic concentrations. Dose titration should be individualized based on age, weight, renal and hepatic function, and other patient-specific factors. Clinicians should be vigilant in evaluating patients' concurrent medications to anticipate possible pharmacokinetic interactions. Patient education and adherence support are crucial, emphasizing the importance of taking the medication consistently and recognizing potential side effects. The choice of AED, therefore, should take into account not only the patient's seizure type and demographics but also their likelihood of being exposed to interacting drugs. In patients with polypharmacy or higher interaction risk, considering alternative medications or adjusting concomitant therapies can help achieve optimal therapeutic levels. Personalized treatment approaches that incorporate drug interaction risk profiles may lead to better seizure control and fewer adverse outcomes.[23]

Limitations

Our study has several limitations due to its retrospective design. It is possible that not all patients for whom drug levels were requested could be reached. Additionally, vital signs monitoring, physical examination findings, clinical outcomes of patients with non-therapeutic drug levels, seizure frequency, and adverse events were not recorded; therefore, the relationship between therapeutic levels and clinical improvement or deterioration could not be evaluated, and the findings cannot be directly translated into prescribing recommendations. Moreover, information regarding patients' medication use includes only drugs prescribed within the hospital system where the study was conducted. Medications obtained from other hospitals, healthcare institutions, or over-the-counter drugs without prescription were not captured in our data. The choice of the Drugs.com database for drug-drug interaction assessment was based on accessibility and practicality; however, compared to alternative platforms such as Micromedex or Lexicomp, it may have limitations in comprehensiveness and classification of interaction severity. The single-center design may limit the generalizability of the findings, and regional prescribing patterns could have influenced both drug selection and attainment of therapeutic levels. Furthermore, "confounding by indication" should be considered, as patients prescribed carbamazepine may have had more refractory epilepsy types, which could influence serum level outcomes. Despite these limitations, since the data were obtained from the largest and sole central hospital in the city, the results may be representative and provide valuable insights for future studies.

CONCLUSION

This study examined the rates at which valproic acid and carbamazepine, used in treating epilepsy, reach therapeutic serum levels, and how these rates relate to patients' demographic characteristics. We found valproic acid is more likely to reach therapeutic levels than carbamazepine.

However, we also observed that demographic differences among patients treated with carbamazepine and potential drug interactions may be associated with lower serum levels. Therefore, careful monitoring and dosage adjustments are essential, particularly for women who are on carbamazepine therapy. Future studies should aim to validate these findings in larger populations and help develop individualized treatment approaches for antiepileptic drugs.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was carried out with the permission of the Kırklareli University Non-interventional Clinical Researches Ethics Committee (Date: 05.07.2024, Decision No: 04).

Informed Consent

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

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

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

Author Contributions

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

REFERENCES

1. Banerjee PN, Filippi D, Allen Hauser W. The descriptive epidemiology of epilepsy-a review. *Epilepsy Res.* 2009;85(1):31-45. doi:10.1016/j.eplepsyres.2009.03.003
2. Turkish Neurological Society. In: Dervent A, Ayta S, Çokar Ö, Uludüz D. Approach Guide to Neurological Diseases in Children and Adolescents. 2015. https://noroloji.org.tr/TNDData/Uploads/files/%C3%87ED-2015(1).pdf
3. Hanaya R, Arita K. The new antiepileptic drugs: their neuropharmacology and clinical indications. *Neurol Med Chir (Tokyo).* 2016;56(5):205-20. doi:10.2176/nmc.ra.2015-0344
4. Iorio ML, Moretti U, Colcera S, et al. Use and safety profile of antiepileptic drugs in Italy. *Eur J Clin Pharmacol.* 2007;63(4):409-415. doi:10.1007/s00228-006-0236-6
5. Elendu C, Jeswani BM, Madekwe CC, et al. Clinical and electroencephalographic correlates of carbamazepine-associated hiccups in epileptic patients. *Ann Med Surg (Lond).* 2024;86(7):4015-4034. doi:10.1097/MS9.0000000000002159
6. Rahman M, Awosika AO, Nguyen H. Valproic Acid. [Updated 2023 Aug 17]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK559112/
7. Maan JS, Duong TvH, Saadabadi A. Carbamazepine. [Updated 2023 Jul 10]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK482455/
8. St Louis EK. Monitoring antiepileptic drugs: a level-headed approach. *Curr Neuropharmacol.* 2009;7(2):115-119. doi:10.2174/157015909788848938
9. Akdağ G, Algın Dİ, Erdinç OO. Epileps. *Osmangazi J Med.* 2016;38(Special Issue 1):35-41. doi:10.20515/otd.88853
10. Giza E, Lefkopoulou M, Dimitrakopoulos D, Karachristianou S, Liasidis C. The impact of the COVID-19 pandemic on a Greek cohort of patients with epilepsy. *Hippokratia.* 2021;25(4):145-150.
11. Shared Health – SBGH Lab Manual. Carbamazepine. Available from: https://apps.sbgh.mb.ca/labmanual/test/view?seedId=1397. Accessed August 2025.
12. Tseng YJ, Huang SY, Kuo CH, Wang CY, Wang KC, Wu CC. Safety range of free valproic acid serum concentration in adult patients. *PLoS One.* 2020;15(9):e0238201. doi:10.1371/journal.pone.0238201
13. Zeng L, Yi Q, Huang L, et al. The guideline for therapeutic drug monitoring guidelines development. *J Evid Based Med.* 2022;15(3):272-283. doi: 10.1111/jebm.12486
14. Liu Y, Li J, Xiao S, et al. Revolutionizing precision medicine: exploring wearable sensors for therapeutic drug monitoring and personalized therapy. *Biosensors (Basel).* 2023;13(7):726. doi:10.3390/bios13070726
15. Irving PM, Gecse KB. Optimizing therapies using therapeutic drug monitoring: current strategies and future perspectives. *Gastroenterology.* 2022;162(5):1512-1524. doi:10.1053/j.gastro.2022.02.014
16. Sharma S, Joshi S, Mukherji S, Bala K, Tripathi CB. Therapeutic drug monitoring: appropriateness and clinical utility in neuropsychiatry practice. *Am J Ther.* 2009;16(1):11-16. doi:10.1097/MJT.0b013e31817fd85f
17. French JA, Perucca E. time to start calling things by their own names? The case for antiseizure medicines. *Epilepsy Curr.* 2020;20(2):69-72. doi: 10.1177/1535759720905516
18. Fratoni AJ, Colmerauer JL, Linder KE, Nicolau DP, Kuti JL. A retrospective case series of concomitant carbapenem and valproic acid use: are best practice advisories working? *J Pharm Pract.* 2023;36(3):537-541. doi:10.1177/08971900211063301
19. Al-Balawi RS, Alshehri MA, Alatawi AS, Al Shehri AM, Alshehry MA, Al-Gayyar MMH. Measuring the appropriateness of carbamazepine and valproic acid prescribing and utilization using a newly implemented online system in the Tabuk Region of Saudi Arabia. *Saudi Pharm J.* 2020;28(7):844-849. doi:10.1016/j.jsps.2020.06.008
20. Dalaklioglu S. Evaluating appropriateness of digoxin, carbamazepine, valproic acid, and phenytoin usage by therapeutic drug monitoring. *Clin Lab.* 2013;59(3-4):325-31. doi:10.7754/clin.lab.2012.120425
21. Bowden CL. Valproate. *Bipolar Disord.* 2003;5(3):189-202. doi:10.1034/j.1399-5618.2003.00031.x
22. Marvanova M. Pharmacokinetic characteristics of antiepileptic drugs (AEDs). *Ment Health Clin.* 2016;6(1):8-20. doi:10.9740/mhc.2015.01.008
23. https://www.drugs.com/drug_interactions.html

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1760149

J Health Sci Med. 2025;8(5):889-895

Prevalence and determinants of orthodontic malocclusions in children: a multifactorial analysis

Yasin Akbulut[1], Semih Ercan Akgün[2]

[1]Department of Orthodontics, Faculty of Dentistry, Kocaeli Health and Technology University, Kocaeli, Turkiye
[2]Department of Dentistry Services, Vocational School of Health Services, Ondokuz Mayıs University, Samsun, Turkiye

Cite this article as: Akbulut Y, Akgün SE. Prevalence and determinants of orthodontic malocclusions in children: a multifactorial analysis. *J Health Sci Med.* 2025;8(5):889-895.

Received: 07.08.2025 ◆ **Accepted**: 26.08.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: To determine the prevalence of orthodontic malocclusions in children and adolescents and to evaluate, through a multifactorial approach, the associations between these malocclusions and mesiodistal dimension loss, presence of dental caries, missing teeth, oral hygiene habits, and dietary behaviors.

Methods: This descriptive and cross-sectional study was conducted on 1092 children aged 4-15 years. Participants were stratified into age groups, and clinical examinations were performed to assess malocclusion type, number of missing and decayed teeth, oral hygiene indices (Gingival Index and Plaque Index), as well as individual oral hygiene and dietary habits. Normality was assessed using the Shapiro-Wilk test. Depending on the data type and distribution, Chi-square tests with Bonferroni correction, two-proportion Z tests, ANOVA with Tukey's HSD, Kruskal-Wallis with Bonferroni-adjusted Dunn, and Mann-Whitney U tests were performed. Statistical significance was set at $p<0.05$.

Results: The prevalence of orthodontic malocclusions increased significantly with age, being most common in the 10-12 and 13-15 age groups ($p<0.001$). All malocclusion types were significantly more prevalent among individuals with mesiodistal dimension loss, with class IV malocclusion observed exclusively in this group ($p<0.001$). Children with class II malocclusion had the highest mean number of missing teeth, which was significantly greater than those with no malocclusion ($p<0.001$). Similarly, the mean number of decayed teeth was significantly higher in class II and class III groups compared to those without malocclusion ($p<0.001$ and $p=0.042$, respectively). Oral hygiene and dietary habits were also significantly associated with malocclusion types. Lower tooth brushing frequency, lack of interdental cleaning, and higher consumption of acidic and sugary foods were more common among those with malocclusion ($p<0.001$). Additionally, poor oral hygiene and diet were strongly correlated with increased rates of caries ($p<0.001$) and missing teeth ($p<0.05$).

Conclusion: Orthodontic malocclusions increase progressively with age and are significantly associated with dental caries, tooth loss, inadequate oral hygiene, and unhealthy dietary habits. The implementation of multidisciplinary oral health strategies at an early age may be effective in reducing the incidence of both orthodontic anomalies and dental caries.

Keywords: Malocclusion, dental caries, tooth loss, oral hygiene, dietary habits, children

INTRODUCTION

Orthodontic malocclusions represent a common oral health issue in children and adolescents, with potential implications within functional, aesthetic, and psychosocial domains.[1] Globally, malocclusion is considered the third most prevalent oral health problem following dental caries and periodontal diseases.[2] Systematic review and meta-analyses conducted among pediatric and adolescent populations report the prevalence of class I malocclusion to be approximately 56%, class II around 31%, and class III about 11%. Large-scale studies conducted in Turkiye have yielded comparable findings, indicating a high prevalence of crowding (41%), increased overjet (34%), and crossbite (11%) anomalies.[3]

Dental caries is closely associated with different types of malocclusion. The Decayed, Missing, and Filled Teeth (DMFT) Index is a widely used epidemiological measure for quantifying dental caries experience. Evidence suggests that children aged 6 to 12 years with a high incidence of caries, as indicated by higher DMFT scores, exhibit a greater prevalence of severe malocclusions and are at increased risk of developing such conditions.[4] Similarly, individuals with high DMFT scores have been identified as belonging to a higher risk group for the development of malocclusions.[5]

Tooth loss during childhood is a critical condition that can significantly affect both masticatory function and craniofacial growth patterns. In particular, early tooth loss may adversely affect the eruption of permanent teeth, reduce arch length, and disrupt normal occlusion. Significant associations have been found between mesiodistal dimension loss resulting

Corresponding Author: Semih Ercan Akgün, semihercanakgun@gmail.com

from congenital tooth agenesis or early primary tooth loss and dental malocclusions.[6,7] This condition has been specifically linked to an increased prevalence of crowding and overjet anomalies, especially in individuals with early loss of primary molars or missing deciduous teeth.[8]

Oral hygiene plays a vital role not only in maintaining periodontal health but also in the development of orthodontic problems. The literature reports a significant association between oral hygiene status and malocclusion. Higher plaque and gingival index scores have been correlated with increased severity of malocclusion.[1,9] Moreover, individual hygiene behaviors such as the frequency of tooth brushing, use of dental floss or interdental brushes, and mouthwash have been identified as key determinants in malocclusion development.[5] Consistent oral care practices may help prevent orthodontic issues associated with caries and plaque accumulation.[5,10]

Dietary habits are among the key behavioral determinants in maintaining oral health. Frequent consumption of acidic beverages and sugary snacks significantly increases the risk of both dental caries and gingivitis by promoting dental plaque accumulation. This, in turn, has been indirectly linked to the development and progression of malocclusion.[11,12]

The aim of this study is to determine the distribution of orthodontic malocclusions across different age groups and to evaluate, from a broad perspective, the relationships between these malocclusions and mesiodistal dimension loss, presence of caries and missing teeth, oral hygiene, and dietary habits. The findings are intended to provide scientific insights that support the development of multidisciplinary oral health strategies.

METHODS

Ethical Approval

This descriptive and cross-sectional study was conducted at the Faculty of Dentistry, Fırat University. The study was approved by the Fırat University Non-interventional Researches Ethics Committee (Date: 10.01.2019, Decision No: 2019/01-24). Written and verbal information about the study was provided to the parents of all participants, and written informed consent forms were obtained based on voluntary participation. Throughout the study, the ethical principles outlined in the World Medical Association (WMA) Declaration of Helsinki for medical research involving human subjects were strictly followed.

Participants

A total of 1092 children aged between 4 and 15 years participated in the study. The participants were randomly selected through clinical examination among pediatric patients who presented to the Department of Orthodontics, Faculty of Dentistry, Fırat University.

Based on developmental characteristics, the participants were divided into four age groups:

- **Group I:** 4-6 years-130 children
- **Group II:** 7-9 years-448 children
- **Group III:** 10-12 years-376 children
- **Group IV:** 13-15 years-138 children

This age-based grouping enabled a comparative analysis of oral and dental health outcomes by considering the children's physiological and behavioral developmental levels.

Data Collection Process

The data obtained through detailed anamnesis and clinical examination included demographic information (age, gender), oral hygiene habits (frequency of tooth brushing, use of dental floss or interdental brushes, use of mouthwash), and dietary behaviors (consumption of acidic beverages and sweet snacks such as chocolate and candy).

Clinical examinations were conducted by the specialist researcher Y.A., on participants who had provided informed consent. The examinations were conducted under standardized lighting conditions using a dental mirror and probe. All procedures were performed in a consistent clinical environment. During the clinical assessments, the presence of missing teeth, number of decayed teeth, status of mesiodistal dimension loss, and type of orthodontic malocclusion were recorded.

Orthodontic Classification

Orthodontic assessment was performed according to the Angle classification. Class I malocclusion was defined as the mesiobuccal cusp of the maxillary first molar being aligned with the buccal groove of the mandibular first molar. Class II malocclusion was characterized by the mesiobuccal cusp of the maxillary first molar being positioned anterior to the buccal groove of the mandibular first molar, whereas class III malocclusion was diagnosed when it was located posterior to this reference point. However, although not included in Angle's original anomaly classification, a class IV category was subsequently introduced to facilitate the categorization of cases with an asymmetric relationship, specifically class II molar relationship on one side and class III on the other, frequently accompanied by midline deviation.[13]

Oral Hygiene Assessment

Additionally, oral hygiene was evaluated using the Gingival Index (GI) and Plaque Index (PI). The GI and PI were recorded according to the criteria of Loe and Silness[14], assessing four surfaces (mesial, distal, buccal, lingual) of all index teeth on a 0-3 scale, where higher scores indicated greater GI or PI.

All index measurements were performed by the calibrated researcher Y.A. To ensure measurement reliability, repeated assessments were conducted on 10% of the sample, yielding a kappa coefficient above 0.80, which indicates strong consistency. All clinical assessments were carried out by the same researcher, thereby minimizing the impact of inter-examiner variability.

Statistical Analysis

Data analyses were performed using IBM SPSS Statistics for Windows, Version 25.0 (IBM Corp., Armonk, NY, USA). Normality of continuous and ordinal variables was assessed using the Shapiro-Wilk test. Normally distributed data were expressed as mean±standard deviation and compared using one-way ANOVA with Tukey's HSD post-hoc test when significant; non-normally distributed or ordinal data were

presented as median (minimum-maximum) and analyzed using the Kruskal-Wallis test with Bonferroni-adjusted Dunn's post-hoc test. Associations between categorical variables were evaluated using the Chi-square test with Bonferroni-adjusted residuals for post-hoc comparisons, and proportional differences between two independent groups were assessed with the two-proportion Z test. For two-group comparisons of non-normally distributed or ordinal variables, the Mann-Whitney U test was applied. Statistical significance was set at $p<0.05$.

RESULTS

Descriptive statistical data on variables such as the number of missing and decayed teeth by age, gingival and plaque indices, oral hygiene habits, and dietary behaviors of the individuals included in the study are presented in Table 1.

The distribution of orthodontic malocclusion types showed a statistically significant difference across all age groups ($p<0.001$) (Table 2). In group I, although the majority of individuals exhibited no malocclusion, class I malocclusion was the most common type among those affected. Class II and class III malocclusions were observed at lower frequencies than class I malocclusion. Groups II and III demonstrated a notable increase in the prevalence of all malocclusion types. In group IV, while malocclusion types remained present, their prevalence decreased compared to the younger age groups.

Statistically significant differences were found in the distribution of orthodontic malocclusion types between individuals with and without mesiodistal (MD) dimension loss ($p<0.05$) (Table 3). The prevalence of class I, II, III, and IV malocclusions was significantly higher among individuals exhibiting MD dimension loss ($p<0.05$). Notably, class IV malocclusion was observed exclusively in individuals with MD loss, and this association was statistically significant ($z=9.660$, $p<0.001$). In contrast, the prevalence of normal occlusion was significantly higher in individuals without MD dimension loss ($z=-11.682$, $p<0.001$).

Table 3. Distribution of orthodontic malocclusion types by presence of mesiodistal loss

	Mesiodistal loss			
Malocclusion type	Positive	Negative	Test statistics	p-value
No malocclusion	140 (25.8%)	338 (60.8%)	-11.682	<0.001
Class I	192 (35.4%)	140 (25.2%)	3.695	<0.001
Class II	80 (14.8%)	48 (8.6%)	3.163	0.002
Class III	46 (8.5%)	30 (5.4%)	2.018	0.044
Class IV	84 (15.5%)	0 (0%)	9.660	<0.001

Two-proportion Z tests; n (%); $p<0.05$ significance level

A statistically significant difference was observed in the mean number of missing teeth across different types of orthodontic

Table 1. Descriptive statistics of orthodontic, behavioral, and dietary variables in the study population

	Gruops (age)			
	Gruop I (4-6)	Gruop II (7-9)	Gruop III (10-12)	Gruop IV (13-15)
Age	5.29±0.78	8.1±0.75	10.84±0.85	13.45±0.60
Gender				
Male	68 (52.3%)	210 (46.9%)	182 (48.4%)	82 (59.4%)
Female	62 (47.7%)	238 (53.1%)	194 (51.6%)	56 (40.6%)
Missing teeth	1.03	0.94	1.22	0.78
Decayed teeth	3.60	4.59	3.31	3.36
Gingival Index	1.14	1.21	1.23	1.14
Plaque Index	1.34	1.38	1.38	1.22
Brushing frequency	0.92	1.03	1.04	1.06
Floss/interdental brush use	0.15	0.27	0.15	0.26
Mouthwash use	0.15	0.34	0.22	0.20
Acidic beverage consumption	2.68	2.87	2.96	3.26
Snack consumption	2.57	3.08	2.98	3.26

Mean±standard deviation

Table 2. Pairwise comparisons of orthodontic malocclusions types within age groups

	Malocclusion groups						
Age groups	No Malocclusion	Class I	Class II	Class III	Class IV	Test statistics	p-value
Group I	86 (66.2)^{c}	32 (24.6)^{a}	6 (4.6)^{b}	0 (0)	6 (4.6)^{b}	131.294	<0.001
Group II	202 (42.5)^{c}	126 (26.5)^{a}	52 (10.9)^{b}	32 (6.7)^{b}	36 (7.5)^{b}	240.661	<0.001
Group III	132 (35.2)^{a}	126 (33.5)^{a}	56 (14.8)^{b}	32 (8.5)^{bc}	30 (8.0)^{c}	134.113	<0.001
Group IV	54 (39.2)^{a}	46 (33.3)^{a}	14 (10.1)^{b}	12 (8.7)^{b}	12 (8.7)^{b}	61.862	<0.001

Chi-square test; n (%); $p<0.05$ significance level
There was a statistically significant difference between the groups shown with different lowercase superscripts in the same row ($p<0.001$).

malocclusion (p<0.001) (Table 4). Individuals with class II malocclusion exhibited the highest average number of missing teeth, followed by those with class III malocclusion. Conversely, individuals with class IV malocclusion had a notably lower average number of missing teeth. Additionally, a statistically significant association was identified between the number of missing teeth and the types of orthodontic malocclusion (X^2=71.951, p<0.001).

A statistically significant difference was observed in the mean number of decayed teeth among the orthodontic malocclusion groups (p<0.001) (Table 5). Individuals with class II and class III malocclusions had significantly higher numbers of decayed teeth. Furthermore, a significant association was identified between categories of decayed teeth and types of malocclusion (X^2=46.052, p<0.001).

As a result of pairwise comparisons, individuals with class II malocclusion were found to have a significantly higher mean number of decayed teeth compared to those without malocclusion (p<0.001). Similarly, individuals with class III malocclusion had significantly more decayed teeth than those without malocclusion (p=0.042). No statistically significant differences were observed in the comparisons among the other malocclusion groups (p>0.05).

No statistically significant differences were observed in GI and PI between individuals with and without orthodontic malocclusion (p>0.05) (Table 6).

Table 6. Comparison of gingival and Plaque Index scores between individuals with and without orthodontic malocclusions

	Orthodontic malocclusion			
Variable	Negative	Positive	Test statistics	p-value
Gingival Index	1.00 (0.5-2.0)	1.00 (0.5-2.5)	143.852	0.324
Plaque Index	1.00 (0.5-2.5)	1.00 (0.5-3.0)	139.900	0.070

Mann-Whitney U test; median (minimum-maximum); p<0.05 significance level

Statistically significant differences were observed among the groups in all behavioral variables, including tooth brushing frequency, use of dental floss/interdental brushes, mouthwash use, consumption of acidic beverages, and snack consumption (p<0.001) (Table 7). Brushing frequency yielded the highest test statistic (H=70.332), suggesting that individuals without orthodontic malocclusion may engage in more regular oral hygiene practices. Likewise, significant differences were observed among the groups in terms of dental floss use (H=51.761) and mouthwash use (H=53.223). Furthermore, dietary behaviors such as acidic beverage (H=25.296) and snack consumption (H=25.307) varied significantly based on the type of orthodontic malocclusion.

Table 4. Distribution of missing teeth count by orthodontic malocclusion types

Malocclusion group	n (%)	Missing tooth count	Test statistics[§]	p-value[§]	Test statistics[α]	p-value[α]
No malocclusion	478 (43.55%)	0.94±1.39[a]				
Class I	332 (30.2%)	0.93±1.34[a]				
Class II	128 (11.7%)	1.67±1.72[b]	8.290	<0.001	71.951	<0.001
Class III	76 (6.9%)	1.24±2.08[ab]				
Class IV	84 (7.7%)	0.76±0.90[a]				

[§]: ANOVA, [α]: Chi-square test; Mean ± Standart Deviation; p<0.05 significance level
There was a statistically significant difference between the groups shown with different lowercase superscripts in the same column (p<0.001).

Table 5. Comparison of mean number of decayed teeth among orthodontic malocclusion types

Malocclusion group	n (%)	Tooth decay count	Test statistics[§]	p-value[§]	Test statistics[α]	p-value[α]
No malocclusion	478 (43.55%)	3.56±2.27[a]				
Class I	332 (30.2%)	4.02±2.86[a]				
Class II	128 (11.7%)	4.42±2.18[ab]	4.843	<0.001	46.052	<0.001
Class III	76 (6.9%)	4.39±1.77[ab]				
Class IV	84 (7.7%)	3.79±2.02[a]				

[§]: ANOVA, [α]: Chi-square test; mean±standard deviation; p<0.05 significance level
There was a statistically significant difference between the groups shown with different lowercase superscripts in the same column (p<0.001).

Table 7. Association between types of orthodontic malocclusions and oral hygiene & dietary habits

Variable	No malocclusion	Class I	Class II	Class III	Class IV	Test statistics	p-value
Brushing frequency	2 (0-5)[ab]	2 (0-5)[a]	1 (0-5)[b]	2 (0-5)[bc]	4 (0-5)[bc]	70.332	<0.001
Floss/interdental brush use	0 (0-4)[b]	0 (0-5)[a]	0 (0-4)[a]	0 (0-4)[ab]	0 (0-4)[a]	51.761	<0.001
Mouthwash use	0 (0-4)[b]	0 (0-4)[a]	0 (0-4)[a]	0 (0-4)[ab]	0 (0-5)[a]	53.223	<0.001
Acidic beverage consumption	4 (0-6)[a]	4 (0-5)[a]	2 (0-5)[b]	3 (0-5)[ab]	3 (0-5)[ab]	25.296	<0.001
Snack consumption	4 (0-5)[ac]	4 (0-5)[ab]	2 (0-5)[b]	3 (0-5)[bc]	4 (0-5)[a]	25.307	<0.001

Kruskal-Wallis test; median (minimum-maximum); p<0.05 significance level
There was a statistically significant difference between the groups shown with different lowercase superscripts in the same row (p<0.001).

Statistically significant differences were found in the relationship between the number of decayed teeth, oral hygiene, and dietary habits (Table 8). Brushing frequency (H=50.692, p<0.001), use of dental floss/interdental brushes (H=20.877, p<0.001), mouthwash use (H=10.429, p=0.034), consumption of acidic beverages (H=57.482, p<0.001), and snack consumption (H=31.086, p<0.001) were all significantly associated with the number of decayed teeth.

Table 8. Association between dental caries and oral hygiene & dietary habits habits

	Dental caries			
Variable	Present	Absent	Test statistics	p-value
Brushing frequency	2 (0-5)	2 (0-4)	50.692	<0.001
Floss/interdental brush use	0 (0-4)	0 (0-5)	20.877	<0.001
Mouthwash use	0 (0-5)	0 (0-4)	10.429	0.034
Acidic beverage consumption	4 (0-6)	4 (0-5)	57.482	<0.001
Snack consumption	4 (0-5)	4 (0-5)	31.086	<0.001

Kruskal-Wallis test; median (minimum-maksimum); p<0.05 significance level

Statistically significant differences were found between the number of missing teeth and individuals' oral hygiene and dietary habits (p<0.05) (Table 9). Brushing frequency significantly affected the number of missing teeth (H=13.961, p=0.016). Individuals with regular brushing habits had a significantly lower number of missing teeth. Similarly, the number of missing teeth was significantly lower among individuals used dental floss or interdental brushes (H=41.616, p<0.001). The frequency of mouthwash use was also associated with the number of missing teeth, showing a statistically significant difference (H=24.357, p<0.001). Regarding dietary habits, consumption of acidic beverages was found to significantly increase the number of missing teeth (H=70.633, p<0.001). Likewise, frequent snack consumption was significantly associated with tooth loss (H=44.136, p<0.001).

Table 9. Association between missing teeth and oral hygiene & dietary habits

	Missing teeth			
Variable	Present	Absent	Test statistics	p-value
Brushing frequency	2 (0-5)	2 (0-5)	13.961	0.016
Floss/interdental brush use	0 (0-5)	0 (0-4)	41.616	<0.001
Mouthwash use	0 (0-5)	0 (0-4)	24.357	<0.001
Acidic beverage consumption	3 (0-5)	4 (0-6)	70.633	<0.001
Snack consumption	3 (0-5)	4 (0-5)	44.136	<0.001

Kruskal-Wallis test; median (minimum-maksimum); p<0.05 significance level

DISCUSSION

In this study, the prevalence of orthodontic anomalies in children and adolescents, along with their relationship to essential oral health parameters such as caries, plaque, and gingival indices, was evaluated. Furthermore, by analyzing the distribution of orthodontic anomalies across different age groups, the study aimed to emphasize the importance of early diagnosis and intervention based on their impact on oral health. The Angle classification traditionally includes class I, II, and III malocclusions. However, although not part of Angle's original anomaly classification, a class IV category has subsequently been introduced, as suggested by Ulgen,[13] to facilitate the identification of asymmetric cases characterized by a class II molar relationship on one side and class III on the other, often accompanied by midline deviation. In this respect, the present study aims to provide a novel contribution to the classification system.

The findings of this study indicate that the low prevalence of orthodontic anomalies in the 4-6 age group and their marked increase in the 10-12 and 13-15 age groups, along with the significant associations between these anomalies and oral health parameters such as caries, PI, and GI, are strongly consistent with findings reported in the literature. In a study by Jafari et al.,[9] it was scientifically demonstrated that as Index of Orthodontic Treatment Need scores increased, DMFT, GI, and OHI-S values also increased, indicating that more severe malocclusion is associated with poorer oral health parameters. Additionally, in a study conducted in China among the 11-14 age group, Wang et al.[15] reported a malocclusion prevalence of 72% and identified factors such as malocclusion, poor oral hygiene, and frequent consumption of sugary drinks as risk factors for caries development. Furthermore, an epidemiological study conducted in Spain emphasized that in 12 and 15-year-old adolescents, dental caries and plaque accumulation were strongly associated with malocclusion, and that the prevalence of malocclusion increased with age.[16]

In the current study, the prevalence of caries was found to be significantly higher among individuals with orthodontic anomalies. Wang et al.[15] reported a caries rate of 68.8% among individuals aged 11-14 and associated this finding with oral hygiene, diet, and genetic predisposition. Similarly, data from this study indicate a high caries rate in the same age group, with orthodontic anomalies appearing to contribute to this outcome. The difficulty in achieving proper interproximal contact and the reduced effectiveness of brushing due to malocclusion may explain this association.

According to the results of this study, class II malocclusions were found to have a particularly strong association with caries development. Similarly, the study by Bernhardt et al.[17] reported that class II malocclusions were more frequently associated with poor oral hygiene and an increased risk of caries. Consistent with these findings, studies by Alrashed et al.[18] and Bakhurji et al.[19] also indicated that malocclusions, particularly in the lower molar region, were directly linked to plaque accumulation and caries. In individuals with class II malocclusion, anterior positioning of the maxillary teeth may impede proper mouth closure, promoting mouth breathing and plaque accumulation in the posterior regions. This condition is considered a key factor contributing to the heightened risk of caries.

Oral hygiene is a crucial factor not only in maintaining periodontal health but also in preserving the functional and morphological integrity of oral tissues. In the present study, PI and GI values were found to be significantly higher in the group with orthodontic anomalies. López-Gómez et al.[20] reported that early loss of primary teeth is associated with poor oral hygiene and may predispose individuals to

malocclusion by affecting jaw development. This finding is largely in agreement with the results of the current study.

In the study conducted by Mahboobi et al.,[21] high carbohydrate consumption was reported to negatively affect oral hygiene, and this condition was directly associated with the development of caries. In contrast, although the present study did not directly assess individuals' dietary habits, the poor oral hygiene observed in the presence of malocclusion and elevated caries scores suggests that a similar pathophysiological mechanism may be involved. The differences between the findings are largely attributable to variations in methodology. Mahboobi et al.[21] assessed carbohydrate consumption using survey-based direct measurements and statistically analyzed the relationship between diet and oral health with a dataset covering a broad age range. In contrast, the current study relied solely on clinical observations and objective intraoral indices. Additionally, factors such as sample size, sociodemographic differences, and environmental influences (e.g., oral hygiene education, parental supervision, and access to healthcare services) may also have contributed to the differing results.

According to the American Academy of Pediatric Dentistry, frequent consumption of sugary drinks and acidic foods increases the risk of dental caries, which may lead to premature tooth loss and indirectly contribute to the development of orthodontic anomalies.[22] Similarly, in this study, individuals who frequently consumed such beverages were found to have significantly higher rates of both caries prevalence and orthodontic anomalies.

Orthodontic anomalies that increase with age underscore the importance of initiating treatment during adolescence. In a study conducted by Patel et al.,[23] among 1290 students aged 13-15 years, 43.9% were found to have a moderate and 22.4% a severe need for orthodontic treatment, indicating that approximately 66% of the individuals were potential treatment candidates. In another study conducted by Kalantari et al.,[24] it was reported that the prevalence of malocclusion increased with age and that the need for orthodontic treatment became more pronounced, especially in individuals aged 13-15 years. These findings are consistent with the current study, which also identified a high proportion of individuals aged 13-15 as candidates for orthodontic treatment. Accelerated craniofacial growth during adolescence, combined with inadequate oral hygiene, contributes to the increased manifestation of malocclusions in this group.

The association between tooth loss and orthodontic anomalies is particularly noteworthy. Gandhi et al.[25] reported that premature loss of primary teeth, particularly in the posterior regions, may predispose individuals to jaw constriction and the development of malocclusion. Similarly, the current study found a higher prevalence of orthodontic anomalies among individuals with missing teeth.

Limitations

This study has several limitations. The cross-sectional design prevents the establishment of causal relationships. Potentially influential variables, such as participants' socioeconomic status and parental education levels, were not evaluated. All clinical evaluations were based solely on visual examination, and objective diagnostic tools such as radiographic imaging or digital model analysis were not utilized. Furthermore, the sample consisted exclusively of individuals from a specific geographic region, limiting the applicability of the findings to other populations. Future studies are recommended to address these limitations by employing multi-centered, longitudinal designs with more comprehensive methodologies.

CONCLUSION

This study revealed that orthodontic malocclusions in children and adolescents increase significantly with age and are strongly associated with various oral health parameters, including caries prevalence, tooth loss, oral hygiene status, and dietary habits. Specifically, class II and class III malocclusions were found to be significantly associated with dental caries and tooth loss. Poor oral hygiene practices and frequent consumption of acidic foods were identified as contributing factors that exacerbate these adverse outcomes. The findings suggest that preventive dentistry strategies implemented through multidisciplinary approaches beginning in early childhood may play an effective role in reducing both the incidence of orthodontic anomalies and dental caries.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was carried out with the permission of the Fırat University Non-interventional Researches Ethics Committee (Date: 10.01.2019, Decision No: 2019/01-24).

Informed Consent

Written and verbal information about the study was provided to the parents of all participants, and written informed consent forms were obtained based on voluntary participation.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

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

Author Contributions

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

REFERENCES

1. Salim NA, Alamoush RA, Al-Abdallah MM, Al-Asmar AA, Satterthwaite JD. Relationship between dental caries, oral hygiene, and malocclusion among Syrian refugee children and adolescents: a cross-sectional study. *BMC Oral Health*. 2021;21(1):629. doi:10.1186/s12903-021-01993-3
2. De Ridder L, Aleksieva A, Willems G, Declerck D, Cadenas de Llano-Pérula M. Prevalence of orthodontic malocclusions in healthy children and adolescents: a systematic review. *Int J Environ Res Public Health*. 2022;19(12):7446. doi:10.3390/ijerph19127446

3. Londono J, Ghasemi S, Moghaddasi N, et al. Prevalence of malocclusion in Turkish children and adolescents: a systematic review and meta-analysis. *Clin Exp Dent Res*. 2023;9(4):689-700. doi:10.1002/cre2.771
4. Adanero A, Baquero L, Berasategui M, et al. Oral health status of 6- to 12-year-old children in Madrid, Spain: a cross-sectional study. *Heliyon*. 2022;8(6):e09557. doi:10.1016/j.heliyon.2022.e09557
5. Singh A, Purohit B. Is malocclusion associated with dental caries among children and adolescents in the permanent dentition? A systematic review. *Community Dent Health*. 2021;38(3):172-177. doi:10.1922/CDH_00340Singh06
6. Shakti P, Singh A, Purohit BM, Purohit A, Taneja S. Effect of premature loss of primary teeth on prevalence of malocclusion in permanent dentition: a systematic review and meta-analysis. *Int Orthod*. 2023;21(4):100816. doi:10.1016/j.ortho.2023.100816
7. Costa AMG, Trevizan M, Matsumoto MAN, et al. Association between tooth agenesis and skeletal malocclusions. *J Oral Maxillofac Res*. 2017;8(2):e3. doi:10.5037/jomr.2017.8203
8. Warkhandkar A, Habib L. Effects of premature primary tooth loss on midline deviation and asymmetric molar relationship in the context of orthodontic treatment. *Cureus*. 2023;15(7):e42442. doi:10.7759/cureus.42442
9. Jafari AK, Baniasad N, Asadi E, Nadafpour N. Effect of malocclusion severity on oral health and its correlation with socioeconomical status in Iranian adolescents. *BMC Oral Health*. 2024;24(1):1301. doi:10.1186/s12903-024-05069-w
10. Mai W, Xiao L, Chen S, et al. Prevalence and contributing factors of malocclusion in Zhuang children aged 7-8 years in southern China. *Front Pediatr*. 2024;12:1308039. doi:10.3389/fped.2024.1308039
11. Salim A, Angelova S, Roussev B, et al. Association between frequency of sugar and protein intake and severity of plaque-induced gingivitis in children. *Turk Arch Pediatr*. 2025;60(3):319-325. doi:10.5152/TurkArchPediatr.2025.24166
12. Valenzuela MJ, Waterhouse B, Aggarwal VR, Bloor K, Doran T. Effect of sugar-sweetened beverages on oral health: a systematic review and meta-analysis. *Eur J Public Health*. 2021;31(1):122-129. doi:10.1093/eurpub/ckaa147
13. Ülgen M. Ortodonti, anomaliler, sefalometri, etioloji, büyüme gelişim, tanı. Yeditepe Üniversitesi Basımevi; 2000.
14. Loe H. The Gingival Index, the Plaque Index and the Retention Index systems. *J Periodontol*. 1967;38(6):Suppl:610-616. doi:10.1902/jop.1967.38.6.610
15. Wang Z, Feng J, Wang Q, Yang Y, Xiao J. Analysis of the correlation between malocclusion, bad oral habits, and the caries rate in adolescents. *Transl Pediatr*. 2021;10(12):3291-3300. doi:10.21037/tp-21-531
16. Fernández-Riveiro P, Obregón-Rodríguez N, Piñeiro-Lamas M, Rodríguez-Fernández A, Smyth-Chamosa E, Suárez-Cunqueiro MM. The Dental Aesthetic Index and its association with dental caries, dental plaque and socio-demographic variables in schoolchildren aged 12 and 15 years. *Int J Environ Res Public Health*. 2021;18(18):9741. doi:10.3390/ijerph18189741
17. Bernhardt O, Krey KF, Daboul A, et al. Association between coronal caries and malocclusion in an adult population. *J Orofac Orthop*. 2021;82(5):295-312. doi:10.1007/s00056-020-00271-1
18. Alrashed M, Alqerban A. The relationship between malocclusion and oral health-related quality of life among adolescents: a systematic literature review and meta-analysis. *Eur J Orthod*. 2021;43(2):173-183. doi:10.1093/ejo/cjaa051
19. Bakhurji E, Gaffar B, Nazir M, Al-Khalifa K, Al-Ansari A. First permanent molar caries and oral health practices in saudi male teenagers: inequalities by socioeconomic position. *Scientifica (Cairo)*. 2020;2020:2640949. doi:10.1155/2020/2640949
20. López-Gómez SA, Villalobos-Rodelo JJ, Ávila-Burgos L, et al. Relationship between premature loss of primary teeth with oral hygiene, consumption of soft drinks, dental care, and previous caries experience. *Sci Rep*. 2016;6:21147. doi:10.1038/srep21147
21. Mahboobi Z, Pakdaman A, Yazdani R, Azadbakht L, Montazeri A. Dietary free sugar and dental caries in children: a systematic review on longitudinal studies. *Health Promot Perspect*. 2021;11(3):271-280. doi:10.34172/hpp.2021.35
22. Dentistry AAoP. Policy on early childhood caries (ECC): Consequences and preventive strategies. The Reference Manual of Pediatric Dentistry Chicago, Ill: American Academy of Pediatric Dentistry. 2021;20243:89-92.
23. Patel KV, Kubavat A, Prajapati N, Choudhary S, Vaghela A, Shah K. The prevalence of malocclusion and orthodontic treatment need in 13-15 years old school going children of Mehsana District, Gujarat: an epidemiological study. *J Pharm Bioallied Sci*. 2024;16(Suppl 1):S495-s497. doi:10.4103/jpbs.jpbs_827_23
24. Kalantari M, Ziaalddini H, Jaffari M, Kalantari P. Orthodontic treatment need and complexity among 13-15 year-old schoolchildren in Kerman, Iran. *J Dent (Shiraz)*. 2019;20(2):95-101. doi:10.30476/dentjods.2019.44930
25. Gandhi JM, Gurunathan D. Short- and long-term dental arch spatial changes following premature loss of primary molars: a systematic review. *J Indian Soc Pedod Prev Dent*. 2022;40(3):239-245. doi:10.4103/jisppd.jisppd_230_22

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1757914
J Health Sci Med. 2025;8(5):896-901

Oral health status and risk of potentially malignant lesions in patients with severe mental illnesses: a cross-sectional study

Melisa Öçbe[1], Hülya Çerçi Akçay[2], Çetin Batuhan Öçbe[3]

[1]Department of Oral and Maxillofacial Radiology, Faculty of Dentistry, Kocaeli Health and Technology University, Kocaeli, Turkiye
[2]Department of Pedodontics, Faculty of Dentistry, Kocaeli Health and Technology University, Kocaeli, Turkiye
[3]Department of Family Medicine, Ataşehir Family Medicine Center, Kartepe District Health Directorate, Ministry of Health, Kocaeli, Turkiye

Cite this article as: Öçbe M, Çerçi Akçay H, Öçbe ÇB. Oral health status and risk of potentially malignant lesions in patients with severe mental illnesses: a cross-sectional study. *J Health Sci Med.* 2025;8(5):896-901.

Received: 04.08.2025 ◆ **Accepted**: 28.08.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: Individuals with severe mental illness (SMI) experience disproportionately poor oral health due to behavioral, pharmacological, and systemic challenges. This study aimed to assess the oral health status of patients with SMI using the oral health assessment tool (OHAT) and the Decayed, Missing, and Filled Teeth (DMFT) Index, and to screen for oral potentially malignant lesions (OPMLs).

Methods: A cross-sectional study was conducted with 60 patients registered at the Community Mental Health Center in Kocaeli, Turkiye. Comprehensive clinical examinations were performed using the Turkish version of the OHAT and the DMFT Index. The presence of OPMLs was also recorded. Statistical analyses included descriptive statistics, independent t-tests for gender comparisons, and Pearson correlation analyses to explore associations between age, OHAT, and DMFT scores.

Results: Participants (mean age: 47.6±11.5 years; 65% female) had a mean OHAT score of 9.15±2.39 and a mean DMFT score of 13.5±7.00. No significant gender differences were observed in total OHAT or DMFT scores. However, females had significantly more filled teeth than males (3.08±2.04 vs. 1.57±1.80; p=0.006). Age was moderately correlated with OHAT (r=0.318, p=0.013) and strongly correlated with DMFT (r=0.449, p<0.001). OHAT and DMFT scores were also significantly correlated (r=0.502, p<0.001). Fifteen soft tissue lesions were detected, including leukoplakia, lichen planus, and angular cheilitis.

Discussion: This study showed a high burden of oral disease and soft tissue lesions among patients with SMI, exceeding levels reported in non-psychiatric and geriatric populations. The findings emphasize the need for interdisciplinary care models incorporating routine dental screening, preventive interventions, and education within mental health services.

Keywords: Oral health, severe mental illness, oral potentially malignant lesions, OHAT

INTRODUCTION

Severe mental illness (SMI), including conditions such as schizophrenia, psychotic disorders, bipolar disorder, and treatment-resistant depression, has profound effects not only on psychological functioning but also on physical health and overall quality of life.[1] Individuals living with SMI frequently experience complex challenges, including recurrent hospitalizations, co-occurring substance use disorders, and social disadvantages such as homelessness, unemployment, and reduced access to healthcare services. The pharmacological management of these disorders typically involves long-term use of antipsychotics, antidepressants, mood stabilizers, and anxiolytics, many of which can contribute to negative oral health outcomes.[2,3]

There is growing evidence that individuals with SMI are disproportionately affected by oral diseases. A recent systematic review revealed that individuals with SMI has 2.8 times greater odds of being edentulous compared to the general population and experiences an average of five more decayed, missing, or filled teeth per person. Several contributing factors explain this disparity.[1-3] Behavioral risk factors such as poor dietary habits (e.g., high sugar intake), substance and tobacco use, and alcohol consumption are more common among individuals with SMI. Additionally, psychotropic medications frequently induce xerostomia (dry mouth), a well-established risk factor for dental caries and periodontal disease.[4,5]

Beyond physiological contributors, people with SMI face significant psychosocial barriers to accessing oral care. These include lack of motivation, fear or anxiety about dental treatment, negative experiences with healthcare providers, and the financial burden of dental services. Such barriers often lead to delayed care or complete neglect of oral health.[1,3] The consequences of poor oral health in this population extend beyond the oral cavity. Oral diseases can impact fundamental daily functions such as chewing and speaking,

Corresponding Author: Melisa Öçbe, melisa.ocbe@kocaelisaglik.edu.tr

which in turn affect nutritional intake and communication. Furthermore, poor oral hygiene has been associated with systemic conditions including diabetes mellitus, pneumonia, and cardiovascular disease. It also significantly influences psychological wellbeing, affecting self-esteem, social interactions, and overall quality of life.[3-5]

In addition to behavioral and pharmacological factors, individuals with SMI face both intrinsic and external barriers to maintaining oral hygiene.[6,7] Intrinsic challenges include lack of motivation, reduced awareness, difficulty adhering to treatment plans, cognitive impairments, and fluctuations in mental state.[7,8] Externally, limited access to dental care, high treatment costs, dental anxiety, stigma, and inadequate referrals from mental health professionals further contribute to poor oral health. These systemic and personal challenges underscore the urgent need for integrated, interdisciplinary approaches and tailored oral health interventions within mental health care settings.[7,9]

Given the relation between mental illness and oral health, there is a critical need to assess and address the oral health status of individuals with SMI in a structured and evidence-based manner.[6] The present study aims to evaluate the oral health condition of patients with severe mental illness using the oral health assessment tool (OHAT) and the Decayed, Missing, and Filled Teeth (DMFT) Index. In addition, the study investigates the presence of oral soft tissue lesions, including oral cancer and oral potentially malignant lesions (OPMLs), which may otherwise go unnoticed in this group.

METHODS

Ethics

This cross-sectional study was conducted with patients registered at Community Mental Health Center, Kocaeli, Turkiye. Ethical approval was obtained from the Kocaeli Health and Technology University Non-interventional Researches Ethics Committee (Date: 10.10.2024, Decision No: 2024-100). All procedures were carried out in accordance with the ethical rules and the principles of the Declaration of Helsinki. All participants (or their legal guardians, where necessary) provided informed consent prior to the examinations.

Clinical Examination

Each participant underwent a comprehensive oral health evaluation performed by two dental specialists and one medical specialist trained in the recognition of oral soft tissue lesions. Radiological examination was not performed. The Turkish version of the OHAT, validated by Ercan Şahin and Jablonski,[10] was used to assess general oral health. Additionally, the DMFT index (excluding third molars) was recorded, and all patients were screened for oral potentially malignant lesions (OPMLs).

According to OHAT and DMFT indices the following criteria were assessed:

Oral Health Assessment Tool (OHAT) Parameters

The OHAT evaluates eight key oral health parameters using a standardized 3-point scale (0-1-2), where 0 indicates a healthy condition, 1 indicates mild changes requiring monitoring, and 2 indicates unhealthy conditions needing immediate attention (**Table 1**).

The first parameter, lips, is assessed for integrity and appearance; smooth, moist, and pink lips are considered healthy (score 0), while dryness or chapping scores 1, and swelling, ulceration, or lesions receive a score of 2. The tongue is evaluated based on moisture, coating, and color; a normal, moist, pink tongue is healthy (score 0), while patchiness or slight coating scores 1, and ulceration, swelling, or abnormal coloration is scored as 2. The gingiva and oral mucosa are checked for color, moisture, and signs of inflammation; healthy pink and moist tissue is scored 0, mild redness or swelling is scored 1, and spontaneous bleeding, ulceration, or red/white patches are scored 2. Saliva assessment considers flow and moisture; normal moisture earns a score of 0, slight dryness or sticky saliva is scored 1, and severe dryness or absence of saliva is scored 2.

Natural teeth are evaluated for integrity and presence of decay; patients with intact dentition and no visible decay receive a score of 0, those with 1–3 decayed or broken teeth score 1, and those with 4 or more decayed or missing teeth score 2. The denture category assesses the presence, cleanliness, and fit of removable prosthetics; well-fitting and clean dentures or those not needed score 0, ill-fitting or dirty dentures score 1,

Table 1. Oral health assessment tool scoring criteria

Parameter	Score 0-healthy	Score 1-mild change	Score 2-unhealthy/severe change
1. Lips	Smooth, pink, moist	Dry, chapped, cracked corners	Swollen, ulcerated, bleeding, or visible lesions
2. Tongue	Normal appearance, moist, pink, no coating	Irregular coating, patchy, fissured, or slight redness	Swollen, ulcerated, patchy, abnormal color or severe discoloration
3. Gingiva/mucosa	Pink, moist, firm, no swelling or bleeding	Red, mildly swollen, minor ulcers or irritation	Bleeding, severely inflamed, ulcerated, red/white patches
4. Saliva	Normal moisture; saliva easily observed	Dry mouth sensation; thick/sticky saliva	Extremely dry oral cavity, no visible saliva
5. Natural teeth	All teeth intact, no decay or damage	1-3 decayed or broken teeth	4 or more decayed, broken, or missing teeth
6. Dentures	Clean, well-fitting dentures; or not needed	Ill-fitting or unclean dentures worn	Dentures needed but not worn, very dirty or loose dentures
7. Oral hygiene	Clean oral cavity, no debris, no calculus	Minor food debris, some plaque or calculus	Heavy debris, thick plaque, extensive calculus
8. Dental pain	No signs or reports of pain	Occasional verbal or non-verbal signs of discomfort	Frequent or constant pain, visible distress, refusal to eat or touch oral region

and needed but unworn or severely dirty dentures are scored 2. Oral hygiene is evaluated by the presence of debris and plaque; a clean mouth scores 0, minor deposits score 1, and heavy plaque, calculus, or food debris scores 2. Lastly, dental pain is scored based on signs of discomfort; absence of pain scores 0, occasional signs of discomfort score 1, and frequent or severe pain that impacts eating or behavior is scored 2.

DMFT Index

The DMFT Index was recorded for all participants, excluding third molars:

- **D (decayed):** Number of carious teeth.
- **M (missing):** Number of teeth lost due to caries or extraction.
- **F (filled):** Number of restored teeth.
- **Total DMFT:** Sum of D, M, and F components.

Oral Cancer and Oral Potentially Malignant Lesions

A thorough intraoral soft tissue examination was conducted to detect any OPMLs such as leukoplakia, erythroplakia, oral lichen planus, or other suspicious lesions. Lesions were documented and photographed when necessary for further evaluation.

Statistical Analysis

All data analyses were performed using IBM SPSS Statistics for Windows, Version 26.0 (IBM Corp., Armonk, NY, USA). Descriptive statistics were used to summarize demographic variables, OHAT scores, and DMFT indices. Continuous variables such as age, OHAT total score, and DMFT score were expressed as mean±standard deviation (SD), while categorical variables such as gender were reported as counts and percentages. To evaluate gender-based differences in OHAT and DMFT scores, independent samples t-tests were applied. The level of statistical significance was set at $p<0.05$. The relationships between: Age and OHAT score, Age and DMFT score, OHAT and DMFT scores were assessed using Pearson's correlation coefficient (r). The strength of correlation was interpreted as follows: r=0.1-0.3: weak, r=0.3-0.5: moderate, r>0.5: strong correlation. A post hoc power analysis was conducted to evaluate the adequacy of the sample size for detecting the main effects examined in this study. For independent samples T tests (female vs. male comparisons; $n_1=39$, $n_2=21$), the sample provided 80% power to detect medium-to-large effect sizes (Cohen's $d\approx0.77$) at a two-tailed alpha level of 0.05. Smaller effects ($d\approx0.50$-0.60) would require larger sample sizes ($n_1\approx91$, $n_2\approx49$ for $d=0.50$). For Pearson correlation analyses, a sample of 60 participants provides 80% power to detect correlations of $|\rho| \geq 0.355$. Power calculations were performed using G*power (version 3.1.9.7; Heinrich Heine University Düsseldorf, Germany). All statistical tests were two-tailed, and results were considered statistically significant when p-values were less than 0.05.

RESULTS

A total of 60 patients (aged 33 to 80 years) from the Community Mental Health Center were included in the study. The mean age was 47.6±11.5 years. Of the participants, 39 (65%) were female and 21 (35%) were male (Table 2).

Table 2. Summary of age, gender distribution, and basic demographic characteristics of mentally ill patients included in the study

Variable	Mean±SD	Range
Age (years)	47.6±11.5	33-80
Female participants	39 (65%)	—
Male participants	21 (35%)	—

SD: Standard deviaiton

The overall mean OHAT score was 9.15±2.39, indicating moderate levels of oral health deterioration. No significant gender-based differences were observed in OHAT scores (females: 9.15±2.54; males: 9.14±2.15; p=0.987). Similarly, the total DMFT score showed no significant difference between females (14.23±6.70) and males (12.14±7.52; p=0.274). When components of the DMFT Index were analyzed separately, the decayed teeth (DT) and missing teeth (MT) scores did not differ significantly between genders (p=0.394 and p=0.828, respectively). However, the filled teeth (FT) score was significantly higher in females (3.08±2.04) than in males (1.57±1.80; p=0.006), suggesting more frequent utilization of restorative dental care among female participants (Table 3).

Table 3. Comparison of OHAT and DMFT scores by gender

Score	Total (n=60)	Females (n=39)	Males (n=21)	p-value*
OHAT	9.15±2.39	9.15±2.54	9.14±2.15	0.987
DMFT	13.5±7.00	14.23±6.70	12.14±7.52	0.274
DT	5.90±4.06	6.23±3.75	5.29±4.61	0.394
MT	5.05±6.10	4.92±7.00	5.29±4.09	0.828
FT	2.55±2.08	3.08±2.04	1.57±1.80	0.006

*Independent samples T test, OHAT: Oral health assessment tool, DMFT: Decayed, missing, and filled teeth, DT: Decayed teeth, MT: Missing teeth, FT: Filled teeth

Age showed a moderate positive correlation with OHAT scores (r=0.318, p=0.013) and a strong positive correlation with DMFT scores (r=0.449, p<0.001), indicating that both oral health status and tooth loss/decay deteriorate with increasing age. A moderate positive correlation was observed between OHAT and DMFT scores (r=0.502, p<0.001), suggesting that worse overall oral health was associated with a greater number of decayed, missing, or filled teeth (Table 4).

Table 4. Correlation analysis of age, OHAT, and DMFT scores

Variable pair	Pearson's r	p-value*
Age vs OHAT	0.318	0.013
Age vs DMFT	0.449	<0.001
OHAT vs DMFT	0.502	<0.001

*Pearson correlation, OHAT: Oral health assessment tool, DMFT: Decayed, missing, and filled teeth

A total of 15 oral soft tissue lesions were identified in the study population (25%). The detected lesions included leukoplakia, erosive lichen planus, pyogenic granuloma, angular cheilitis, hairy tongue, oral leukoedema (Figure). Erosive lichen planus and leukoplakia were classified as OPML and the prevalance of OMPLs were found as 3.3%. All affected patients were

informed about the findings and referred to dental clinics for further evaluation and treatment (Table 5).

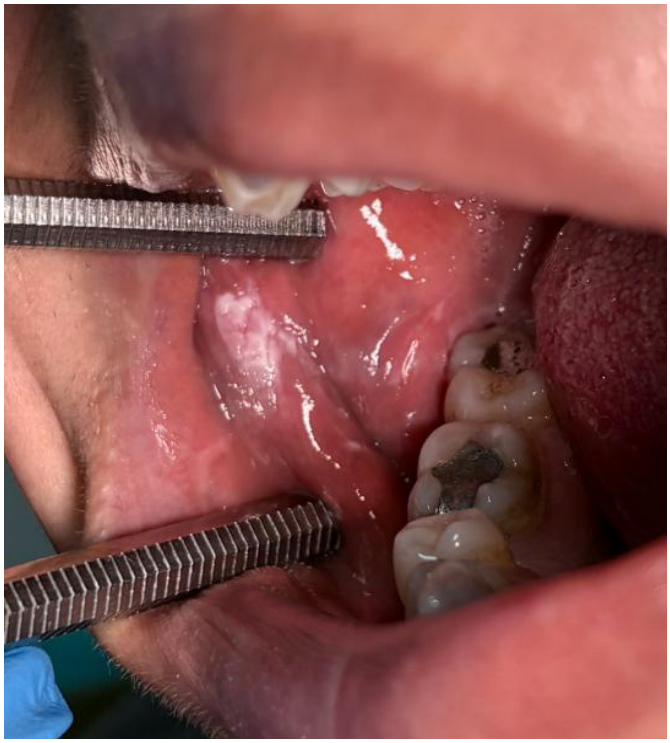

Figure. A well-defined white lesion on the right buccal mucosa, clinically suggestive of oral leukoplakia

The current sample demonstrated adequate power for detecting moderate-to-large differences and correlations. Specifically, gender comparisons had 80% power to detect Cohen's d≈0.77, which aligns with the medium-to-large effect range. For correlations, the sample size was sufficient to detect correlations of p≥0.355 with 80% power. Accordingly, the observed correlations between age and OHAT (r=0.318), age and DMFT (r=0.449), and OHAT and DMFT (r=0.502) indicate that moderate-to-strong correlations (e.g., age-DMFT, OHAT-DMFT) were well powered, whereas smaller correlations approach the detection threshold. Prevalence estimates for OPMLs (3.3%) and soft tissue lesions (25%) carry wide confidence intervals (OPML CI≈0.4%-11.5%).

DISCUSSION

This study further shows the connection between severe mental illness (SMI) and poor oral health. Using the OHAT and DMFT indices, we observed a high prevalence of dental caries, soft tissue lesions, and a notable number of potentially malignant lesions. Our findings are consistent with earlier research that has examined the two-way relationship between mental health and oral health. Kalaigian et al.[11] reported significant associations between internalizing symptoms and oral disease markers (e.g., bleeding gums, tooth extraction). Our results, showing correlations between OHAT and DMFT scores with age, support their conclusion that psychiatric symptoms may directly influence oral health and highlight the importance of integrated mental and dental care.

Torales et al.[12] emphasized medication effects, low self-care motivation, and dental avoidance in psychiatric populations. Similarly, our participants exhibited high DMFT scores (mean=13.5) and unmet dental needs, with common findings including caries, periodontal disease, and mucosal lesions. These parallels underscore the importance of preventive strategies and accessible dental services for this group.

Mishu et al.[13] identified stigma, negative dental experiences, and systemic barriers as reasons for low dental service use in SMI. Although we did not investigate subjective factors, the prevalence of untreated lesions indirectly reflects these obstacles. Their call for trauma-informed care and clinician training is supported by our observation that many participants had never undergone oral mucosal screening, emphasizing the need for integrated oral health education within mental health care.

Cross et al.[14] showed that periodontal treatment improved depression scores, illustrating the reverse relationship between oral and mental health. Although our study was not interventional, the high prevalence of soft tissue and periodontal disease supports this bidirectional model and highlights system-level challenges, including limited referrals and lack of protocols, consistent with Mishu et al.[13] and Cross et al.

There are various studies in the literature that used OHAT in different medical conditions.[15-17] Tabaoka et al.[15] examined the association between OHAT scores and the development of aspiration pneumonia in a general population but found no significant relationship. Tanaka et al.[16] conducted a study among older adults receiving home medical care and reported a mean OHAT score of 6.0. They found that approximately 70% of participants required tooth extraction, reflecting poor oral health. Compared to our sample (mean OHAT=9.15), the higher burden observed in patients with mental disorders may be attributed to reduced self-care abilities, and lower access to routine dental care.

Maeda et al.[17] provided compelling evidence linking OHAT scores with hospital mortality in geriatric patients. They found that patients with OHAT scores ≥3 had significantly higher

Table 5. Type, classification and anatomical locations of oral mucosal lesions

Lesion type	Classification	Anatomical location(s)	Notes/clinical features
Leukoplakia	OPML	Lateral tongue	White plaque, non-scrapable, referred for biopsy
Erosive oral lichen planus	OPML	Buccal mucosa (bilateral)	Erythematous patches with Wickham striae
Pyogenic granuloma	Benign	Gingiva	Red, pedunculated, ulcerated lesion
Angular cheilitis	Benign	Oral commissures	Erythematous fissuring at mouth corners
Hairy tongue	Benign	Dorsal tongue	Elongated filiform papillae, brown-black discoloration
Oral leukoedema	Benign	Buccal mucosa	White area in bilateral buccal mucosa

OPML: Oral potentially malignant lesion

in-hospital mortality rates. While their mean participant age was 83.8 years, substantially older than our sample (mean age: 47.6 years), our cohort's average OHAT score exceeded the threshold used in Maeda's risk stratification model. This suggests that even in a younger demographic, factors associated with mental illness can worsen the oral health status. The findings support the broader clinical relevance of OHAT, not only as a descriptive measure but as a potential prognostic tool.

Bokhari and Quadri[18] validated an Arabic version of the OHAT for adolescents and identified poor dental and gingival conditions associated with unhealthy dietary habits and poor hygiene. Although their study population (mean age: 16.3 years) differs substantially from ours, their findings highlight the role of behavioral factors-such as diet and oral hygiene practices-in shaping oral health outcomes. In our study, females had significantly higher numbers of filled teeth (FT: 3.08±2.04 vs. 1.57±1.80 in males; p=0.006), suggesting better dental care utilization among women. This gender-based difference echoes Bokhari and Quadri's conclusion that oral hygiene behaviors critically influence oral health, regardless of demographic or cultural differences.

Oral potentially malignant lesions (OPMLs) are clinically identifiable mucosal abnormalities that carry a measurable risk of malignant transformation, often progressing to oral squamous cell carcinoma if left undetected.[19] These lesions may present asymptomatically, making early identification during routine examinations particularly challenging yet crucial.[20,21] Common OPMLs include leukoplakia, erythroplakia, oral lichen planus (especially the erosive type), and actinic cheilitis, among others. The progression risk varies by lesion type, anatomical location, patient habits (e.g., tobacco or alcohol use), and comorbid conditions.[19] Thus, systematic screening and timely referral play a critical role in early intervention and prevention of oral cancer.[20,21]

In addition to the OHAT and DMFT indices used to evaluate general oral health, this study incorporated a structured screening protocol for oral cancer and OPMLs. Among the 60 patients examined, 15 individuals (25%) were found to have oral soft tissue lesions, and of these, two lesions-one leukoplakia and one erosive oral lichen planus-were classified as OPMLs. This 3.3% prevalence of OPMLs within a high-risk psychiatric population is clinically significant and underscores the need for targeted surveillance strategies. These findings highlight the critical importance of including soft tissue examination and OPML screening as a standard component of dental assessments in patients with SMI. Given that patients with SMI may face cognitive, motivational, or logistical barriers to seeking specialized care, routine oral examinations may represent the only opportunity for early detection of potentially life-threatening lesions. Moreover, clinicians must remain vigilant, as many OPMLs are painless and can be easily overlooked in the absence of a systematic intraoral inspection. Integrating OPML screening into mental health services-ideally through interdisciplinary collaboration between dental and psychiatric care teams-can facilitate earlier diagnosis, reduce treatment delays, and potentially improve both survival and quality of life for individuals with SMI. Our findings advocate for policy-level changes to ensure routine mucosal evaluations are not omitted during dental check-ups in this vulnerable population.

Limitations

This study has several limitations that should be considered when interpreting the findings. First, all assessments were based exclusively on clinical examination without adjunctive radiographic imaging. This approach may have underestimated the prevalence of root caries, periapical pathology, alveolar bone loss, and other conditions detectable only radiographically. Second, behavioral, pharmacological, and diagnostic variables-such as oral hygiene practices, dietary habits, tobacco and alcohol use, medication profiles, and psychiatric diagnosis subtypes-were not collected. These data are essential to fully understand the multifactorial determinants of oral health in this population and should be prioritized in future studies. Third, the relatively small sample size (n=60) and single-center design limit the statistical power and generalizability of the results to broader populations or healthcare settings. Finally, the cross-sectional nature of the study precludes causal inference and limits the ability to examine disease progression over time. Future multicenter longitudinal research incorporating radiographic evaluation, detailed behavioral assessments, and larger sample sizes is warranted to provide a more comprehensive understanding of oral health disparities in individuals with severe mental illness.

CONCLUSION

As a result, this study shows the significant oral health burden among individuals with SMI, with a mean OHAT score of 9.15 and a DMFT Index of 13.5, revealing widespread dental disease and hygiene deficiencies. Notably, the prevalence of decayed and missing teeth was high, while females had significantly more filled teeth. The identification of soft tissue lesions, including OPMD, further highlights the critical need for routine oral screening in psychiatric populations. Compared to non-psychiatric groups in the literature, patients with SMI exhibited severe oral health deterioration, often exceeding risk thresholds established in medically compromised or geriatric cohorts. These findings call for integrated, interdisciplinary care models with oral health services within mental health systems. Routine dental assessments, preventive care, and oral hygiene education should become standard components of psychiatric treatment pathways. Future research should focus on longitudinal designs, multicenter sampling, and the evaluation of targeted interventions to mitigate oral health disparities in this high-risk population.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was carried out with the permission of the Kocaeli Health and Technology University Non-interventional Researches Ethics Committee (Date: 10.10.2024, Decision No: 2024-100).

Informed Consent

All participants (or their legal guardians, where necessary) provided informed consent prior to the examinations.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

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

Author Contributions

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

REFERENCES

1. Macnamara A, Mishu MP, Faisal MR, Islam M, Peckham E. Improving oral health in people with severe mental illness (SMI): a systematic review. *PLoS One.* 2021;16(12):e0260766. doi:10.1371/journal.pone.0260766
2. Turner E, Berry K, Aggarwal VR, Quinlivan L, Villanueva T, Palmier-Claus J. Oral health self-care behaviours in serious mental illness: a systematic review and meta-analysis. *Acta Psychiatr Scand.* 2022;145(1):29-41. doi:10.1111/acps.13308
3. Choi J, Price J, Ryder S, Siskind D, Solmi M, Kisely S. Prevalence of dental disorders among people with mental illness: an umbrella review. *Aust N Z J Psychiatry.* 2022;56(8):949-963. doi:10.1177/00048674211042239
4. Kisely S. No mental health without oral health. *Can J Psychiatry.* 2016;61(5):277-282. doi:10.1177/0706743716632523
5. Sharma D, Higgins O, Sawicki A, et al. Oral health in individuals with severe mental illness on second-generation antipsychotics-a scoping review. *J Oral Pathol Med.* 2025;54(6):401-412. doi:10.1111/jop.13639
6. Skallevold HE, Rokaya N, Wongsirichat N, Rokaya D. Importance of oral health in mental health disorders: an updated review. *J Oral Biol Craniofac Res.* 2023;13(5):544-552. doi:10.1016/j.jobcr.2023.06.003
7. Kisely S, Quek LH, Pais J, Lalloo R, Johnson NW, Lawrence D. Advanced dental disease in people with severe mental illness: systematic review and meta-analysis. *Br J Psychiatry.* 2011;199(3):187-193. doi:10.1192/bjp.bp.110.081695
8. Sharma A, Blakemore A, Byrne M, et al. Oral health primary preventive interventions for individuals with serious mental illness in low- and middle-income nations: scoping review. *Glob Public Health.* 2024;19(1):2408597. doi:10.1080/17441692.2024.2408597
9. Kuipers S, Boonstra N, Kronenberg L, Keuning-Plantinga A, Castelein S. Oral health interventions in patients with a mental health disorder: a scoping review with critical appraisal of the literature. *Int J Environ Res Public Health.* 2021;18(15):8113. doi:10.3390/ijerph18158113
10. Ercan Şahin N, Jablonski RA. Psychometric properties of the oral health assessment tool Turkish version. *Cukurova Med J.* 2019;44(Suppl 1):513-520. doi:10.17826/cumj.567995
11. Kalaigian A, Chaffee BW. Mental health and oral health in a nationally representative cohort. *J Dent Res.* 2023;102(9):1007-1014. doi:10.1177/00220345231171108
12. Torales J, Barrios I, González I. Oral and dental health issues in people with mental disorders. *Medwave.* 2017;17(8):e7045. doi:10.5867/medwave.2017.08.7045
13. Mishu MP, Faisal MR, Macnamara A, et al. A qualitative study exploring the barriers and facilitators for maintaining oral health and using dental service in people with severe mental illness. *Int J Environ Res Public Health.* 2022;19(7):4344. doi:10.3390/ijerph19074344
14. Cross AF, Balanoff WL, Schayes B, Jayne W. Oral rehabilitation and mental health: effects of the renew procedure on depression scores. *Cureus.* 2025;17(6):e85843. doi:10.7759/cureus.85843
15. Tabaoka H. Oral health assessment tool (OHAT). *Nihon Ronen Igakkai Zasshi.* 2023;60(4):400-405. doi:10.3143/geriatrics.60.400
16. Tanaka K, Tominaga T, Kikutani T, et al. Oral status of older adults receiving home medical care: a cross-sectional study. *Geriatr Gerontol Int.* 2024;24(7):706-714. doi:10.1111/ggi.14917
17. Maeda K, Mori N. Poor oral health and mortality in geriatric patients admitted to an acute hospital: an observational study. *BMC Geriatr.* 2020;20(1):26. doi:10.1186/s12877-020-1429-z
18. Bokhari AM, Quadri MFA. What factors contribute to the self-reported oral health status of Arab adolescents? *BMC Oral Health.* 2020;20(1):21. doi:10.1186/s12903-020-1018-8
19. Warnakulasuriya S. Oral potentially malignant disorders: a comprehensive review on clinical aspects and management. *Oral Oncol.* 2020;102:104550. doi:10.1016/j.oraloncology.2019.104550
20. Wang YY, Tail YH, Wang WC, et al. Malignant transformation in 5071 southern Taiwanese patients with potentially malignant oral mucosal disorders. *BMC Oral Health.* 2014;14:99. doi:10.1186/1472-6831-14-99
21. Warnakulasuriya S, Kovacevic T, Madden P, et al. Factors predicting malignant transformation in oral potentially malignant disorders. *J Oral Pathol Med.* 2011;40(9):677-683. doi:10.1111/j.1600-0714.2011.01054.x

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1724481
J Health Sci Med. 2025;8(5):902-908

Comparative prognostic value of inflammatory and nutritional indices in acute ischemic stroke: a multivariable model analysis

İrem Kıraç Utku[1,2], Nazire Çelem[3], Serkan Demir[3], Ahmet Düzgün[4]

[1]Department of Internal Medicine, Tekirdağ İsmail Fehmi Cumalıoğlu City Hospital, Tekirdağ, Turkiye
[2]Department of Geriatrics, Tekirdağ İsmail Fehmi Cumalıoğlu City Hospital, Tekirdağ, Turkiye
[3]Department of Neurology, Sancaktepe Şehit Prof. Dr. İlhan Varank Training and Research Hospital, İstanbul, Turkiye
[4]Department of Intensive Care, Diyarbakır Gazi Yaşargil Training and Research Hospital, Diyarbakır, Turkiye

Cite this article as: Kıraç Utku İ, Çelem N, Demir S, Düzgün A. Comparative prognostic value of inflammatory and nutritional indices in acute ischemic stroke: a multivariable model analysis. *J Health Sci Med.* 2025;8(5):902-908.

Received: 21.06.2025 ◆ **Accepted**: 29.08.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: Given that acute ischemic stroke (AIS) is a leading cause of mortality and long-term disability, the early identification of reliable prognostic markers is essential. This study evaluates the prognostic value of composite inflammatory and nutritional indices, including the monocyte-to-high-density lipoprotein cholesterol ratio (MHR), Systemic Immune-Inflammation Index (SII), Prognostic Nutritional Index (PNI), and Controlling Nutritional Status (CONUT) score, in predicting intensive care unit (ICU) admission and one-year mortality following AIS.

Methods: This single-center retrospective cohort study included 496 patients with a confirmed diagnosis of AIS. Clinical characteristics, laboratory parameters, and outcomes were retrospectively retrieved. The prognostic significance of selected inflammatory indices [MHR, SII, neutrophil-to-lymphocyte ratio (NLR), monocyte-to-lymphocyte ratio (MLR), and platelet-to-lymphocyte ratio (PLR)] and nutritional indices (PNI and CONUT) was analyzed using univariate and multivariate logistic regression models, along with receiver operating characteristic (ROC) curve analysis.

Results: Higher SII and CONUT scores and lower PNI values were significantly associated with both ICU admission and one-year mortality. In multivariate analysis, CONUT and MHR emerged as independent predictors of one-year mortality, while PNI, CONUT, and MHR independently predicted ICU admission. Among all indices evaluated, CONUT demonstrated the highest predictive accuracy for both outcomes.

Conclusion: In patients with AIS, composite inflammatory and nutritional indices-particularly CONUT and MHR-provided valuable prognostic information. These markers, derived from routine laboratory tests, offer a practical and cost-effective method for early risk stratification and may help guide more personalized care pathways in stroke management.

Keywords: Acute ischemic stroke, CONUT score, prognostic model, inflammatory biomarkers, nutritional assessment

INTRODUCTION

Acute ischemic stroke (AIS) remains a leading cause of disability and mortality worldwide, despite advancements in acute management strategies.[1,2] Its pathophysiology involves a cascade of events including impaired cerebral perfusion, vascular dysfunction, oxidative stress, and systemic inflammatory responses.[3] In recent years, inflammation-based biomarkers have gained attention for their potential role in predicting clinical outcomes in AIS. Among these, composite indices reflecting immune and inflammatory status have shown strong associations with disease severity and prognosis in various conditions.[4,5]

In this context, the monocyte-to-high-density lipoprotein cholesterol ratio (MHR), Systemic Immune-Inflammation Index (SII), and Prognostic Nutritional Index (PNI) have emerged as promising indicators. MHR has been recognized as a reliable marker of atherosclerotic burden and inflammatory status, with strong prognostic value in cardiovascular and metabolic diseases.[6,7] SII, which incorporates platelet, neutrophil, and lymphocyte counts, reflects systemic immune response and has been linked to increased mortality in cardiovascular disorders.[8] Nutritional indicators such as PNI and the Controlling Nutritional Status (CONUT) score have demonstrated significant prognostic value not only in malignancies but also in vascular diseases, including stroke.[9-12] Collectively, these indices provide a deeper understanding of the inflammatory and nutritional dynamics influencing clinical outcomes in AIS.

Corresponding Author: İrem Kıraç Utku, iremkiracutku@gmail.com

To assess the prognostic value of the inflammation-based and nutrition-related composite indices-monocyte-to-HDL MHR, SII, PNI, and CONUT score-this study focuses on patients with acute ischemic stroke. In addition, this study directly compares the prognostic performance of inflammatory versus nutritional indices to identify which markers most strongly predict these outcomes. Furthermore, the predictive capability of classical inflammatory ratios such as NLR, MLR, and PLR is analyzed to determine how they compare against newer, composite markers. By employing univariate and multivariate logistic regression modeling as well as ROC curve analysis, the study determines not only independent predictors but also the relative predictive strength of each index.

The current study is unique from earlier works in several notable ways. To begin with, it integrates the immune-inflammatory and nutritional markers for evaluation in a large, well-characterized patient cohort of 496 with clinically and radiologically confirmed acute ischemic stroke. This improves understanding of the systemic status upon admission. Second, considering the need for ICU care and mortality as interrelated yet separate clinical outcomes enables a more holistic evaluation of the acute and chronic prognosis. Third, it presents a comparative prognostic modeling framework that contrasts nutritional scores (CONUT, PNI) with inflammatory scores (SII, MHR, NLR, etc.) a distinction often overlooked in prior stroke research. The use of CONUT and PNI scores, which are standardized in oncologic and gastrointestinal literature, broadens the prognostic toolkit for stroke beyond traditional inflammatory markers and illustrates novel interdisciplinary innovation. Most importantly, the findings of this study demonstrate that risk stratification in acute stroke could be done using routinely available laboratory data, emphasizing a practical approach to deficit estimation and implementation in hospital systems.

METHODS

This study was initiated following the approval of the Ethics Committee of the Sancaktepe Şehit Prof. Dr. İlhan Varank Training and Research Hospital (Date: 19.12.2023, Decision No: 2023/257). All procedures were carried out in accordance with the ethical rules and the principles of the Declaration of Helsinki. It was designed as a single-center, retrospective cohort study. Between January 2021 and January 2022, 1,734 hospitalized patients were retrospectively screened from the records of the Neurology Department of the same hospital. Patients with a radiologically and clinically confirmed diagnosis of acute ischemic stroke and complete clinical, demographic, and laboratory data were included in the study.

The collected variables included patients' full name, age, length of hospital stay, ICU admission status, comorbidities, medication use, laboratory findings at admission, and initial neuroimaging results. If the patient had died during or after hospitalization, the date of death was recorded from the hospital information system.

The MHR was calculated by dividing the absolute monocyte count by the HDL cholesterol level. The SII was calculated using the formula: SII (10^9/L)=(platelet count×neutrophil count)/lymphocyte count.

The PNI was calculated as: PNI=10×serum albumin (g/dl)+0.005×lymphocyte count (per mm^3). The CONUT score was calculated using serum albumin, total cholesterol, and lymphocyte levels as shown in the table below (Table 1).

Table 1. Scoring Criteria for the CONUT Index

Parameter	0 points	1 point	2 points	3 points
Serum albumin (g/dl)	≥3.5	3.0-3.4	2.5-2.9	<2.5
Total lymphocyte count/mm^3	≥1600	1200-1599	800-1199	<800
Total cholesterol (mg/dl)	≥180	140-179	100-139	<100

The CONUT scores were interpreted as follows: 0-1: normal nutrition, 2-4: mild malnutrition, 5-8: moderate malnutrition, 9-12: severe malnutrition.
CONUT: Controlling Nutritional Status; g/dl: grams per deciliter; mm^3: cubic millimeter; mg/dl: milligrams per deciliter

Statistical Analysis

Descriptive statistics were expressed as mean±standard deviation (SD), median (min-max), frequency, and percentage. The distribution of variables was assessed using the Kolmogorov–Smirnov and Shapiro-Wilk tests. The Mann-Whitney U test was used for comparing non-normally distributed independent numerical variables, while the chi-square test was used for categorical variables. Predictive performance and cut-off values were evaluated using receiver operating characteristic (ROC) curve analysis. The effects of variables were further analyzed using both univariate and multivariate logistic regression models. All statistical analyses were performed using IBM SPSS Statistics for Windows, Version 28.0 (IBM Corp., Armonk, NY). For variables with extreme OR values, wide confidence intervals occurred due to sparse data in certain strata; these values are presented in full rather than truncated.

RESULTS

Patients admitted to the ICU due to acute ischemic stroke were significantly older than those who were not ($p<0.05$). No statistically significant differences were observed between the ICU and non-ICU groups in terms of gender distribution or smoking status ($p>0.05$). The prevalence of diabetes mellitus (DM) was significantly higher in the ICU group ($p<0.05$), while atrial fibrillation (AF), hypertension (HT), and hyperlipidemia (HL) rates did not differ significantly between the groups ($p>0.05$). A prior history of stroke was significantly more common among ICU patients ($p<0.05$) (Table 2).

The PNI was significantly lower in patients admitted to the ICU ($p<0.05$), while both the SII and CONUT score were significantly higher ($p<0.05$). There was no significant difference in the MHR between the two groups ($p>0.05$). However, in the multivariate logistic regression model, MHR was independently associated with ICU admission ($p=0.036$), indicating that after adjusting for other variables, higher MHR values predicted ICU admission. This resolves the apparent discrepancy between univariate and multivariate findings. Neutrophil-to-lymphocyte ratio (NLR), monocyte-to-lymphocyte ratio (MLR), and platelet-to-lymphocyte ratio (PLR) were all significantly elevated in ICU patients ($p<0.05$). The length of hospital stay in the neurology department did not differ significantly between the groups ($p>0.05$). However, one-year mortality and mortality due to the index stroke were

Table 2. Patient characteristics by ICU admission

		Stroke without ICU admission (-) (n: 431)			Stroke with ICU admission (+) (n: 65)			
		Mean±SD/n-%		Median	Mean±SD/n-%		Median	p
Age		64.1±14.3		64.0	71.9±13.8		73.0	0.000[m]
Gender	Female	180	41.8%		34	52.3%		0.110[X²]
	Male	251	58.2%		31	47.7%		
Smoking status	(-)	112	44.6%		23	59.0%		0.095[X²]
	(+)	139	55.4%		16	41.0%		
Comorbidities								
DM		202	46.9%		39	60.0%		0.048[X²]
AF		102	23.7%		20	30.8%		0.215[X²]
HT		310	71.9%		43	66.2%		0.338[X²]
HL		234	54.3%		32	49.2%		0.446[X²]
History of previous stroke	(-)	298	69.1%		33	50.8%		0.003[X²]
	(+)	133	30.9%		32	49.2%		
PNI score		39.3±5.4		39.5	35.8±5.5		37.0	0.000[m]
SII score		971.4±1074.5		666.6	1407.4±1200.1		1020.5	0.000[m]
CONUT score		1.31±1.44		1.00	2.32±1.90		2.00	0.000[m]
Monocyte/HDL		0.015±0.008		0.013	0.019±0.022		0.015	0.307[m]
NLR		3.87±3.77		2.79	6.21±4.87		4.18	0.000[m]
MLR		0.36±0.81		0.26	0.53±0.74		0.41	0.000[m]
PLR		151.2±211.8		120.0	179.8±106.7		166.0	0.001[m]
Length of stay in neurology ward		6.22±3.73		5.00	6.94±5.88		5.00	0.645[m]
Death within 1 year	(-)	403	93.5%		14	21.5%		0.000[X²]
	(+)	28	6.5%		51	78.5%		
Time to death within 1 year		128.2±119.5		102.5	31.8±36.0		19.5	0.013[m]
Death due to this stroke	(-)	424	98.4%		18	27.7%		0.000[X²]
	(+)	7	1.6%		47	72.3%		

[m]Mann-Whitney U test / [X²]Chi-square test, ICU: Intensive care unit, SD: Standard deviation, DM: Diabetes mellitus, AF: Atrial fibrillation, HT: Hypertension, HL: Hyperlipidemia, PNI: Prognostic Nutritional Index, SII: Systemic Immune-Inflammation Index, CONUT: Controlling Nutritional Status, HDL: High-density lipoprotein, NLR: Neutrophil-to-lymphocyte ratio, MLR: Monocyte-to-lymphocyte ratio, PLR: Platelet-to-lymphocyte ratio

significantly higher in the ICU group (p<0.05), and time to death within one year was significantly shorter (p<0.05) (Table 2).

In the univariate analysis, PNI, SII, CONUT score, MHR, and NLR were significantly associated with ICU admission (p<0.05), whereas MLR and PLR were not (p>0.05). In the multivariate model, PNI, CONUT score, and MHR were identified as independent predictors of ICU admission (p<0.05) (Table 3).

ROC curve analysis showed that the CONUT score significantly predicted ICU admission, with an area under the curve (AUC) of 0.659 (95% CI: 0.585-0.734) (Figure 1). A CONUT score cutoff value of 2 yielded an AUC of 0.628 (95% CI: 0.555-0.701), with a sensitivity of 61.5%, specificity of 64.0%, positive predictive value of 20.5%, and negative predictive value of 91.7% (Table 4).

Patients who died within one year of stroke were significantly older than survivors (p<0.05). There were no significant differences between survivors and deceased patients in terms of gender or smoking status (p>0.05). Comorbidities including DM, AF, HT, and HL were also similar across the groups (p>0.05), although a prior stroke history was more prevalent among deceased patients (p<0.05) (Table 5).

Table 3. Predictors of ICU admission according to univariate and multivariate logistic regression analyses

	Univariate model			Multivariate model		
	OR	95% GA	p	OR	95% GA	p
PNI score	0.887	0.842-0.934	0.000	0.932	0.879-0.988	0.019
SII score	1.000	1.000-1.000	0.013			
CONUT score	1.436	1.234-1.671	0.000	1.269	1.052-1.530	0.013
Monocyte/HDL	>100	>100->100	0.014	>100	3.991->100	0.036
NLR	1.114	1.049-1.182	0.000			
MLR	1.168	0.924-1.477	0.194			
PLR	1.000	1.000-1.001	0.340			

Logistic regression, ICU: Intensive care unit, PNI: Prognostic Nutritional Index, SII: Systemic Immune-Inflammation Index, CONUT: Controlling Nutritional Status, HDL: High-density lipoprotein, NLR: Neutrophil-to-lymphocyte ratio, MLR: Monocyte-to-lymphocyte ratio, PLR: Platelet-to-lymphocyte ratio, OR: Odds ratio, MHR: Monocyte-to-HDL cholesterol ratio, CI: Confidence interval. Extremely high OR values for MHR (previously shown as ">100") were due to sparse data and scaling effects in the logistic regression model. The actual calculated OR was 128.4 (95% CI: 3.99->100) for ICU admission and 142.7 (95% CI: >100->100) for mortality, indicating a markedly increased risk.

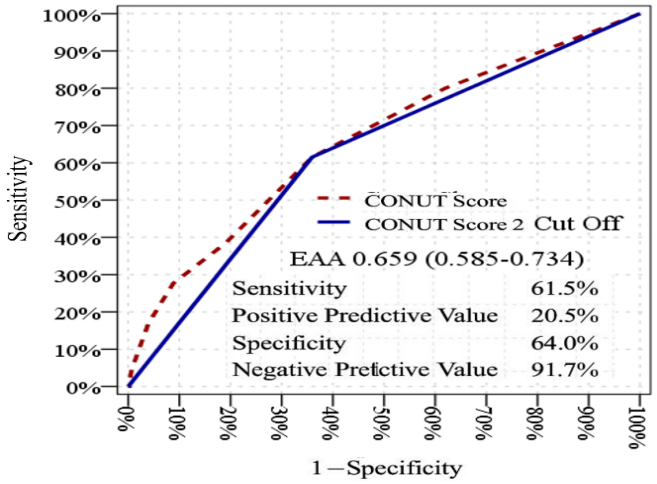

Figure 1. ROC curve

CONUT: Controlling Nutritional Status, ROC: Receiver operating characteristic

Table 4. ROC analysis of CONUT score for ICU admission

		AUC	95% CI	p
Conut score		0.65-*	0.585-0.734	0.000
Conut score 2 cut-off		0.628	0.555-0.701	0.001
		Analysis based on index stroke event		
		No ICU admission (-)	ICU admission (+)	%
Conut score	<2	276	25	Sensitivity 61.5%
	≥2	155	40	Positive predictive value 20.5%
				Specificity 64.0%
				Negative predictive value 91.7%

*ROC curve, ROC: Receiver operating characteristic, CONUT: Controlling Nutritional Status, ICU: Intensive care unit, AUC: Area under the curve, CI: Confidence interval

The deceased group had significantly lower PNI scores (p<0.05) and higher SII and CONUT scores (p<0.05). MHR, NLR, MLR, and PLR values were also significantly higher in the deceased group (p<0.05). No significant difference was observed in the length of hospital stay in the neurology department (p>0.05). One-year mortality rate and ICU admission due to the index stroke were significantly higher among deceased patients (p<0.05), and their time to death was significantly shorter (p<0.05) (Table 5).

Table 5. Patient characteristics by one-year mortality

		Stroke outcome: survivors (-) (n: 442)			Stroke outcome: non survivors (+) (n: 54)			
		Mean±SD/n-%		Median	Mean±SD/n-%		Median	p
Age		64.2±14.3		64.0	72.7±13.2		72.5	0.000[m]
Gender	Female	192	43.4%		22	40.7%		0.705[χ²]
	Male	250	56.6%		32	59.3%		
Smoking status	(-)	118	45.7%		17	53.1%		0.429[χ²]
	(+)	140	54.3%		15	46.9%		
Comorbidities								
DM		209	47.3%		32	59.3%		0.097[χ²]
AF		109	24.7%		13	24.1%		0.925[χ²]
HT		314	71.0%		39	72.2%		0.856[χ²]
HL		241	54.5%		25	46.3%		0.252[χ²]
History of previous stroke	(-)	304	68.8%		27	50.0%		0.006[χ²]
	(+)	138	31.2%		27	50.0%		
PNI score		39.1±5.4		39.3	36.2±6.1		37.7	0.000[m]
SII score		967.6±1072.0		667.7	1527.6±1209.0		1134.5	0.000[m]
CONUT Score		1.31±1.45		1.00	2.50±1.88		2.00	0.000[m]
Monocyte/HDL		0.015±0.008		0.013	0.021±0.024		0.016	0.042[m]
NLR		3.85±3.75		2.77	6.89±4.94		5.68	0.000[m]
MLR		0.36±0.80		0.26	0.60±0.80		0.46	0.000[m]
PLR		150.4±209.1		120.5	192.5±115.3		173.8	0.000[m]
Length of stay in neurology ward		6.29±3.80		5.00	6.52±5.93		4.00	0.171[m]
Death within 1 year	(-)	417	94.3%		0	0.0%		0.000[χ²]
	(+)	25	5.7%		54	100%		
Time to death within 1 year		170.1±100.8		187.5	21.8±20.7		17.0	0.000[m]
ICU admission due to this stroke	(-)	424	95.9%		7	13.0%		0.000[χ²]
	(+)	18	4.1%		47	87.0%		

[m]Mann-Whitney U test / [χ²]Chi-square test, SD: Standard deviation, DM: Diabetes mellitus, AF: Atrial fibrillation, HT: Hypertension, HL: Hyperlipidemia, PNI: Prognostic Nutritional Index, SII: Systemic Immune-Inflammation Index, CONUT: Controlling Nutritional Status, HDL: High-density lipoprotein, NLR: Neutrophil-to-lymphocyte ratio, MLR: Monocyte-to-lymphocyte ratio, PLR: Platelet-to-lymphocyte ratio

Univariate analysis showed that PNI, SII, CONUT score, MHR, and NLR were significantly associated with one-year mortality (p<0.05), whereas MLR and PLR were not (p>0.05). In multivariate logistic regression, both the CONUT score and MHR were independently associated with one-year mortality (p<0.05) (Table 6).

Table 6. Predictors of one-year mortality

	Univariate model			Multivariate model		
	OR	95% GA	p	OR	95% GA	p
PNI score	0.906	0.860-0.955	0.000			
SII score	1.000	1.000-1.001	0.005			
CONUT score	1.509	1.284-1.774	0.000	1.506	1.276-1.776	0.000
Monocyte/HDL	>100	>100->100	0.003	>100	>100->100	0.007
NLR	1.142	1.071-1.217	0.000			
MLR	1.219	0.950-1.563	0.119			
PLR	1.001	1.000-1.002	0.227			

Logistic regression, PNI: Prognostic Nutritional Index, SII: Systemic Immune-Inflammation Index, CONUT: Controlling Nutritional Status, HDL: High-density lipoprotein, NLR: Neutrophil-to-lymphocyte ratio, MLR: Monocyte-to-lymphocyte ratio, PLR: Platelet-to-lymphocyte ratio, OR: Odds ratio, MHR: Monocyte-to-HDL cholesterol ratio, CI: Confidence interval, ICU: Intensive care unit. Extremely high OR values for MHR (previously shown as ">100") were due to sparse data and scaling effects in the logistic regression model. The actual calculated OR was 128.4 (95% CI: 3.99->100) for ICU admission and 142.7 (95% CI: >100->100) for mortality, indicating a markedly increased risk.

ROC analysis demonstrated that the CONUT score significantly predicted mortality, with an AUC of 0.689 (95% CI: 0.610-0.768) (Figure 2). A cutoff value of 2 for the CONUT score yielded an AUC of 0.664 (95% CI: 0.588-0.740), with 68.5% sensitivity, 64.3% specificity, 19.0% positive predictive value, and 94.4% negative predictive value (Table 7).

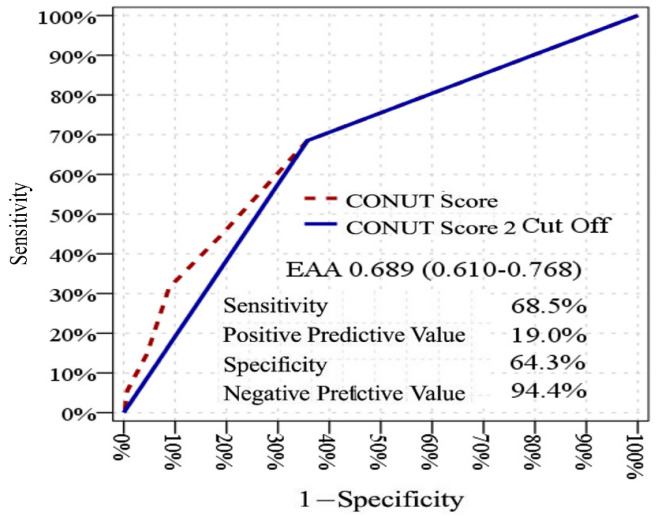

Figure 2. ROC curve

CONUT: Controlling Nutritional Status, ROC: Receiver operating characteristic

Table 7. ROC analysis of CONUT score for mortality

		AUC		95% CI	p
CONUT score		0.689		0.610-0.768	0.000
CONUT score 2 cut off		0.664		0.588-0.740	0.000
		Survived (-)	Deceased (+)		%
CONUT score	<2	284	17	Sensitivity	68.5%
	≥2	158	37	Positive predictive value	19.0%
				Specificity	64.3%
				Negative predictive value	94.4%

ROC curve, ROC: Receiver operating characteristic, CONUT: Controlling Nutritional Status, AUC: Area under the curve, CI: Confidence interval

DISCUSSION

This study examined the prognostic importance of inflammation-based indices and nutrition-related composite indices such as MHR, SII, PNI, and the CONUT score in predicting intensive care unit (ICU) admission and one-year mortality post- AIS. Our results underpin the clinical significance of these markers, which can be obtained from routine laboratory tests, and bolster the argument towards their proactive use in standard protocols for stratifying stroke risk.

The marked relationship already established within our cohort between elevated SII levels and both the need for ICU care, as well as one-year mortality, has previously been noted concerning SII's reputation as a powerful marker of immune activation and adverse prognosis. For example, Yang et al.[5] reported the association of elevated SII with the occurrence of adverse cardiovascular events, an observation that was later confirmed by Ye et al.[8] in a systematic review and meta-analysis of cardiovascular disease populations. In addition, Xue et al.[13] reported in a cross-sectional analysis of NHANES data that higher SII values increased the risk of stroke. We have shown that SII, as an index which combines neutrophil, platelet, and lymphocyte counts, provides a measure of systemic inflammation, integrates predictive capability regarding the severity of AIS and its long-term outcomes.

In relation to the MHR, more speculative work by Ganjali et al.[6] and Mi et al.[7] proposed MHR as a predictor of cardiovascular and inflammatory disorders such as gout. Our study, however, attempts to further refine these observations by examining their differential prognostic impact in AIS. Although univariate analysis showed no significant difference in MHR between ICU and non-ICU patients, multivariate analysis identified MHR as an independent predictor of ICU admission (p=0.036) and one-year mortality (p=0.007). This underlines the importance of multivariate modeling to account for confounding effects. This discrepancy might be due to the more chronic MHR-related pathophysiological processes that are systemically conditioned rather than acute. Such results align with those of Deng et al.[14] who reported that some immunonutritional markers, including MHR, were associated with long-term mortality risk in patients after stroke.

The influence of a patient's nutritional status concerning certain clinical outcomes in AIS patients has received greater scrutiny, particularly the PNI and CONUT scores which have been analyzed in diverse populations. As pointed out by Nozoe et al.[9] who validated the prognostic significance of PNI in colorectal carcinoma and subsequently expanded by Ho et al.[12] to hepatocellular carcinoma, our findings that lower PNI and higher CONUT scores were associated with worse

clinical outcomes also converge with these prior findings. In the stroke literature, more pertinent to our focus, Pan et al.[15] demonstrated lower PNI significantly increased the likelihood of stroke and worsened outcomes in a representative U.S. cohort. Huang et al.[16] proposed the AIS datasets with inflammation and nutrition variables could better integrate stroke prognosis with a Derived Inflammation-Nutrition Index sparking further conversation. This study further emphasizes the need for holistic models of stroke prognosis while integrating inflammatory as well as nutritional variables along with clinical factors.

The comparison of inflammatory and nutritional indices as predictors of outcome in the context of multivariable logistic regression is what sets this study apart from previous work. Unlike other studies that focus on individual markers, our study seeks to determine which of these scores-their confounder-adjusted values-serve as the most independent predictors for ICU admission and mortality within one year. We found that inflammatory markers, SII, NLR, and MLR, are outperformed by nutritional indices, especially the CONUT score, across both endpoints. This reinforces the notion that although systemic inflammation is a key component of stroke pathophysiology, malnutrition in this context denotes a greater susceptibility to prolonged compromised states that are more intimately associated with poor prognosis among patients with AIS. In this regard, the approach taken in this study improves prognostic stratification and highlights the clinical relevance of each index.

The CONUT score, specifically, demonstrated the strongest prognostic capability in our cohort analysis outperforming other indices in ROC curve analysis for both ICU admission and one year mortality. This reaffirms the findings of Song et al.,[17] who demonstrated that CONUT, alongside other inflammatory markers, predicted adverse outcomes in patients with acute coronary syndrome undergoing percutaneous coronary intervention remarkably well. Its ability to stratify malnutrition risk based on serum albumin, total cholesterol, and lymphocyte count makes CONUT an especially practical and integrative tool for system-wide evaluations, including those conducted during hospitalization in stroke units.[18]

In contrast to earlier studies that focused on short-term or long-term outcomes individually, our study takes a more holistic approach. It is comprehensive and comparative in its analysis. Given the well-defined AIS cohort, we were able to assess multiple composite markers simultaneously in addition to employing univariate and multivariate modeling to establish a hierarchy of prognostic indicators. Additionally, the focus on ICU admission as a marker of acute severity and one-year mortality as a long-term outcome deepens the clinical relevance of our results.[19]

Our findings suggest that integrating inflammatory and nutritional evaluation may offer a clearer insight into patient vulnerability, informing both acute care and strategies for managing the post-discharge period. For instance, the identification of patients with high CONUT scores along with elevated SII values could enable proactive monitoring and tailored interventions that enhance survival and functional recovery.[20]

Regardless, some important limitations should be noted. Even with multivariate adjustments, the retrospective design is still susceptible to selection bias and unmeasured confounding factors, which remain difficult to account for and control in single-center studies. Consequently, the external validity of our findings is limited, necessitating multi-center, prospective validation studies to establish broader applicability. Lastly, while our indices are based upon easily obtainable laboratory parameters, they did not capture time-dependent dynamics which could illuminate temporal patterns and treatment response.

As a concluding remark, our study reinforces the prognostic significance of SII, MHR, PNI, and particularly the CONUT score in patients with AIS. These markers, both individually and in combination, enhance predictive accuracy regarding ICU admission and long-term mortality risk. Further investigation is warranted to confirm these findings in larger patient populations and evaluate their incorporation into clinical decision-support frameworks for targeted, tailored stroke management.

Future studies should adopt prospective, multi-center designs to validate these findings across diverse populations. Integrating dynamic (serial) measurements of inflammatory and nutritional indices, along with detailed clinical parameters such as thrombolysis/thrombectomy status, lesion characteristics, and rehabilitation outcomes, could improve prognostic modeling. Additionally, incorporating quality-of-life and disability metrics would expand the clinical applicability of these indices beyond mortality prediction.

Limitations

The generalizability of the findings is constrained by the retrospective and single-center design of the study. Inflammatory and nutritional indices were only evaluated upon admission, and longitudinal changes were not analyzed. Furthermore, rehabilitation outcomes related to disability, as well as health-related quality of life measures, were also omitted from the assessment. Since information on thrombolytic treatment, lesion localization and volume were not systematically recorded in retrospective file data, these variables could not be included in the analysis. Since information on thrombolytic therapy, lesion localization and volume were not systematically recorded in retrospective file data, these variables could not be included in the analysis. However, our study has shown that mortality prediction can be made with biochemical parameters that can be easily obtained in the early period, independent of such clinical details. In this respect, it offers a practical contribution, especially in time-limited clinical environments such as emergency departments or intensive care units. Additionally, important clinical variables such as thrombolytic treatment status, mechanical thrombectomy status, lesion localization and volume, stroke severity scores (e.g., NIHSS), and functional outcome measures (mRS, Barthel Index) were not included due to incomplete data in the retrospective records. The absence of these factors may limit the ability to fully adjust for confounders. Rehabilitation outcomes and quality-of-life metrics were also not assessed, which narrows the scope of long-term prognostic implications.

CONCLUSION

As a result, this study demonstrates that inflammation- and nutrition-based composite indices, particularly the CONUT score, SII, and MHR, are valuable prognostic tools for assessing the risk of ICU admission and one-year mortality in patients with acute ischemic stroke. These biomarkers, derived from routine blood tests, offer a practical, low-cost, and effective method for early risk stratification. The findings suggest that combining immunological and nutritional assessments at admission may enhance prognostic precision and guide clinical decision-making. Further prospective, multicenter studies are recommended to validate these results and explore the integration of these indices into stroke care algorithms and discharge planning strategies.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was carried out with the permission of the Sancaktepe Şehit Prof. Dr. İlhan Varank Training and Research Hospital (Date: 19.12.2023, Decision No: 2023/257).

Informed Consent

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

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

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

Author Contributions

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

Acknowledgment

We thank who voluntarily participated in the study.

Data Availability Statement

The original contributions presented in the article; further inquiries can be directed to the corresponding author.

REFERENCES

1. Herpich F, Rincon F. Management of acute ischemic stroke. *Crit Care Med*. 2020;48(11):1654-1663. doi:10.1097/CCM.0000000000004597
2. Powers WJ. Acute ischemic stroke. *N Engl J Med*. 2020;383(3):252-260. doi:10.1056/NEJMcp1917030
3. González RG, Hirsch JA, Koroshetz WJ, Lev MH, Schaefer PW, editors. Acute ischemic stroke. Berlin: Springer-Verlag; 2011.
4. Zhong JH, Huang DH, Chen ZY. Prognostic role of Systemic Immune-Inflammation Index in solid tumors: a systematic review and meta-analysis. *Oncotarget*. 2017;8(43):75381-75388. doi:10.18632/oncotarget.18856
5. Yang YL, Wu CH, Hsu PF, et al. Systemic Immune-Inflammation Index (SII) predicted clinical outcome in patients with coronary artery disease. *Eur J Clin Invest*. 2020;50(5):e13230. doi:10.1111/eci.13230
6. Ganjali S, Gotto AM Jr, Ruscica M, et al. Monocyte-to-HDL-cholesterol ratio as a prognostic marker in cardiovascular diseases. *J Cell Physiol*. 2018;233(12):9237-9246. doi:10.1002/jcp.27028
7. Mi L, He X, Gao J, Xu K. Monocyte-to-HDL cholesterol ratio (MHR) as a novel Indicator of gout risk. *Sci Rep*. 2025;15(1):12188. doi:10.1038/s41598-025-97373-w
8. Ye Z, Hu T, Wang J, et al. Systemic Immune-Inflammation Index as a potential biomarker of cardiovascular diseases: a systematic review and meta-analysis. *Front Cardiovasc Med*. 2022;9:933913. doi:10.3389/fcvm.2022.933913
9. Nozoe T, Kohno M, Iguchi T, et al. The prognostic nutritional index can be a prognostic indicator in colorectal carcinoma. *Surg Today*. 2012;42(6):532-535. doi:10.1007/s00595-011-0061-0
10. Buzby GP, Mullen JL, Matthews DC, Hobbs CL, Rosato EF. Prognostic Nutritional Index in gastrointestinal surgery. *Am J Surg*. 1980;139(1):160-167. doi:10.1016/0002-9610(80)90246-9
11. Zheng Y, Wang K, Ou Y, et al. Prognostic value of a baseline Prognostic Nutritional Index for patients with prostate cancer: a systematic review and meta-analysis. *Prostate Cancer Prostatic Dis*. 2024;27(4):604-613. doi:10.1038/s41391-023-00689-9
12. Ho CT, Chia-Hui Tan E, Lee PC, et al. Prognostic Nutritional Index as a prognostic factor for very early-stage hepatocellular carcinoma. *Clin Transl Gastroenterol*. 2024;15(4):e00678. doi:10.14309/ctg.0000000000000678
13. Xue H, Zeng Y, Zou X, Li Y. Systemic immune inflammation index and risk of stroke: a cross-sectional study of the National Health and Nutrition Examination Survey 2005-2018. *Front Neurol*. 2024;15:1431727. doi:10.3389/fneur.2024.1431727
14. Deng K, Feng S, Cheng F, et al. Association of immunonutritional indicators with all-cause mortality in adult stroke patients. *Sci Rep*. 2025;15(1):14762. doi:10.1038/s41598-025-99158-7
15. Pan L, Peng Y, Jiang L. Association between prognostic nutritional index and stroke: a nationally representative cross-sectional study from NHANES. *J Stroke Cerebrovasc Dis*. 2025;34(1):108165. doi:10.1016/j.jstrokecerebrovasdis.2024.108165
16. Huang Y, Wang X, Li Z, Yin X. A novel Nutritional Inflammation Index for predicting mortality in acute ischemic stroke patients: insights into Advanced Lung Cancer Inflammation Index from the Medical Information mart for intensive care-IV database. *Front Nutr*. 2024;11:1408372. doi:10.3389/fnut.2024.1408372
17. Song G, Wang X, Wei C, et al. The complex inflammatory and nutritional indices to predict prognostic risk for patients with acute coronary syndrome undergoing percutaneous coronary intervention. *Immun Inflamm Dis*. 2025;13(3):e70180. doi:10.1002/iid3.70180
18. Zhang Y, Jiang M, Gao Y, et al. "No-reflow" phenomenon in acute ischemic stroke. *J Cereb Blood Flow Metab*. 2024;44(1):19-37. doi:10.1177/0271678X231208476
19. Pacchiano F, Tortora M, Criscuolo S, et al. Artificial intelligence applied in acute ischemic stroke: from child to elderly. *Radiol Med*. 2024;129(1):83-92. doi:10.1007/s11547-023-01735-1
20. Liu L, Zhao YB, Cheng ZT, Li YH, Liu Y. Development and validation of a prognostic model for critically ill type 2 diabetes patients in ICU based on composite inflammatory indicators. *Sci Rep*. 2025;15(1):3627. doi:10.1038/s41598-025-87731-z

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1721193

J Health Sci Med. 2025;8(5):909-913

Diagnostic and therapeutic approaches of pediatricians to scabies: a nationwide cross-sectional study from Turkiye

İlkay Can[1], Ahmet Soysal[2], Arzu Kılıç[1], Pelin Hızlı[1], Jale Aylin Akdaşli[1]

[1]Department of Dermatology, Faculty of Medicine, Balıkesir University, Balıkesir, Turkiye

[2]Department of Pediatric Infectious Diseases, Memorial Ataşehir Hospital, İstanbul, Turkiye

Cite this article as: Can İ, Soysal A, Kılıç A, Hızlı P, Akdaşli JA. Diagnostic and therapeutic approaches of pediatricians to scabies: a nationwide cross-sectional study from Turkiye. *J Health Sci Med.* 2025;8(5):909-913.

Received: 17.06.2025 ◆ **Accepted**: 01.09.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: Scabies is a highly contagious ectoparasitic infestation commonly observed in childhood. Its diagnosis and treatment can be challenging due to atypical clinical presentations that overlap with other dermatological conditions. This study aims to evaluate the diagnostic and therapeutic approaches of pediatricians toward scabies and highlight the variabilities in their clinical practices.

Methods: Online survey forms were distributed to pediatric specialists and residents practicing across Turkiye. The questionnaire included items on demographic characteristics, clinical approaches, treatment preferences, and management of close contacts.

Results: A total of 459 pediatricians participated in the study. The most commonly preferred first-line treatment agent was permethrin lotion (60.3%), while the use of ivermectin was limited (1.3%). In addition, 64.5% of the participants reported prescribing magistral preparations in their daily practice, and 90% stated that they recommended treatment for family members. Clinical findings were most frequently used for diagnosis (96.5%), whereas laboratory confirmation was rarely requested. Approximately one-quarter of the participants believed that scabies could also be transmitted from animals. These results demonstrate considerable variability in knowledge and clinical practice.

Conclusion: This study has revealed notable variabilities among pediatricians in the diagnostic and therapeutic approaches used to treat scabies. Practical challenges in the clinical setting underscore the need for improved integration of current guidelines and implementation of targeted educational programs.

Keywords: Scabies, pediatrics, treatment

INTRODUCTION

Scabies (*Sarcoptes scabiei* var. hominis) is a contagious ectoparasitic infestation of humans characterized by itching and inflammatory skin lesions that worsen especially at night. In recent years, there has been an increase in cases of scabies worldwide even in developed countries. In 2017, the World Health Organization (WHO) included scabies in the Neglected Tropical Diseases (NTDs) group, demonstrating the seriousness of the situation.[1] Scabies is very common not only in adults but also in pediatric patients. Communal areas such as kindergartens and schools where children are populated make the risk of transmission easier. Unlike adults, we can see its atypical presentations more frequently in children. Among these, the involvement of areas such as palms, soles and scalp, eczematous appearance of the lesions and accompanying secondary infections cause difficulties in diagnosis. Delays in diagnosis adversely affect both the patient's quality of life and also public health by leading to chain infestations due to the increased risk of transmission.[2,3]

Pediatricians play a multifaceted role in the management of scabies, yet they often face challenges including diagnostic difficulties, treatment resistance, non-compliance of families, and contact management.[4] Despite the diagnosis and treatment protocols specified in scabies management guidelines, there are significant differences in daily practices of pediatricians. This is due not only to a lack of direct knowledge but also to difficulties encountered in field conditions.[4,5] There are many epidemiological studies on pediatric scabies cases in the literature. However, field-based studies directly examining the clinical approach of pediatricians are quite limited.[6-8]

To the best of our knowledge, this is one of the first nationwide questionnaire-based studies specifically designed to evaluate the diagnostic and therapeutic approaches of pediatricians toward scabies in Turkiye, which emphasizes its originality and contribution to the literature.

With this study, we have aimed to evaluate the level of knowledge, attitudes and practices of pediatricians and residents actively working in the management of scabies in Turkiye. We believe that the data we obtained from this study will shed light on the treatment guidelines and training programs to be developed in the future

Corresponding Author: İlkay Can, ilkaycan@balikesir.edu.tr

METHODS

This is a descriptive and cross-sectional study designed to evaluate the attitudes and practices of 459 pediatricians actively working to establish the diagnosis and treatment of scabies in Turkiye. The ethics committee approval of the study was obtained from the Memorial Ataşehir Hospital Ethics Committee (Date: 13.05.2025, Decision No: 2025/16) and the study was conducted based on the ethical principles of the WMA Declaration of Helsinki.

The questionnaire was developed by the research team to collect clinical and demographic information on pediatricians' approaches to scabies, and therefore no validity or reliability analysis was performed. The data used in the study were collected through a questionnaire form developed by the researchers on the Google Forms platform. The questionnaire collected information from pediatric physicians currently working in Turkiye. Participating physicians were reached through various professional communication channels (telegrams, e-mail groups, social media, WhatsApp, etc.) and online access was provided. Informed consent was obtained from the participants. Since this study was designed as a descriptive, nationwide survey, all accessible pediatricians were invited to participate. Therefore, no formal sample size calculation was performed, and the final sample size reflects voluntary participation.

The questionnaire forms contain basic data such as demographic information of physicians (age, gender, duration of professional experience, institution, academic title, geographical region), total number of referrals per day and number of patients presenting with itching complaints. In addition, clinical findings of the patients suggesting the diagnosis of scabies, use of diagnostic laboratory tests, treatment preferences (initial treatment approach, prescribing of magistral medications, recommended treatment by family members, repeat treatment), calling patients for follow-up visits, and seasonal awareness about scabies were evaluated. In addition, patients' knowledge level about scabies and the content of information given to the patients were also questioned.

Pediatricians, residents or subspecialists actively practicing in Turkiye were included in the study. Patients with missing data, and non-pediatric physicians were excluded from the study.

Statistical Analysis

The research data were analyzed using SPSS (Statistical Package for the Social Sciences for Windows, version 26.0; SPSS Inc., Chicago, IL). Descriptive statistics were presented as frequency distributions and percentages. Categorical variables were evaluated using the Chi-square test. A p-value of <0.05 was considered statistically significant.

RESULTS

A total of 459 pediatricians from different regions of Turkiye participated in the study. Participants showed a wide distribution in terms of age, gender, years of professional experience, institutional setting, and academic title. The largest proportion were in the 30-39 age group, and female physicians were slightly predominant. About one-third of the physicians reported more than 20 years of professional experience. Nearly half were working in public hospitals, while the remainder were employed in private or university hospitals. The majority held specialist titles. Geographically, participation was highest from the Marmara region, followed by Central Anatolia and the Aegean region (Table 1).

Table 1. Demographic characteristics of participating pediatricians (n=459)

n: 459		**n (%)**
Age (years)	20-29	24 (5.2)
	30-39	207 (45.1)
	40-49	133 (29.0)
	50-59	95 (20.7)
Gender	Female	254 (55.3)
	Male	205 (44.7)
Professional experience (years)	1-4	31 (6.8)
	5-9	111 (24.2)
	10-14	103 (22.4)
	15-19	70 (15.3)
	20<	144 (31.4)
Employed institution	University	65 (14.2)
	Private hospital	168 (36.6)
	State hospital	226 (49.2)
Academic title of the participants	Assistant doctor	38 (8.3)
	Academician	42 (9.2)
	Specialist	379 (82.6)
Geographical regions of the study participants	Marmara region	198 (43.1)
	Aegean region	53 (11.5)
	Mediterranean region	37 (8.1)
	Black Sea region	38 (8.3)
	Central Anatolia region	53 (11.5)
	Southeastern Anatolia region	55 (12.0)
	Eastern Anatolia region	25 (5.4)

Most physicians reported examining 15-59 patients per day, and two-thirds indicated that 1-4 patients presented daily with itching complaints (Table 2).

In terms of treatment approaches, permethrin lotion was the most commonly prescribed agent, followed by magistral preparations and sulfur cream. Family treatment was recommended by 90% of participants, 59.9% scheduled follow-up visits, and 64.1% advised repeat treatment within 7-15 days. A total of 29.2% of physicians reported a seasonal increase in cases during winter. However, the largest proportion of respondents reported no seasonal difference (38.8%). When the literature is examined, it is seen that incidence rates of scabies increase in winter.[11,12] Table 3 summarizes the data on treatment approaches.

Table 2. Daily patient density of pediatricians and number of patients presenting with itching complaints

n: 459		n (%)
Average number of patients cared for per day	0-14	68 (14.8)
	15-24	90 (19.6)
	25-34	58 (12.6)
	35-44	63 (13.7)
	45-59	84 (18.3)
	60-74	58 (12.6)
	75-89	23 (5.0)
	90<	15 (3.3)
Number of patients presenting daily with itching complaints	1-4	313 (68.2)
	5-9	122 (26.6)
	10-14	19 (4.1)
	15-19	3 (0.7)
	20<	2 (0.4)

Table 3. Treatment approaches used by pediatricians to manage scabies

n: 459		n (%)
Initial recommendation for drug therapy	Permethrin lotion	277 (60.3)
	Sulfur cream	55 (12.0)
	Magistral drugs	39 (8.5)
	Ivermectin tablets	6 (1.3)
	No treatment recommended	65 (14.2)
	Ready-made magistral drugs	17 (3.7)
Prescribing magistral medications	Yes	296 (64.5)
	No	163 (35.5)
Magistral drugs used	Sulfur	96 (20.9)
	Balsam of Peru	75 (16.3)
	Benzyl benzoate	9 (2.0)
	Other	116 (25.3)
Recommending family treatment	Yes	413 (90.0)
	No	46 (10.0)
Status of calling patients for follow-up	Yes	275 (59.9)
	No	184 (40.1)
Repeat treatment within 7-15 days	Yes	294 (64.1)
	No	165 (35.9)
Seasonal differences	Spring	35 (7.6)
	Summer	48 (10.5)
	Autumn	64 (13.9)
	Winter	134 (29.2)
	No	178 (38.8)

With respect to diagnosis, 96.5% of physicians relied on clinical findings without laboratory confirmation. In addition, 23.7% believed that scabies could be transmitted from animals (Table 4).

Subgroup analyses according to years of professional experience revealed significant differences. Physicians with <10 years of experience prescribed permethrin lotion more frequently compared to those with ≥10 years (68.3% vs. 56.8%; p=0.029). Conversely, the rate of not recommending any treatment was higher among physicians with ≥10 years of experience (17.7% vs. 6.3%). Belief in animal-to-human transmission was also more common among less experienced physicians (45.3% vs. 32.8%; p=0.012). No significant difference was observed between the two groups in terms of recommending family treatment (p=0.228). These findings are presented in Table 5.

Table 4. Findings on diagnosis and level of knowledge

n:459		n (%)
Laboratory test required	Yes	16 (3.5)
	No	443 (96.5)
Diagnostic examination findings	No benefit from treatment and diagnosis is doubtful	439 (95.6)
	Erythematous papule, vesicle	20 (4.4)
Transmission of scabies from animals	Yes	109 (23.7)
	No	350 (76.3)

Table 5. Treatment preferences of pediatricians by years of professional experience

n: 459	Experience <10 years n (%)	Experience ≥10 years n (%)	p*
Treatment			
Permethrin lotion	97 (68.3)	180 (56.8)	0.029
Sulfur cream	14 (9.9)	41 (12.9)	
Magistral drugs	14 (9.9)	25 (7.9)	
Ivermectin tablets	2 (1.4)	4 (1.3)	
No treatment recommended	9 (6.3)	56 (17.7)	
Ready-made magistral drugs	6 (4.2)	11 (3.5%)	
Belief in animal-to-human transmission			
Yes	64 (45.3)	104 (32.8)	0.012
No	78 (54.9)	213 (67.2)	
Family treatment			
Yes	138 (97.2)	300 (94.4)	0.228
No	4 (2.8)	17 (5.4)	

*: Chi-square test

DISCUSSION

Scabies has an increasing prevalence not only in Turkiye but also all over the world.[9,10] Although the reason for the recent increase in the number of cases is not clearly known, deficiencies in diagnosis and treatment contribute to higher rates of scabies. Whereas non-responsiveness to medications on a case-by-case basis may have contributed to increased prevalence of scabies.[11-13]

To the best of our knowledge, this is one of the few nationwide field-based studies evaluating the approaches of pediatricians working throughout Turkiye towards the diagnosis and treatment of scabies.[14,15] Although the data obtained show that a significant portion of physicians exhibit an approach to these patients in parallel with the current guidelines, it is

noteworthy that differences in knowledge and practice among physicians persist in some areas.

The fact that permethrin lotion (60.3%), which is the most preferred agent in the treatment of scabies, stands out in the first line, shows a trend that is generally in line with national and international guidelines.[16,17] However, the limited use of systemic agents such as ivermectin (1.3%) may reflect factors such as availability, physician experience, or confidence in this treatment option, although these aspects were not directly evaluated in our study. In addition, the high rates of prescribing of medicinal products (64.5%) and the diversity of ingredients (e.g. balsam of Peru, benzyl benzoate) indicate the heterogeneity of practice in the field, which suggests that standardization of scabies treatment remains a challenge.

The high rate (90.0%) of recommendation of family treatment indicates that physicians are aware of the importance of contact management in breaking the chain of transmission. However, the variabilities in the rates of calling patients for follow-up (59.9%) and recommending repeat treatment (64.1%) suggest that the guideline recommendations are not uniformly adopted by all physicians. This finding is particularly noteworthy for the management of resistant or recurrent cases. The importance of repeat treatment in scabies is emphasized in many international guidelines. Reapplication of topical treatments and systemic ivermectin therapy is recommended after 7-14 days. This approach aims to eliminate both mites that survived the initial treatment or newly hatched eggs.[16,17] This variability in follow-up and re-treatment strategies may also reflect institutional or regional differences. Previous studies have reported that suboptimal application of topical therapy, poor adherence, and lack of systematic follow-up are common barriers to effective management, particularly in high-burden or resource-limited settings.[7,18] These findings underline the need for standardized approaches to follow-up and re-treatment across different healthcare institutions in Turkiye.

Diagnostic findings revealed that the vast majority of pediatricians made the diagnosis by clinical observation (96.5%) and laboratory tests were rarely used. Although this failure to resort to diagnostic laboratory tests may be related to the need for rapid diagnosis in practice, it may also bring the risk of misdiagnosis or delayed diagnosis in atypical cases. Considering the atypical presentations that may be seen especially in pediatric patients, it can be said that supportive diagnostic methods need to be used more effectively in appropriate cases.

Another noteworthy point in the study is that approximately one fourth of the physicians (23.7%) thought that scabies could be transmitted from animals. This lack of knowledge indicates a widespread misconception and suggests that professional trainings should focus not only on treatment but also on transmission routes and infection control. Based on literature findings, humans can rarely contract scabies from domestic dogs (*S. scabiei var canis*) and cats (*Notoedres cati*). However, zoonotic scabies differs from classical scabies in that it has a short incubation period, the distribution of lesions is limited to areas of contact with the animal and tunnels are not seen. Animal mites cannot reproduce in human hosts and can only survive for a few days.[19] As a result, human-to-human transmission does not occur and only treatment of the animals is sufficient for cure.[20,21]

Previous studies conducted in Turkiye and other countries also provide valuable insights into the management of scabies and support the findings of the present study. Özçelik[14] retrospectively analyzed pediatric scabies cases and emphasized the diagnostic difficulties and distinct clinical presentations in children, including palmoplantar and scalp involvement. Similarly, Etgu and Önder[15] evaluated the knowledge levels of primary care physicians and demonstrated significant gaps in awareness regarding transmission and treatment compliance, underlining the need for targeted educational programs. In addition, Oba et al.[22] investigated factors influencing treatment success in pediatric scabies and highlighted the importance of family treatment and correct application of topical agents. On an international level, Romani et al.[7] and Bernigaud et al.[18] reported that inadequate adherence to guidelines, lack of systematic follow-up, and treatment resistance are common barriers in scabies control. These findings indicate that the challenges identified in our study are also reported both nationally and globally, reinforcing the need for standardized management strategies.

The findings of our study also highlight the need for structured physician training in the management of scabies. Previous reports have indicated that inconsistencies in adherence to treatment and follow-up recommendations, as well as suboptimal practices, may adversely affect patient outcomes.[18] International guidelines emphasize the importance of re-treatment intervals, contact management, and the use of supportive diagnostic methods in atypical cases.[16,23] Incorporating these aspects into continuing medical education programs and developing updated national guidelines harmonized with international standards would be beneficial. Such efforts could help address knowledge gaps, correct misconceptions, and promote standardization in clinical practice.

Limitations

This study has several limitations. First, participation was based on voluntary recruitment, which may have led to an over-representation of certain physician profiles or geographical regions and therefore limits the generalizability of the findings. Second, data were obtained through self-report questionnaires, which carry an inherent risk of recall bias and socially desirable responding. Third, the recruitment of participants through online channels such as social media and WhatsApp groups may have introduced sampling bias, as physicians less engaged in such platforms could be underrepresented. Despite these limitations, the study provides valuable insights into the current diagnostic and therapeutic practices of pediatricians in Turkiye and highlights important areas for improvement.

CONCLUSION

As a result, this nationwide survey highlights significant variability in the diagnostic and therapeutic approaches of pediatricians toward scabies in Turkiye. While permethrin lotion remains the most preferred treatment in line with

national and international guidelines, gaps persist in the use of systemic therapies, re-treatment practices, follow-up strategies, and knowledge about transmission. The study underscores the importance of standardized approaches, improved integration of current guidelines into clinical practice, and targeted educational initiatives such as continuing medical education and updated national guidelines. Addressing these gaps will be critical for optimizing scabies management in pediatric practice and reducing the overall disease burden.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was conducted with the approval of the Memorial Ataşehir Hospital Ethics Committee (Date: 13.05.2025, Decision No: 2025/16).

Informed Consent

All participating physicians were informed about the study and provided voluntary informed consent before participation.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors declare that they have no conflicts of interest.

Financial Disclosure

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

Author Contributions

All authors declare that they participated in the conception and design of the study, data collection, analysis, and interpretation, drafting and revising of the manuscript, and approved the final version of the article.

REFERENCES

1. Chosidow O, Hay RJ. Control of scabies and secondary impetigo: optimising treatment effectiveness in endemic settings. *Lancet Infect Dis*. 2019;19(5):454-456. doi:10.1016/s1473-3099(19)30068-4
2. Korycinska J, Dzika E, Kloch M. Epidemiology of scabies in relation to socio-economic and selected climatic factors in north-east Poland. *Ann Agric Environ Med*. 2020;27(3):374-378. doi:10.26444/aaem/109319
3. Bassi A, Magnano M, Filippeschi C, et al. Tuscany consensus for pediatric scabies: update on clinical management and the urgent need for new guidelines. *Ital J Dermatol Venerol*. 2025;160(1):12-19. doi:10.23736/s2784-8671.24.07991-x
4. Thompson R, Westbury S, Slape D. Paediatrics: how to manage scabies. *Drugs Context*. 2021;10:12-13. doi:10.7573/dic.2020-12-3
5. Ferhatosmanoğlu A, Baykal Selçuk L, Arıca İE. Risk factors and lesion patterns in treatment-resistant scabies: impact of sex, age, and comorbidities. *Acta Parasitol*. 2025;70(3):116. doi:10.1007/s11686-025-01055-6
6. Matthews A, Le B, Amaral S, et al. Prevalence of scabies and impetigo in school-age children in Timor-Leste. *Parasit Vectors*. 2021;14(1):156. doi:10.1186/s13071-021-04645-1
7. Romani L, Steer AC, Whitfeld MJ, Kaldor JM. Prevalence of scabies and impetigo worldwide: a systematic review. *Lancet Infect Dis*. 2015;15(8):960-967. doi:10.1016/s1473-3099(15)00132-2
8. Osti MH, Sokana O, Phelan S, et al. Prevalence of scabies and impetigo in the Solomon Islands: a school survey. *BMC Infect Dis*. 2019;19(1):803. doi:10.1186/s12879-019-4382-8
9. Etiz P, Altunsu AT. The scabies epidemic during the COVID-19 pandemic. *Turkiye Parazitol Derg*. 2023;47(4):235-239. doi:10.4274/tpd.galenos.2023.44154
10. Gupta S, Thornley S, Morris A, Sundborn G, Grant C. Prevalence and determinants of scabies: a global systematic review and meta-analysis. *Trop Med Int Health*. 2024;29(12):1006-1017. doi:10.1111/tmi.14058
11. Mbuagbaw L, Sadeghirad B, Morgan RL, et al. Failure of scabies treatment: a systematic review and meta-analysis. *Br J Dermatol*. 2024;190(2):163-173. doi:10.1093/bjd/ljad308
12. Azzolina V, Schauer F, Pilz JF, Zink A, Eyerich K, Pilz AC. Scabies management outcomes: identification of risk factors for treatment success or failure. *Dermatol Pract Concept*. 2025;15(2):5077. doi:10.5826/dpc.1502a5077
13. Aydemir S, Kul Z, Barlik F, Saygin M, Ekici A, Yilmaz H. Increase in scabies cases and permethrin nonresponse in Turkiye. *Indian J Public Health*. 2025;69(1):63-66. doi:10.4103/ijph.ijph_137_24
14. Özçelik S. A neglected disease: scabies a retrospective study on children. *Turk Çocuk Hast Derg*. 2022;16(2):121-126.
15. Etgu F, Onder S. Evaluating knowledge level about scabies in primary care physicians during the scabies outbreak of Turkey. *Turk J Dermatol*. 2023;17(2):57-63.
16. Salavastru CM, Chosidow O, Boffa MJ, Janier M, Tiplica GS. European guideline for the management of scabies. *J Eur Acad Dermatol Venereol*. 2017;31(8):1248-1253. doi:10.1111/jdv.14351
17. Uzun S, Durdu M, Yürekli A, et al. Clinical practice guidelines for the diagnosis and treatment of scabies. *Int J Dermatol*. 2024;63(12):1642-1656. doi:10.1111/ijd.17327
18. Bernigaud C, Fischer K, Chosidow O. The management of scabies in the 21st century: past, advances and potentials. *Acta Derm Venereol*. 2020;100(9):adv00112. doi:10.2340/00015555-3468
19. Heukelbach J, Feldmeier H. Scabies. *Lancet*. 2006;367(9524):1767-1774. doi:10.1016/s0140-6736(06)68772-2
20. Aydıngöz IE, Mansur AT. Canine scabies in humans: a case report and review of the literature. *Dermatology*. 2011;223(2):104-106. doi:10.1159/000327378
21. Morris G, Haddow L, Sashidharan PN, et al. British Association for Sexual Health and HIV national guideline on the management of scabies in adults 2025. *Int J STD AIDS*. 2025;36(7):516-532. doi:10.1177/09564624251321264
22. Oba MC, Ozkoca D, Basara Sahin R, Kazan DS, Guldiken G, Kara Esen B. Factors affecting topical treatment success in pediatric scabies cases: a cross-sectional study. *Dermatol Pract Concept*. 2023;13(4):e2023307. doi:10.5826/dpc.1304a307
23. Engelman D, Yoshizumi J, Hay RJ, et al. The 2020 international alliance for the control of scabies consensus criteria for the diagnosis of scabies. *Br J Dermatol*. 2020;183(5):808-820. doi:10.1111/bjd.18943

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1762968
J Health Sci Med. 2025;8(5):914-918

Vancomycin vs teicoplanin antibiotic lock therapy for pediatric coagulase negative staphylococcal central line associated bloodstream infections

Zeynep Ergenç[1,2,3], Aylin Dizi Işık[1], Sevgi Aslan Tuncay[1], Seyhan Yılmaz[1], Pınar Canizci Erdemli[1], Eda Kepenekli[1,4]

[1]Division of Pediatric Infectious Diseases, Department of Pediatrics, Faculty of Medicine, Marmara University, İstanbul, Turkiye
[2]Department of Pediatric Infectious Diseases and Immunology, Great Ormond Street Hospital for Children, London, United Kingdom
[3]Centre for Neonatal and Paediatric Infection and Vaccine Institute, City St George's University of London, London, United Kingdom
[4]Division of Pediatric Infectious Diseases, Department of Pediatrics, Faculty of Medicine, Biruni University, İstanbul, Turkiye

Cite this article as: Ergenç Z, Dizi Işık A, Aslan Tuncay S, Yılmaz S, Canizci Erdemli P, Kepenekli E. Vancomycin vs teicoplanin antibiotic lock therapy for pediatric coagulase negative staphylococcal central line associated bloodstream infections. *J Health Sci Med.* 2025;8(5):914-918.

Received: 12.08.2025 ◆ **Accepted**: 01.09.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: Central line associated bloodstream infections (CLABSIs) are the major complication of central lines. Coagulase-negative *Staphylococcus* (CoNS) are the leading cause of CLABSI often necessitating line removal. While antibiotic lock therapy (ALT) is widely utilized in adults, pediatric specific data remains limited. In this study we aimed to evaluate the ALT outcomes in children with CoNS CLABSIs.

Methods: Children with CoNS CLABSIs who received ≥72 hours of ALT with either vancomycin or teicoplanin between January 2020-2023 were retrospectively reviewed. Data on demographic and clinical characteristics, management strategies, and outcomes were analyzed. ALT success was defined as clinical resolution, negative follow-up blood cultures, and catheter retention.

Results: Nineteen patients were included (median age 50 months; 58% female). Methicillin-sensitive CoNS (MSCoNS) were isolated in 4 (21%) cases and methicillin-resistant CoNS (MRCoNS) in 15 (79%). ALT regimens comprised vancomycin (n=9, 47%), teicoplanin (n=8, 42%), or sequential use (n=2, 11%). Overall, ALT success was 63% (12/19). Vancomycin-based ALT succeeded in 56% (5/9) vs. 62.5% (5/8) for teicoplanin (p=1.0). Success was higher in MSCoNS CLABSIs (100%) than MRCoNS (53%), though not statistically significant (p=0.245). Younger age (p=0.003) and persistent positive cultures (p=0.013) were associated with catheter loss. No infection-related mortality occurred. Recurrence occurred in 3 (16%) patients and reinfection in 5 (26%).

Conclusion: ALT achieved satisfactory catheter salvage rates in pediatric CoNS CLABSI, with comparable efficacy between vancomycin and teicoplanin. Younger age and persistent bacteremia predicted failure. Vancomycin or teicoplanin based ALT can be used in selected children with CoNS CLABSI.

Keywords: Catheter-related infections, coagulase, *Staphylococcus*, teicoplanin, vancomycin

INTRODUCTION

Central lines play an essential role in the care of hospitalized pediatric patients, facilitating the administration of medications, parenteral nutrition, and blood sampling over extended periods. However, bloodstream infections remain the most common complication, contributing significantly to morbidity, mortality, prolonged hospitalization, and increased healthcare costs.[1-4]

Coagulase-negative *Staphylococcus* (CoNS) are the leading causative organisms of central line associated bloodstream infection (CLABSI), largely due to their capacity to adhere to catheter surfaces and form biofilms.[4-7] These biofilms confer protection from host defences and systemic antimicrobial agents, rendering eradication difficult without catheter removal.[6,7]

Conventional management of CoNS CLABSIs often necessitates catheter removal, yet this is not always feasible in paediatric patients who rely on long-term vascular access.[8-10] In such cases, antibiotic lock therapy (ALT) has emerged as a valuable adjunctive strategy. ALT, involving the instillation of highly concentrated antimicrobial solutions into the catheter lumen, has long been utilized as a targeted strategy to treat intraluminal infections and preserve catheter function in adults.[8,11,12] ALT is typically used in combination with systemic antibiotics and aims to sterilize the catheter lumen,

Corresponding Author: Zeynep Ergenç, drzeynepergenc@gmail.com

disrupt biofilms, and allow catheter salvage, particularly in stable patients without signs of metastatic infection.[8,11,13] Despite growing clinical use, there is limited pediatric specific evidence to guide ALT protocols especially regarding optimal antimicrobial agents, dwell times, and treatment duration.[12-14]

In this study, we hypothesised that ALT with vancomycin or teicoplanin, when combined with parenteral therapy is an effective and safe strategy for managing CoNS CLABSIs in children by enabling catheter salvage and reducing line removal. The study retrospectively evaluated the clinical outcomes of paediatric patients treated with this approach.

METHODS

Study Design

The study was approved by the Marmara University Faculty of Medicine Clinical Researches Ethics Committee (Date: 07.04.2023, Decision No: 09.2023.621). All procedures were carried out in accordance with the ethical rules and the principles of the Declaration of Helsinki.

This retrospective study evaluated patients aged 1 months to 18 years with CoNS CLABSIs who received ALT between January 2020 and January 2023 at a tertiary hospital. Demographic, clinical and laboratory data were extracted from electronic medical records, including infection characteristics, management strategies, and treatment outcomes.

Clinical Management

According to the institutional policy, children with central lines and either peripheral or catheter-derived blood cultures positive for gram-positive organisms were empirically initiated with systemic vancomycin or teicoplanin. Following the detection of a positive blood culture, all cases were evaluated by the pediatric infectious diseases (PID) team, who determined whether the episode met criteria for CLABSI and made individualized decisions regarding the initiation of ALT. The choice of systemic antibiotic and lock solution was guided by the antimicrobial susceptibility profile of the isolated organism and clinical profile of the patient.

In line with hospital policy, C-reactive protein (CRP) levels were not routinely measured during febrile episodes in children receiving anti-cancer chemotherapy. Clinical assessment and complementary laboratory investigations were used to guide management, given that malignancy or chemotherapy can result in non-specific elevations in CRP, thereby limiting its diagnostic utility.

Patient Selection and Definitions

A central line associated bloodstream infection (CLABSI) was defined according to CDC guidelines as a primary bloodstream infection (BSI) occurring in a patient who had a central venous catheter (CVC) in place within the preceding 48 hours and in the absence of an identifiable source of infection elsewhere.[15]

Patients were eligible for inclusion if they met the following criteria: (1) confirmed CoNS CLABSI, (2) received systemic antimicrobial therapy with vancomycin or teicoplanin, and (3) received ALT for ≥72 hours.

Patients were excluded if they had Staphylococcal colonization without systemic infection signs, ALT administered for less than 72 hours, or polymicrobial infections. For children with multiple episodes of CoNS CLABSI during the study period, only the first episode was included in the analysis.

Treatment success was defined as clinical improvement and resolution of fever within 72 hours of ALT initiation with at least two consecutive sterile blood cultures obtained from the catheter, and retention of the long-term central line.[16]

Recurrence was defined as a new episode of CLABSI with the same pathogen ≥30 days after documented clearance. Reinfection was defined as a new CLABSI episode caused by a different organism following clinical resolution and sterile blood cultures.[17-19]

Antibiotic Lock Therapy (ALT)

ALT was administered in addition to the systemic therapy. Vancomycin lock solution was prepared by using vancomycin at 5 mg/ml with 2500 IU/ml heparin, and teicoplanin lock solution was prepared with teicoplanin at 10 mg/ml and 2500 IU/ml heparin according to local protocol. ALT was administered by instilling the antibiotic-heparin solution into the catheter lumen for at least 12 hours each day, and the solution was renewed every 24 hours. Patients who were on systemic vancomycin (60 mg/kg/day divided to four doses) therapy received vancomycin lock therapy while patients who were on systemic teicoplanin (10 mg/kg every 12 hours for three doses, then 10 mg/kg once daily) therapy received teicoplanin lock therapy.

Statistical Analysis

Data analyses were performed using SPSS Statistics Version 22.0 for Windows (IBM Corp., Armonk, NY). Categorical variables were presented as frequencies and percentages, while continuous variables were expressed as either mean±standard deviation or median with interquartile range, depending on the normality of their distribution. Continuous variables were compared using the independent-samples T test or the Mann-Whitney U test based on normality of distribution. The Pearson Chi-square test or Fisher exact test was used to compare categorical variables, with statistical significance set at $p<0.05$.

RESULTS

During the study period a total of 20 children with CoNS CLABSIs were treated with ALT. One patient with a polymicrobial infection involving methicillin resistant coagulase negative *Staphylococcus* (MRCoNS) and *Enterococcus faecium* infection was excluded. The final analysis included 19 patients.

Of these, 11 (58%) were female. The median age of patients was 50 months [interquartile range (IQR): 16-116 months]. Underlying conditions included solid organ tumours (n=4), hematologic malignancies (n=9), of whom four had undergone hematopoietic stem cell transplantation (HSCT), and rare comorbidities (n=6), such as I-cell disease, short bowel syndrome, autoimmune enteropathy, epileptic encephalopathy, trichohepatoenteric syndrome, and combined cystic fibrosis with osteogenesis imperfecta.

The median duration of catheter use prior to infection onset was 29.5 days (IQR: 17.75-79 days). Catheter types included 15 (79%) skin tunnelled central venous catheters, and four (21%) implanted port catheters.

Microbiological testing identified methicillin sensitive CoNS (MSCoNS) in 4 (21%) patients and methicillin resistant CoNS (MRCoNS) in 15 (79%) patients. Species level identification was available in 10 (52.6%) cases, comprising *Staphylococcus epidermidis* in 7 (36.8%), *Staphylococcus haemolyticus* in 2 (10.5%), and *Staphylococcus hominis* in 1 (5.3%).

At the time of infection, 7 (36.8%) patients were neutropenic. Among the non-neutropenic group (n=12, 63.2%), the median leukocyte count was 6.900/mm^3 (IQR: 2.225-13.550/mm^3), and the median absolute neutrophil count was 4.000/mm^3 (IQR: 1.500-7.600/mm^3). CRP levels were measured in 9 (47.3%) patients with a median value of 11 mg/L (IQR: 8.5-74.5 mg/L).

ALT regimens included teicoplanin in 8 (42.1%) patients, vancomycin in 9 (47.4%) patients, and sequential use of both agents in 2 (10.5%) patients. The median duration of ALT was 7 days (IQR: 5-8 days), while systemic intravenous antibiotics were administered for a median of 13 days (IQR: 13-14 days).

ALT was successful in 12/19 (63.2%) patients. Among those treated with vancomycin-based ALT, 5/9 (55.6%) patients achieved catheter salvage, while teicoplanin ALT was successful in 5/8 (62.5%) patients. The difference was not statistically significant. (p=1.0) When stratified by methicillin susceptibility, ALT was successful in all 4 cases (100%) of MSCoNS CLABSI, compared with 8 of 15 (53.3%) cases of MRCoNS CLABSI. The difference was not found statistically significant (p=0.245).

Younger age and a higher number of repeated positive blood cultures were found significantly associated with catheter loss (p=0.003 and p=0.013, respectively). Catheter instillation duration and CRP levels were not associated with catheter loss. (p=0.44 and p=0.90, respectively)

Subgroup analyses for underlying conditions and catheter types was not performed due to the limited sample size.

One patient with port catheter developed concurrent endocarditis. No infection related mortality occurred during the study period. Recurrence of CoNS CLABSI was observed in 3 (15.8%) patients with implanted ports and all underwent port catheter removal. Reinfection occurred in 5 (26.3%) patients, caused by *Rothia mucilaginosa* (n=1), *Ralstonia pickettii* (n=1), *Pseudomonas aeruginosa* (n=1), *Stenotrophomonas maltophilia* (n=1), and *Ralstonia insidiosa* (n=1).

DISCUSSION

In this study, ALT achieved an overall success rate of 63% in children with *CoNS* CLABSIs, allowing catheter salvage in the majority of cases. This finding is consistent with previously reported ALT success rates in children, which range from 50% to 80%, depending on pathogen type, underlying disease, and treatment protocol.[16,18-21]

We found no significant difference in success between vancomycin and teicoplanin based ALT. Both regimens achieved comparable catheter salvage rates, suggesting that the choice between these glycopeptides may be guided by organism susceptibility, drug availability, and patient-specific considerations rather than anticipated efficacy. This aligns with prior studies in which both agents demonstrated similar outcomes against CoNS.[13,17]

Only a limited number of studies in the literature have investigated the use of teicoplanin based ALT, and these have generally involved small case series. Del Pozo et al.[22] examined 44 adults with CoNS related bacteremia, of whom 17 received teicoplanin based ALT and 27 vancomycin-based ALT. None of the teicoplanin-treated patients required catheter removal, whereas 23% of those treated with vancomycin failed therapy and subsequently required catheter removal. Okur Acar et al.[21] recently reported a success rate of 72.7% with teicoplanin ALT in children with port related CoNS bacteremia. Conversely, Guédon et al.[23] found a considerably lower success rate of 37.5% in 24 adults with catheter related infections, which was attributed to the use of a low teicoplanin lock concentration (2.5 mg/ml).

Reported concentrations of teicoplanin based ALT vary considerably across studies. While Shan et al.,[12] in their review, recommended 10 mg/ml, and Okur Acar et al.[21] employed a teicoplanin lock concentration of 2.0 mg/ml and Guédon et al.[23] used 2.5 mg/ml, and Blanco-Di Matteo et al.,[24] in their study on haemodialysis related catheter infections, also applied a 10 mg/ml concentration. In line with our institutional protocol, we adopted a 10 mg/ml teicoplanin based ALT in this study. By contrast, vancomycin-based ALT concentrations are more consistent in the literature, with most reports supporting effective catheter salvage at 5 mg/ml.[12,24,25] Accordingly, this concentration was adopted in our protocol.

Methicillin resistance is a well-recognized predictor of poorer outcomes in CLABSI. Castagnola et al.[13] reported that MRCoNS isolates required higher vancomycin minimum inhibition concentrations (MICs), with standard intravenous therapy achieving success in 75% of cases when the MIC was ≤1 mg/L, but in only 6% when the MIC was 2-4 mg/L; catheter salvage improved modestly with adjunctive vancomycin lock therapy at 3 mg/ml. Similarly, Buonsenso et al.,[26] in a systematic review, found lower success rates and higher recurrence in MR *Staphylococcus aureus* related CLABSIs. We found higher success rate observed in MSCoNS CLABSIs (100% vs. 53%) compared to MRCoNS CLABSIs. While this trend is clinically relevant, our limited sample size may have restricted our ability to detect statistical significance.

Risk factor analysis identified younger age and repeated positive blood cultures as predictors of catheter loss aligning with other studies in literature.[4,19,27] These findings likely reflect a combination of host and infection related factors, such as the immature immune response in younger children and the increased bacterial burden associated with persistent bacteremia, both of which may diminish the likelihood of successful catheter salvage. They underscore the importance of careful patient selection and vigilant monitoring, particularly

in younger children or in those who continue to have positive cultures during ALT.

Limitations

Our study has several limitations, including its retrospective design, small sample size, non-randomized selection of ALT agents, incomplete CRP data, and potential confounding from heterogeneous underlying conditions and catheter duration.

CONCLUSION

To the best of our knowledge this is the first pediatric study comparing outcomes of vancomycin vs teicoplanin lock therapy in CONS CLABSIs. In conclusion, ALT achieved satisfactory catheter salvage rates in pediatric CoNS CLABSI, with similar outcomes for vancomycin and teicoplanin. Vancomycin or teicoplanin based ALT may be considered in selected children with CoNS CLABSI. Greater caution is warranted in younger patients, in those with persistent bacteraemia, and in cases caused by MRCoNS. Future prospective studies with larger sample sizes are needed to validate these risk factors, refine selection criteria, and optimize ALT protocols, including lock duration, concentration, and combination strategies.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was carried out with the permission of the Marmara University Faculty of Medicine Clinical Researches Ethics Committee (Date: 07.04.2023, Decision No: 09.2023.621).

Informed Consent

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

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

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

Author Contributions

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

REFERENCES

1. Hsu HE, Mathew R, Wang R, et al. Health care-associated infections among critically ill children in the US, 2013-2018. *JAMA Pediatr.* 2020; 174(12):1176-1183. doi:10.1001/jamapediatrics.2020.3223
2. Kopsidas I, Collins M, Zaoutis T. Healthcare-associated infections-can we do better? *Pediatr Infect Dis J.* 2021;40(8):e305-e309. doi:10.1097/INF.0000000000003203
3. Fratino G, Molinari AC, Parodi S, et al. Central venous catheter-related complications in children with oncological/hematological diseases: an observational study of 418 devices. *Ann Oncol.* 2005;16(4):648-654. doi:10.1093/annonc/mdi111
4. Walker LW, Visweswaran S, Nowalk AJ. Outcomes in pediatric central line-associated bloodstream infections treated with antimicrobial locks: a 14-year retrospective analysis. *Pediatr Infect Dis J.* 2023;42(6):473-478. doi:10.1097/INF.0000000000003885
5. Rabelo BS, de Alvarenga KAF, Miranda JWFB, et al. Risk factors for catheter-related infection in children with cancer: a systematic review and meta-analysis. *Am J Infect Control.* 2023;51(1):99-106. doi:10.1016/j.ajic.2022.05.005
6. Kiedrowski MR, Horswill AR. New approaches for treating staphylococcal biofilm infections. *Ann N Y Acad Sci.* 2011;1241:104-121. doi:10.1111/j.1749-6632.2011.06281.x
7. Otto M. Staphylococcal biofilms. *Microbiol Spectr.* 2018;6(4):10.1128/microbiolspec.gpp3-0023-2018. doi:10.1128/microbiolspec.GPP3-0023-2018
8. Alfieri A, Di Franco S, Passavanti MB, et al. Antimicrobial lock therapy in clinical practice: a scoping review. *Microorganisms.* 2025;13(2):406. doi:10.3390/microorganisms13020406
9. Mermel LA, Allon M, Bouza E, et al. Clinical practice guidelines for the diagnosis and management of intravascular catheter-related infection: 2009 update by the Infectious Diseases Society of America. *Clin Infect Dis.* 2009;49(1):1-45. doi:10.1086/599376
10. Members of the Sous les Verrous Study Group. Antibiotic lock therapy for the conservative treatment of long-term intravenous catheter-related infections in adults and children: when and how to proceed? Guidelines for clinical practice 2020. *Infect Dis Now.* 2021;51(3):236-246. doi:10.1016/j.idnow.2021.02.004
11. O'Grady NP, Alexander M, Dellinger EP, et al. Guidelines for the prevention of intravascular catheter-related infections. *Clinical Infectious Diseases.* 2002;35(11):1281-1307. doi:10.1086/344188
12. Shan A, Santos-Sebastian M, Saavedra-Lozano J. Lock therapy for treatment and prevention of catheter-related infections. *Pediatr Infect Dis J.* 2024;43(10):e358-e362. doi:10.1097/INF.0000000000004457
13. Buonsenso D, Salerno G, Sodero G, et al. Catheter salvage strategies in children with central venous catheter-related or -associated bloodstream infections: a systematic review and meta-analysis. *J Hosp Infect.* 2022;125:1-20. doi:10.1016/j.jhin.2022.03.010
14. Hecht SM, Ardura MI, Yildiz VO, Ouellette CP. Central venous catheter management in high-risk children with bloodstream infections. *Pediatr Infect Dis J.* 2020;39(1):17-22. doi:10.1097/INF.0000000000002495
15. Centers for Disease Control and Prevention. Device-associated module CLABSI. Available at: http://www.cdc.gov/nhsn/pdfs/pscmanual/4psc_clabscurrent.pdf [last accessed October 2021].
16. Megged O, Shalit I, Yaniv I, Fisher S, Livni G, Levy I. Outcome of antibiotic lock technique for persistent central venous catheter-associated coagulase-negative *Staphylococcus* bacteremia in children. *Eur J Clin Microbiol Infect Dis.* 2010;29(2):157-161. doi:10.1007/s10096-009-0831-4
17. Adler A, Yaniv I, Solter E, et al. Catheter-associated bloodstream infections in pediatric hematology-oncology patients: factors associated with catheter removal and recurrence. *J Pediatr Hematol Oncol.* 2006; 28(1):23-28. doi:10.1080/10910340500534282
18. Celebi S, Sezgin ME, Cakir D, et al. Catheter-associated bloodstream infections in pediatric hematology-oncology patients. *Pediatr Hematol Oncol.* 2013;30(3):187-194. doi:10.3109/08880018.2013.772683
19. Asrak HK, Belet N, Tüfekçi Ö, Özlü C, Baysal B, İnce D. Investigating the risk factors for antibiotic lock therapy failure in pediatric cancer: a single center retrospective analysis. *Turk J Pediatr.* 2021;63(1):86-94. doi: 10.24953/turkjped.2021.01.010
20. Qureshi S, Fatima P, Mukhtar A, Zehra A, Qamar FN. Clinical profile and outcome of antibiotic lock therapy for bloodstream infections in pediatric hematology/oncology patients in a tertiary care hospital, Karachi, Pakistan. *Int J Pediatr Adolesc Med.* 2019;6(1):25-28. doi:10.1016/j.ijpam.2019.01.004
21. Okur Acar S, Tahta N, Böncüoğlu E, et al. Efficacy of teicoplanin lock therapy in the treatment of port-related coagulase-negative *Staphylococcus* bacteremia in pediatric oncology patients. *J Pediatr Hematol Oncol.* 2023;45(1):e17-e20. doi:10.1097/MPH.0000000000002502
22. Del Pozo JL, Cenoz MG, Hernáez S, et al. Effectiveness of teicoplanin versus vancomycin lock therapy in the treatment of port-related coagulase-negative *Staphylococcus* bacteraemia: a prospective case-series analysis. *Int J Antimicrob Agents.* 2009;34(5):482-485. doi:10.1016/j.ijantimicag.2009.06.020

23. Guédon C, Nouvellon M, Lalaude O, Lerebours E. Efficacy of antibiotic-lock technique with teicoplanin in *Staphylococcus* epidermidis catheter-related sepsis during long-term parenteral nutrition. *J Parenter Enteral Nutrit*. 2002;26(2):109-113. doi:10.1177/0148607102026002109

24. Blanco-Di Matteo A, Garcia-Fernandez N, Aguinaga Pérez A, et al. In vivo effectiveness of several antimicrobial locks to eradicate intravascular catheter coagulase-negative *Staphylococcus* biofilms. *Antimicrob Agents Chemother*. 2023;67(1):e0126422. doi:10.1128/aac.01264-22

25. Kara TT, Özdemir H, Erat T, et al. Is antibiotic lock therapy effective for the implantable long-term catheter-related bloodstream infections in children? *Turk J Pediatr*. 2019;61(6):895-904. doi:10.24953/turkjped.2019.06.011

26. Castagnola E, Bandettini R, Lorenzi I, Caviglia I, MacRina G, Tacchella A. Catheter-related bacteremia caused by methicillin-resistant coagulase negative *Staphylococcus* with elevated minimal inhibitory concentration for vancomycin. *Pediatr Infect Dis J*. 2010;29(11):1047-1048. doi:10.1097/INF.0b013e3181e8b74d

27. Chen SH, Yang CP, Jaing TH, Lai JY, Hung IJ. Catheter-related bloodstream infection with removal of catheter in pediatric oncology patients: a 10-year experience in Taiwan. *Int J Clin Oncol*. 2012;17(2):124-130. doi:10.1007/s10147-011-0268-5

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1764625

J Health Sci Med. 2025;8(5):919-925

The effect of commonly used pediatric iron supplements on the color stability of restorative materials

Hülya Çerçi Akçay[1], Çağan Taş[2], Eda Sır[2], Nefise Şeker[3]

[1]Department of Pediatric Dentistry, Faculty of Dentistry, Kocaeli Health and Technology University, Kocaeli, Turkiye
[2]Dentistry Student, Faculty of Dentistry, Kocaeli Health and Technology University, Kocaeli, Turkiye
[3]Department of Public Health, Faculty of Medicine, Ankara University, Ankara, Turkiye

Cite this article as: Çerçi Akçay H, Taş Ç, Sır E, Şeker N. The effect of commonly used pediatric iron supplements on the color stability of restorative materials. *J Health Sci Med.* 2025;8(5):919-925.

Received: 13.08.2025 ◆ **Accepted**: 01.09.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: This in vitro study aimed to evaluate the effect of commonly prescribed pediatric iron supplements-ferric polymaltose (Ferrum®) and ferrous sulfate (Ferro Sanol B®)-on the color stability (ΔE) of different restorative materials.

Methods: One hundred eighty specimens from composite resin, compomer, flowable composite resin, and glass ionomer cement were assigned to distilled water (control), ferric polymaltose, or ferrous sulfate groups (n=15). All samples were stored at 37°C, immersed in the test solutions for 2 minutes daily, and evaluated at baseline, 7 days, and 28 days. Between immersions, specimens were kept in distilled water. Color measurements were performed using a spectrophotometer (VITA Easyshade, Germany) in the CIE Lab* system and ΔE values were calculated. Data were analyzed using Kruskal-Wallis and Mann-Whitney U tests ($p<0.05$).

Results: Both iron supplements caused significant discoloration in all restorative materials compared to the control ($p<0.05$). GIC exhibited the highest color change, while composite resin showed the lowest. Ferric polymaltose produced greater discoloration than ferrous sulfate in most material groups, particularly in GIC.

Conclusion: Pediatric iron supplements have a high potential for discoloration of restorative materials, with the iron's chemical form influencing the extent of staining. The development of formulations with reduced staining potential may contribute to improved esthetic outcomes in pediatric dental care.

Keywords: Color stability, dental materials, iron supplements, pediatric dentistry, spectrophotometry

INTRODUCTION

In recent years, advancements in pediatric restorative dentistry have increasingly emphasized both functional and esthetic outcomes.[1] Achieving a restoration that closely mimics the natural appearance of primary teeth is essential for both psychological and clinical success, particularly in anterior regions.[2] Due to anatomical differences such as thinner enamel and dentin layers, flat proximal contacts, and short retention periods, restorative procedures in primary teeth require materials that are not only durable but also esthetically pleasing and minimally invasive.[3]

Fluoride-releasing materials with favorable esthetic properties-including compomers, resin-modified glass ionomer cements (RMGICs), and composite resins-are widely preferred in pediatric clinical practice.[1] However, among the various criteria for successful outcomes, color stability remains a critical factor.[4] The optical behavior of restorative materials is influenced by intrinsic properties such as filler distribution, polymer matrix composition, and surface texture, as well as clinical factors like finishing/polishing techniques.[5] Over time, discoloration can occur due to environmental influences such as plaque accumulation, dietary pigments, and acidic conditions. These changes may present as external, superficial, or intrinsic staining and can compromise the longevity and esthetic quality of restorations.[6]

Iron deficiency anemia, frequently encountered during childhood, often necessitates treatment with oral iron supplements in the form of syrups or drops.[7] While these preparations are effective in restoring systemic iron levels, they may also have undesirable oral side effects.[8] Among them, tooth discoloration is of particular concern in pediatric patients. Ferric ions can form dark precipitates (such as Fe_2S_3) through chemical interactions in the oral environment, leading to both enamel staining and potential alterations in restorative materials.[9] In addition to extrinsic effects, oxidative reactions and ion complexation may also influence the material's optical properties and surface chemistry.[4]

Corresponding Author: Hülya Çerçi Akçay, hulya.akcay@kocaelisaglik.edu.tr

Although several studies have investigated the staining potential of iron supplements on both natural dentition and restorative materials, there remains a scarcity of research directly comparing the discoloration effects of different iron formulations across multiple commonly used pediatric restorative materials under standardized in vitro conditions.[6,10-13] The present study addresses this gap by simultaneously evaluating ferric polymaltose and ferrous sulfate on four distinct restorative materials, providing a clinically relevant comparative analysis.

Therefore, this in vitro study aimed to evaluate the effect of exposure to two commonly prescribed pediatric iron supplements (Ferrum® and Ferro Sanol B®) on the color stability of four restorative materials; composite resin, compomer, flowable composite resin, and conventional glass ionomer cement. By measuring the color change (ΔE) using spectrophotometry, this study seeks to provide evidence-based guidance for material selection in pediatric patients frequently receiving iron therapy.

The null hypothesis (H_0) of this study was that exposure to pediatric iron supplements would not cause a significant color change (ΔE) in any of the tested restorative materials.

The alternative hypothesis (H_1) was that exposure to pediatric iron supplements would result in a significant color change (ΔE) in at least one type of restorative material.

METHODS

Study Design

This in vitro experimental study was conducted at the Department of Pediatric Dentistry, Kocaeli Health and Technology University, Turkiye. Ethical approval was not required as the study involved only in vitro procedures using restorative dental materials, with no human or animal subjects. All procedures were carried out in accordance with the ethical rules and the principles. The aim was to evaluate the color stability of four different restorative materials after exposure to commonly prescribed pediatric iron supplements (Table 1, 2).

Sample Preparation

Four direct restorative materials were used: a conventional composite resin, a compomer, a flowable composite resin, and a conventional glass ionomer cement (GIC). A total of 180 cylindrical specimens were prepared (5 mm diameter × 3 mm height), with 15 specimens in each subgroup (n=15/group). Each material was divided into three subgroups based on the immersion medium:

- Control (immersed in distilled water)
- Ferrum® pediatric iron supplement
- Ferro Sanol B® pediatric iron supplement

This resulted in twelve experimental groups. For resin-based materials, specimens were placed in cylindrical silicone molds, covered with a glass slide to ensure a flat surface, and light-cured for 20 seconds using an LED curing device, following the manufacturer's instructions. GIC specimens were self-cured according to manufacturer guidelines (Table 3).

Sample Size Calculation

The sample size was determined using G*power software (version 3.1.9.4). Based on an assumed effect size of f=0.294, with α=0.05 and a desired statistical power of 90%, the

Table 1. Restorative materials used in this study

Material	Classification	Composition	Manufacturer
3M ESPE Filtek Z250 universal composite resin	A conventional composite resin	UDMA + Bis-EMA + Bis-GMA resin matrix; silane-treated ceramic fillers (approx. 75-85% wt)	3M ESPE, Dental products 2510 Conway Avenue St. Paul, MN 55144-1000 USA
Imicryl nova compomer flow	Compomer (flowable)	Resin-modified glass ionomer (compomer); resin matrix+fluoride-releasing glass ionomer components (ticarboxylic acid, glass powder)	Imicryl, Turkiye
3M ESPE filtek ultimate flowable composite	Flowable composite resin	Flowable composite Nanofiller composite: resin matrix with nanotechnology fillers for high polishability, radiopacity	33M ESPE, Dental products 2510 Conway Avenue St. Paul, MN 55144-1000 USA
Imicryl nova glass F glass ionomer filling cement	Conventional glass ionomer cement	Powder-liquid system: silicate glass powder+polyalkenoic acid; self-curing, radiopaque, high fluoride release	Imicryl, Turkiye

Table 2. Iron supplements used in this study

Product	Composition	Manufacturer
Ferrum®	Ferric (III) hydroxide polymaltose complex as iron source; excipients include sucrose, sorbitol, and flavoring agents	Vifor Pharma, Switzerland (or local license holder)
Ferro Sanol B®	Iron (II) glycine sulfate complex; formulation may also contain riboflavin, sodium phosphate, thiamine hydrochloride and pyridoxine hydrochloride depending on presentation	UCB Pharma, Germany (or local license holder)

Table 3. Specimen preparation and polymerization procedures recommended by the manufacturer for the restorative materials tested in this study

Material	Specimen preparation	Polymerization procedures
3M ESPE Filtek Z250 Universal Composite Resin	Placed into cylindrical silicone molds (5 mm×3 mm), covered with a glass slide to obtain a flat surface	Light-cured with an LED curing unit for 20 s
Imicryl nova compomer flow	Placed into molds, surface flattened with a glass slide	Light-cured with an LED curing unit for 20 s
3M ESPE Filtek ultimate flowable composite	Injected into molds, surface flattened with a glass slide	Light-cured with an LED curing unit for 20 s
Imicryl nova glass F glass ionomer filling cement	Hand-mixed powder and liquid placed into molds; surface flattened with a glass slide	Self-cured (setting time per manufacturer's instructions)

required sample size was calculated to be 168. To compensate for potential specimen loss, the final sample size was increased to 180, with 15 specimens allocated to each subgroup.

Each material group was randomly divided into three subgroups according to the immersion medium: a control group (immersed in distilled water), a group exposed to Ferrum® iron syrup, and a group exposed to Ferro Sanol B® iron syrup. As a result, twelve experimental groups were established, each comprising 15 specimens.

Immersion Protocol

Following specimen preparation, all samples were stored in an incubator at 37°C until the immersion procedures commenced. For the 1-day evaluation, specimens were immersed in their respective test solutions for 2 minutes on day 1, after which immediate color measurements were performed. For the 7-day evaluation, specimens were immersed for 2 minutes daily over seven consecutive days (days 1-7), with the final color measurement taken on day 7. For the 28-day evaluation, specimens underwent the same daily immersion protocol for 28 consecutive days, and the final color measurement was recorded on day 28. Representative images of the restorative material specimens immersed in pediatric iron supplement solutions (Ferrum® and Ferro Sanol B®) are shown in **Figure 1**.

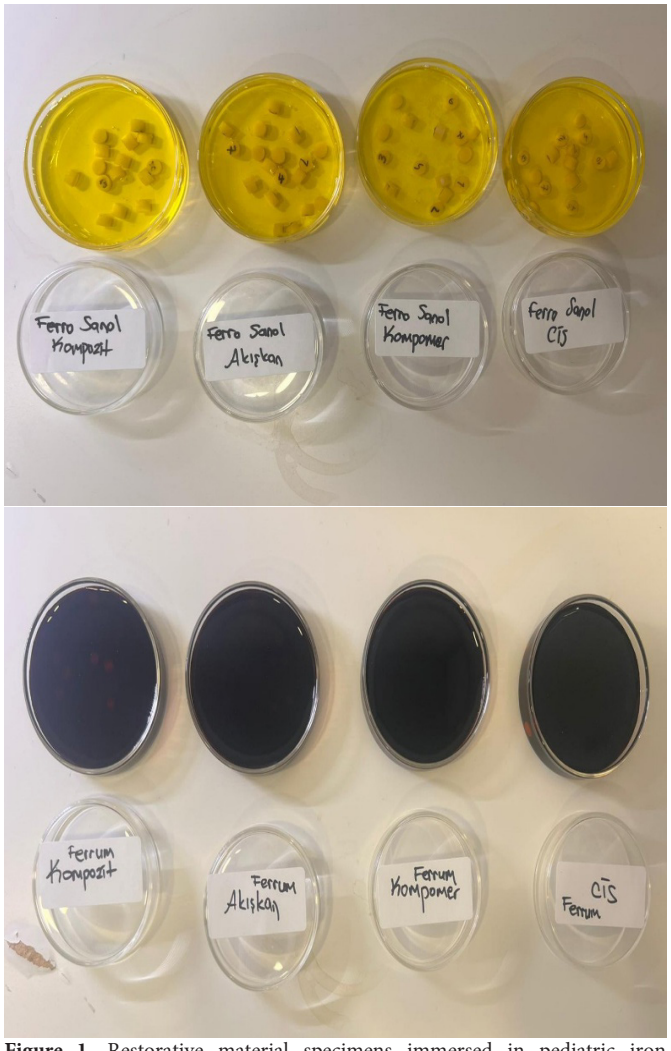

Figure 1. Restorative material specimens immersed in pediatric iron supplement solutions (Ferrum® and Ferro Sanol B®) during the experimental procedure

Between immersion periods, all specimens were stored in distilled water at room temperature to prevent dehydration and surface contamination. Prior to each color measurement, specimens were gently dried with absorbent paper to remove excess surface moisture. Measurements were performed against a standardized white background, and three consecutive readings were taken for each specimen using the spectrophotometer. The mean of these readings was used as the representative ΔE value for statistical analysis.

Color Measurements

Baseline color measurements (day 0) and subsequent measurements at days 1, 7, and 28 were performed using a VITA Easyshade V spectrophotometer (VITA Zahnfabrik, Bad Säckingen, Germany) in the CIE Lab* color space. Prior to each session, the device was calibrated according to the manufacturer's instructions. Measurements were taken by placing the specimens on a neutral white background to minimize reflection.

Color change (ΔE) was calculated using the formula:

$\Delta E=(\Delta L)2+(\Delta a)2+(\Delta b)2\Delta E = \sqrt{(\Delta L)^2 + (\Delta a)^2 + (\Delta b)^2}\Delta E=(\Delta L)2+(\Delta a)2+(\Delta b)2$

The following perceptibility and acceptability thresholds were applied: ΔE=1.2 (perceptible to 50% of observers) and ΔE=2.7 (clinically unacceptable).[14-16]

Statistical Analysis

Data analyses were performed using SPSS software (IBM SPSS Statistics, version 30.0; IBM Corp., Armonk, NY, USA). Normality was assessed using the Shapiro-Wilk test, and homogeneity of variances was evaluated using Levene's test. When the assumption of normality was not met, the Kruskal-Wallis test was applied; when assumptions were satisfied, Welch one-way ANOVA was used. Welch's test was applied with appropriate adjustments in cases of variance heterogeneity. For the Kruskal-Wallis test, effect sizes were calculated using the η^2 formula:

$\eta2=H–k+1n–k\eta^2 = \frac{H - k + 1}{n - k}\eta2=n–kH–k+1$

Where H is the test statistic, k is the number of groups, and n is the total sample size.

For Welch ANOVA, omega squared (ω^2) was calculated as the effect size. Effect sizes were interpreted according to Cohen[17] and Tomczak and Tomczak[18] as follows:

- Small effect: η^2 or $\omega^2 \geq 0.01$
- Medium effect: η^2 or $\omega^2 \geq 0.06$
- Large effect: η^2 or $\omega^2 \geq 0.14$

When a statistically significant difference was found, post-hoc tests were performed:

- For Kruskal-Wallis: Mann-Whitney U test with Bonferroni correction.
- For Welch ANOVA: Tamhane's T2 test.

The significance level was set at $p<0.05$ for all analyses, with Bonferroni-adjusted p-values applied where appropriate.

RESULTS

The present study evaluated the color stability (ΔE) of four restorative materials-composite resin, compomer, flowable composite, and glass ionomer cement-following exposure to three pediatric iron supplement conditions (Ferrum, Ferro Sanol, control) across five-time intervals (0-1, 0-7, 0-28, 1-7, and 7-28 days) as summarized in Table 4.

In the composite group, Ferrum produced significantly higher ΔE values than control at 0-1 days (mean: 32.9±17.9 vs. 13.5±5.0; p<0.001) and 0-7 days (mean: 52.7±38.2 vs. 27.5±8.3; p=0.004). Ferrum also exceeded Ferro Sanol at 0-1 days (p=0.002). Moreover, Ferro Sanol exhibited significantly higher discoloration than control at 0-28 days (median: 29 vs. 20; p=0.020) and 7-28 days (median: 24 vs. 29; p=0.044). The largest effects in this group occurred in the early evaluation period (0-1 days; η^2=0.421, large effect).

For compomer, large differences were recorded at 0-1 days (mean: Ferrum 32.0±14.1 vs. control 37.3±88.8; p<0.001), with Ferrum surpassing Ferro Sanol (p=0.006). At 1-7 days, Ferrum again produced greater discoloration than control (p=0.005), although the magnitude of change was smaller (median: 15 vs. 39).

The flowable composite demonstrated marked susceptibility at multiple intervals. At 0-1 days, both Ferrum (21.0±8.6) and Ferro Sanol (15.7±7.4) produced significantly greater ΔE than control (59.1±71.4; p=0.026 and p<0.001, respectively). At 0-7 days, Ferrum exceeded control (p=0.004) and Ferro Sanol (p=0.030). At 1-7 days, large differences were observed (η^2=0.442), with Ferrum and Ferro Sanol both differing significantly from each other (p<0.001) and from control (p=0.010). In the 7-28-day interval, all pairwise comparisons were statistically significant (p<0.05), with median values ranging from 22 in control to 53 in Ferrum, indicating a substantial long-term effect.

The glass ionomer cement group showed the highest early discoloration. At 0-1 days, Ferrum (mean: 197.1±127.7) induced significantly greater ΔE than control (92.3±101.2; p=0.011) and Ferro Sanol (37.7±21.0; p<0.001). Significant differences persisted at 0-7 days (Ferrum vs. Ferro Sanol, p<0.001) and 0-28 days (Ferrum vs. Ferro Sanol, p=0.003). Additionally, Ferro Sanol exceeded control at 0-28 days (p=0.001), despite having lower baseline values.

Overall, the most pronounced statistical differences were concentrated in the early (0-1 days) and cumulative (0-28 days) intervals. Flowable composite and glass ionomer cement exhibited the highest susceptibility to long-term (7-28 days) discoloration. Ferrum consistently induced greater ΔE values in short and medium terms, whereas the effect of Ferro Sanol varied by material type and was more prominent in extended exposure periods as detailed in Table 5, and illustrated in Figure 2.

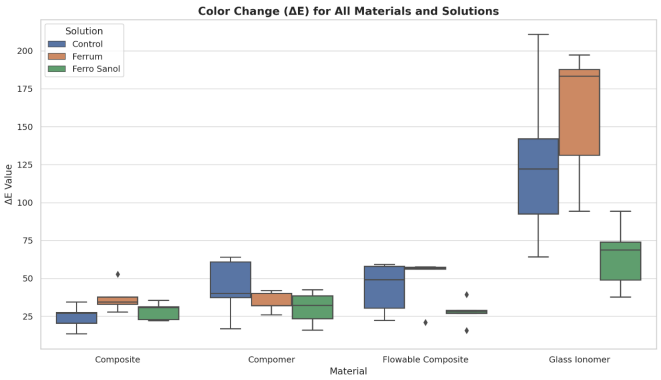

Figure 2. Boxplot representation of color change (ΔE) values for all tested restorative materials (composite resin, compomer, flowable composite resin, and glass ionomer cement) after exposure to control (distilled water), ferric polymaltose (Ferrum®), and ferrous sulfate (Ferro Sanol B®) pediatric iron supplement solutions

DISCUSSION

This in vitro study evaluated the color stability (ΔE) of four restorative materials-composite resin, compomer, flowable composite resin, and glass ionomer cement (GIC)-following exposure to two commonly prescribed pediatric iron supplements: ferric polymaltose (Ferrum Hausmann, Fe^{3+}) and ferrous sulfate (FerroSanol B, Fe^{2+}). These formulations were selected based on data from the Turkish Medicines and Medical Devices Agency, indicating that they are among the

Table 4. Descriptive statistics of the study groups (mean±SD)

Material	Group	0-1 day change mean±SD	0-1 day change median (Q1-Q3)	0-7 day change mean±SD	0-7 day change median (Q1-Q3)	0-28 day change mean±SD	0-28 day change median (Q1-Q3)	1-7 day change mean±SD	1-7 day change median (Q1-Q3)	7-28 day change mean±SD	7-28 day change median (Q1-Q3)
Composite	Control	13.5±5	14 (9-16)	27.5±8.3	27 (22-32)	20.5±7	20 (14-28)	27±10.3	28 (18-32)	34.3±12.5	29 (27-40)
	Ferrum	32.9±17.9	27 (20-37)	52.7±38.2	36 (32-47)	27.7±15.7	26 (16-39)	37.7±36	21 (16-60)	34.4±32.2	27 (15-31)
	Ferro Sanol	22±29	13 (10-22)	35.5±9.9	32 (27-48)	31.3±12.1	29 (21-41)	30.8±17.5	26 (23-37)	22.9±7.3	24 (22-27)
Compomer	Control	37.3±88.8	10 (4-19)	63.9±82.4	43 (35-51)	60.9±83.9	40 (36-46)	40.1±10.1	39 (33-50)	16.7±6.4	15 (13-21)
	Ferrum	32±14.1	27 (23-41)	42±24.4	35 (27-50)	40.1±30.4	30 (22-44)	25.9±26.7	15 (9-43)	32±28.6	16 (12-55)
	Ferro Sanol	15.9±9.1	13 (9-20)	42.4±10.5	43 (36-47)	38.4±13	34 (31-46)	32.1±7.3	30 (28-37)	23.4±9.3	24 (16-32)
Flowable Composite	Control	59.1±71.4	27 (25-62)	49±68.6	22 (16-26)	57.9±67.9	31 (27-43)	30.3±4.9	32 (25-33)	22.3±5.5	22 (20-25)
	Ferrum	21±8.6	19 (15-26)	57.5±30.9	50 (37-72)	57.3±30	55 (30-89)	56.1±28.1	50 (36-70)	56.6±28.1	53 (29-73)
	Ferro Sanol	15.7±7.4	14 (10-21)	27.9±6.7	27 (23-34)	39.3±6.7	39 (33-44)	26.9±6.3	27 (21-29)	28.9±6.6	26 (24-33)
Glass Ionomer	Control	92.3±101.2	48 (26-123)	122.2±111.4	79 (51-177)	210.7±106.2	200 (101-301)	64.1±69.5	46 (29-64)	141.9±123.5	114 (38-247)
	Ferrum	197.1±127.7	174 (86-250)	183.2±135.3	173 (64-212)	187.7±86.8	192 (120-265)	94.2±65	78 (41-126)	131.1±103.2	94 (47-209)
	Ferro Sanol	37.7±21	38 (21-54)	48.9±23	44 (30-72)	73.9±59.8	56 (42-91)	68.7±35.5	53 (44-96)	94.2±70.3	79 (44-125)

Table 5. Summary of statistical comparisons for color change (ΔE)

Material	Time interval	H/welch	df	p-value	Eta²	Significant pairwise comparisons (p<0.05)
Composite	0-1 day	19.710	2	<0.001	0.421	Control-Ferrum (<0.001); Ferrum-Ferro Sanol (0.002)
	0-7 days	10.697	2	0.005	0.207	Control-Ferrum (0.004)
	0-28 days	4.873	2	0.016	0.085	Control-Ferro Sanol (0.020)
	1-7 days	1.930	2	0.381	-0.002	-
	7-28 days	6.410	2	0.041	0.105	Control-Ferro Sanol (0.044)
Compomer	0-1 day	16.478	2	<0.001	0.345	Control-Ferrum (<0.001); Ferrum-Ferro Sanol (0.006)
	0-7 days	2.642	2	0.267	0.015	-
	0-28 days	4.310	2	0.116	0.055	-
	1-7 days	9.798	2	0.007	0.186	Control-Ferrum (0.005)
	7-28 days	3.754	2	0.153	0.042	-
Flowable composite	0-1 day	17.540	2	<0.001	0.370	Control-Ferro Sanol (<0.001); Control-Ferrum (0.026)
	0-7 days	11.438	2	0.003	0.225	Control-Ferrum (0.004); Ferrum-Ferro Sanol (0.030)
	0-28 days	3.423	2	0.181	0.034	-
	1-7 days	20.547	2	<0.001	0.442	Control-Ferrum (0.010); Ferrum-Ferro Sanol (<0.001)
	7-28 days	13.416	2	<0.001	0.421	Control-Ferro Sanol (0.017); Ferrum-Ferro Sanol (0.006)
Glass Ionomer	0-1 day	21.271	2	<0.001	0.459	Ferrum-Ferro Sanol (<0.001)
	0-7 days	16.117	2	<0.001	0.336	Ferrum-Ferro Sanol (<0.001)
	0-28 days	16.756	2	<0.001	0.351	Ferrum-Ferro Sanol (0.003)
	1-7 days	3.888	2	0.143	0.045	-
	7-28 days	0.822	2	0.663	-0.028	-

Note: Only statistically significant pairwise comparisons (p<0.05, adjusted) are shown. '-' indicates no significant difference

most frequently prescribed iron syrups for pediatric patients in the country, thereby ensuring the clinical relevance of the experimental design.[19]

Ferrous sulfate (Fe^{2+}) was included in this study due to its higher bioavailability compared to ferric salts, as consistently reported in clinical literature, and its well-documented potential to cause dental staining in pediatric populations.[8]

The null hypothesis (H_0), which assumed that exposure to pediatric iron supplements would not result in significant color change in any of the tested materials, was rejected, while the alternative hypothesis (H_1) was supported. Both supplements caused significant discoloration in all tested materials, with the extent of change depending on the type of material and the iron formulation. The greatest discoloration was observed in GIC, while composite resin exhibited the least. Ferric polymaltose consistently produced higher ΔE values than ferrous sulfate, particularly in GIC and compomer groups.

These results are consistent with previous findings. Kaya et al.[11] reported that Fe^{2+}-containing pediatric syrups caused greater discoloration in primary teeth than Fe^{3+} formulations, while Singh et al.[19] emphasized the role of low pH, acidic excipients, and synthetic dyes in the staining potential of pediatric liquid medications. Similarly, Pani et al.[20] observed that ferric polymaltose produced significantly higher ΔE values than ferrous fumarate after 72 h of exposure in primary teeth. The present findings align with this observation, as ferric polymaltose (Fe^{3+}) caused more intense staining than ferrous sulfate (Fe^{2+}), possibly due to the higher oxidative reactivity of ferric ions, which may promote faster pigment formation within and on the surface of restorative materials.

In addition to ion type, formulation technology also appears to play a critical role in discoloration potential. Lokhande et al.[12] demonstrated that conventional pediatric iron drops induced significantly greater discoloration (ΔE=40.6) compared to liposomal microencapsulated iron drops (ΔE=34.84), with the difference being statistically significant. This suggests that microencapsulation technology may reduce pigment adherence and penetration into dental tissues, potentially offering a clinically relevant strategy for minimizing esthetic compromise in pediatric patients requiring iron supplementation. The convergence of these findings underscores that both the chemical form of iron and its delivery vehicle substantially influence staining outcomes.

Consistent with our observations, Tüzüner et al.[6] reported that among pediatric formulations, Ferrosanol B induced the highest degree of discoloration. However, differences were observed in the ranking of restorative materials: in their study, composites exhibited the highest and glass ionomers the lowest discoloration susceptibility, whereas in the present study, glass ionomers showed the greatest color change and composites the least.

Differences among restorative materials can be explained by their intrinsic physicochemical properties. Glass ionomer cement (GIC), due to its hydrophilic polyalkenoate matrix, high surface porosity, and acid-base setting reaction, facilitates deeper penetration of chromogenic compounds and metallic ions into the material bulk, thereby increasing its susceptibility to irreversible discoloration.[21] In the present study, GIC exhibited the highest color change values at all time intervals, with ΔE values exceeding clinically perceptible thresholds as early as day 7, consistent with prior findings.[11]

Additionally, the continuous release and uptake of ions through its porous structure may act as channels for pigment infiltration, particularly in low pH environments.

In contrast, composite resins, which demonstrated the lowest ΔE values in our results, contain a predominantly hydrophobic resin matrix (e.g., bis-GMA, UDMA) and a highly cross-linked polymer network, which limits water sorption and restricts pigment uptake, thereby providing the greatest resistance to staining.[12] Compomers and flowable composites exhibited intermediate staining susceptibility, in line with their water absorption capacity being higher than that of conventional composites but lower than that of GIC.[22] In our data, compomers showed significantly less discoloration than GIC but more than composite resin, whereas flowable composites, due to their lower filler content and smaller filler size, tended to stain slightly more than compomers over prolonged exposure.

The acidic pH of pediatric iron syrups can soften and erode the resin matrix or the polysalt matrix in GIC, increasing surface roughness and creating microretentive sites for pigment deposition.[10] This synergistic effect of acidic degradation and pigment-rich solution exposure has been documented in several in vitro studies assessing the color stability of restorative materials following immersion in pediatric medications.[10-13] The present findings support these mechanisms, indicating that both material composition and the physicochemical properties of the staining agent critically determine the extent of discoloration.

Color measurements were performed using a spectrophotometer (VITA Easyshade) and the CIE L*a*b* color system, which is widely used in dentistry for its reproducibility and strong correlation with human visual perception.[23] The CIE L*a*b* system quantifies color in a three-dimensional space (L*, a*, b*), allowing objective detection of minimal changes and avoiding the subjectivity of visual assessment.[19] This methodological approach enhances the reliability of the present findings.

The measurement intervals of 0-1, 0-7, and 0-28 days were chosen to capture short-, medium-, and long-term effects. The 28-day period is widely used in in vitro color stability studies as it approximates 2-3 years of clinical service under accelerated conditions.[22] This duration allows observation of the plateau phase of pigment penetration, and in our study, prolonged exposure accentuated the differences among materials, with GIC and flowable composite exhibiting the most marked long-term discoloration.[24]

Clinically, discoloration caused by iron supplements may compromise the esthetic outcomes of restorations, particularly in anterior teeth of pediatric patients.[25] Given the psychological and social importance of esthetics in children, clinicians should consider using materials with greater color stability for patients likely to require long-term iron therapy and counsel parents about the risk of discoloration.

Limitations

The present study is subject to certain limitations that should be acknowledged. The in vitro design employed does not fully replicate the complex oral environment, where factors such as saliva composition, dental plaque accumulation, intraoral pH fluctuations, and the mechanical effects of toothbrushing may substantially influence the discoloration process of restorative materials. In addition, the scope of the investigation was limited to a specific number of iron supplement formulations and fixed exposure durations, which may not reflect the diversity of clinical scenarios. Another important aspect concerns the excipients commonly contained in pediatric iron syrups-such as flavoring agents, colorants, and pH modifiers-the influence of which on color stability could not be comprehensively assessed within the current model. Future research should therefore aim to incorporate a broader range of formulations, evaluate longer-term exposure periods, and integrate in vivo conditions to enhance both the clinical relevance and generalizability of the findings.

CONCLUSION

This in vitro study demonstrated that pediatric iron supplements possess a considerable staining potential on primary teeth and restorative materials. The chemical form of iron plays a significant role in the extent of discoloration, with ferric polymaltose (Fe^{3+}) producing greater color change than ferrous sulfate (Fe^{2+}) in most tested materials. As the formulation of iron syrups directly influences their staining potential, developing pediatric iron supplements with reduced discoloration properties may serve as an effective public health strategy to preserve the esthetic and functional integrity of restorations in pediatric patients.

ETHICAL DECLARATIONS

Ethics Committee Approval

Ethical approval was not required for this study, as it was conducted entirely in vitro using restorative dental materials and did not involve human or animal subjects.

Informed Consent

Because the study has no study with human and human participants, no written informed consent form was obtained.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

This research was funded by the Scientific Research Projects (BAP) Commission of Kocaeli Health and Technology University (Project No: 2025/2).

Author Contributions

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

Acknowledgment

The authors would like to thank the Faculty of Dentistry, Kocaeli Health and Technology University, for their support during the study.

REFERENCES

1. American Academy of Pediatric Dentistry. Pediatric restorative dentistry. In: The Reference Manual of Pediatric Dentistry. Chicago, IL: American Academy of Pediatric Dentistry; 2024:452-465.
2. American Academy of Pediatric Dentistry. Best practices for restorative dentistry. In: The Reference Manual of Pediatric Dentistry. Chicago, IL: American Academy of Pediatric Dentistry; 2019:340-352.
3. Mutluay MS. Süt dişlerinde restoratif materyal seçimi ve etkileyen faktörler. *Selcuk Dent J*. 2016;3(3):151-158. doi:10.15311/1441.300071
4. Genç G, Toz T. A review of the color stability of resin composites: the etiology, classification and the treatment of composite staining. *Ege Univ Dent Fac J*. 2017;38(2):68-79. doi:10.5505/eudfd.2017.43265
5. Yamanel K. Farklı içeceklerin diş rengindeki restoratif materyallerin renk stabiliteleri üzerine etkisi. *Süleyman Demirel Univ Saglik Bilim Derg*. 2018;9(2):26-31. doi:10.22312/sdusbed.409547
6. Tüzüner T, Turgut S, Baygin O, Yıldız M. Effects of different pediatric drugs on the color stability of various restorative materials applicable in pediatric dentistry. *BioMed Res Int*. 2017;1:9684193. doi:10.1155/2017/9684193
7. Özdemir N. Iron deficiency anemia from diagnosis to treatment in children. *Turk Pediatri Ars*. 2015;50(1):11-19. doi:10.5152/tpa.2015.2331
8. Yılmaz R, Aral YZ, Dallar Y. The comparison of the efficiency of once, twice or three times daily oral doses of ferrous sulphate in the treatment of iron deficiency anemia among children. *J Child Health Dis*. 2011;11(3):102-107. doi:10.5222/jchild.2011.102
9. Babaei N, Molaei T, Belyad S, et al. Relationship of pH and the viscosity of five different iron supplements with the absorption of iron ions and enamel discoloration in the anterior primary teeth (an in vitro study). *Dent Res J (Isfahan)*. 2021;18:7.
10. İnci MA, Özer H, Ercan VB. Diş rengindeki restoratif materyallerin farklı içeceklerde bekletilmesi sonrası renk stabilitelerinin ve yüzey pürüzlülüklerinin değerlendirilmesi. *Ege Univ Dis Hekim Fak Derg*. 2023;44(2):93-99.
11. Singh A, Grover C, Raina D, Pandey A, Krishna ASC. Color stability of pediatric restorative material over pediatric drug formulation. *Cureus*. 2023;15(8):e42953. doi:10.7759/cureus.42953
12. Lokhande RS, Nagarathna PJ, Deoghare A, Chhatani N, Busi S, Malladi S. Effect of different iron supplements on color stability of nanocomposite restorative materials. *Int J Clin Pediatr Dent*. 2024;17(3):274-278. doi:10.5005/jp-journals-10005-2790
13. Harza Y, Rao A, Natarajan S, Srikrishna SB. Effect of herbal toothpaste on the colour stability, surface roughness, and microhardness of aesthetic restorative materials-an in vitro study. *BDJ Open*. 2024;10(1):95. doi:10.1038/s41405-024-00280-x
14. Paravina RD, Pérez MM, Ghinea R. Acceptability and perceptibility thresholds in dentistry: a comprehensive review of clinical and research applications. *J Esthet Restor Dent*. 2019;31(2):103-112. doi:10.1111/jerd.12465
15. Paravina RD. Critical appraisal. Color in dentistry: match me, match me not. *J Esthet Restor Dent*. 2009;21(2):133-139. doi:10.1111/j.1708-8240.2009.00259.x
16. International Organization for Standardization. Dentistry-Guidance on Color Measurement. ISO/TR 28642. Geneva, Switzerland: ISO; 2016.
17. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
18. Tomczak M, Tomczak E. The need to report effect size estimates revisited: an overview of some recommended measures of effect size. *Trends Sport Sci*. 2014;21(1):19-25.
19. Kaya E. The effect of surface sealants on discoloration of primary teeth related to iron syrups. *Yeditepe J Dent*. 2021;17(3):171-176. doi:10.5505/yeditepe.2021.64426
20. Pani SC, Alenazi FM, Alotain AM, Alanazi HD, Alasmari AS. Extrinsic tooth staining potential of high dose and sustained release iron syrups on primary teeth. *BMC Oral Health*. 2015;15:90. doi:10.1186/s12903-015-0075-6
21. Janisch FAS, Aguilar MF, Aguiar FHB, França FMG, Bastos RT, Vieira-Junior WF. Surface roughness and color stability of conventional and bulk-fill resin composite with S-PRG fillers after coffee exposure: an in-vitro study. *Clin Cosmet Investig Dent*. 2025;17:21-29. doi:10.2147/CCIDE.S507164
22. Aktaş N, Akın Y, Bal C, Bani M, Bankoğlu Güngör M. Effect of the different dietary supplements on the average surface roughness and color stability of direct restorative materials used in pediatric dentistry. *Children (Basel)*. 2024;11(6):645. doi:10.3390/children11060645
23. Yılmaz N, Baygin O, Tüzüner T, Turgut SN, Erbek SM. Evaluation of the effect of pediatric drugs and an oral rinse on primary teeth discoloration. *Dent Med Probl*. 2022;59(2):225-231. doi:10.17219/dmp/133406
24. Valera B, Bhatt R, Patel M, Patel C, Makwani D, Goyal S. Effect of different pediatric medications on various tooth colored restorative materials used in pediatric dentistry: a comparative study. *Int J Health Sci*. 2022;6(S3):578-591. doi:10.53730/ijhs.v6nS3.5318
25. Kale YJ, Lingayat SD, Dadpe MV, et al. Assessment of detrimental effects of pediatric drug syrups on different esthetic dental restorative materials used in children: a systematic review. *World J Dent*. 2025;16(4):391-399. doi:10.5005/jp-journals-10015-2631

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1767263
J Health Sci Med. 2025;8(5):926-933

COVID-19, influenza, and pneumococcal vaccination awareness and uptake in geriatric patients: a cross-sectional study

Enes Dalmanoğlu[1], Ferda Özdemir[2], Hatice Engin[2], Mustafa Tekkoyun[2], Hilmi Berkay Aslan[2], Öykü Torun[2], Büşra Hacı[2], Abdulkerem Alegöz[2]

[1]Department of Infectious Diseases and Clinical Microbiology, Faculty of Medicine, Balıkesir University, Balıkesir, Turkiye
[2]Student, Faculty of Medicine, Balıkesir University, Balıkesir, Turkiye

Cite this article as: Dalmanoğlu E, Özdemir F, Engin H, et al. COVID-19, influenza, and pneumococcal vaccination awareness and uptake in geriatric patients: a cross-sectional study. *J Health Sci Med.* 2025;8(5):926-933.

Received: 19.08.2025 ◆ **Accepted**: 03.09.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: Respiratory tract infections are leading causes of morbidity and mortality in the geriatric population. The COVID-19 pandemic has reinforced the critical importance of vaccination against preventable respiratory diseases in elderly individuals. This study aimed to assess awareness levels, vaccination rates, and factors influencing vaccine acceptance for COVID-19, influenza, and pneumococcal vaccines among individuals aged ≥65 years.

Methods: This cross-sectional study was conducted between January 08, 2025, and February 28, 2025, involving patients aged ≥65 years who presented to the hospital for any reason. Structured interviews assessed vaccination history and awareness levels. Unvaccinated participants received standardized educational information about vaccine benefits and risks, followed by assessment of vaccination willingness. Demographic characteristics, educational level, and reasons for vaccine hesitancy were recorded. Statistical analysis was performed using SPSS v30.0, with Chi-square tests for categorical variables and $p<0.05$ considered significant.

Results: Among 168 participants (mean age 71.9±6.3 years, 52.4% female), vaccination rates were 96.4% for COVID-19, 59.5% for influenza (past year), and 14.9% for pneumococcal vaccine (past five years). Primary reasons for non-vaccination included lack of risk group awareness for pneumococcal (62.0%) and influenza vaccines (39.4%), while perceived vaccine inefficacy dominated COVID-19 hesitancy (85.7%). However, the COVID-19 unvaccinated subgroup was very small (n=6), limiting the robustness of statistical analysis for this vaccine. Higher educational level significantly correlated with influenza ($p=0.032$) and pneumococcal vaccination ($p=0.018$). Post-education, willingness to be vaccinated increased substantially: influenza 64.1%, pneumococcal 74.2%, and COVID-19 100%. It should be noted that this study measured vaccination intention rather than actual vaccine uptake.

Conclusion: Significant disparities exist in vaccination awareness and uptake across different vaccines in the geriatric population. However, as a single-center study focusing only on hospital-attending elderly adults, the generalizability of findings may be limited. Targeted educational interventions demonstrate substantial potential for improving vaccine acceptance intention. These findings highlight the need for healthcare provider-led education programs and policy initiatives to address knowledge gaps and enhance preventive care in this high-risk population.

Keywords: Geriatrics, COVID-19 vaccine, influenza vaccine, pneumococcal vaccine, vaccination awareness, vaccine hesitancy

INTRODUCTION

The geriatric population, defined as individuals aged 65 years and older, represents a particularly vulnerable group to infectious diseases due to age-related immune system decline (immunosenescence) and increased prevalence of chronic comorbidities. Immunosenescence involves progressive deterioration of immune function, including decreased T-cell proliferation and differentiation, reduced B-cell antibody production, impaired antigen presentation, and diminished vaccine-induced antibody responses, all of which contribute to increased susceptibility to infections and reduced vaccine efficacy in elderly individuals.[1] This demographic faces elevated risks from respiratory tract infections, including COVID-19, influenza, and pneumococcal diseases, which can result in significant morbidity and mortality.[2] These infections contribute to increased hospitalization rates, heightened complication risks, and impose substantial burdens on healthcare systems among elderly populations.[3]

Vaccination is recognized as one of the most effective evidence-based public health strategies for preventing these diseases, playing a critical role in establishing both individual and community immunity.[4] However, maximizing the protective benefits of vaccines depends on adequate awareness among

Corresponding Author: Enes Dalmanoğlu, enesdalmanoglu@gmail.com

individuals, comprehensive understanding of sociocultural, economic, and psychological factors influencing vaccine acceptance, and development of targeted interventions addressing these factors.[5] In Turkiye, vaccination services for the elderly are primarily delivered through family medicine practices as part of the national immunization program, with influenza and pneumococcal vaccines provided free of charge to individuals aged 65 and older. Family physicians play a crucial role in identifying at-risk populations, providing vaccination counseling, and maintaining immunization records through the national health information system.

In recent years, particularly following the emergence of the COVID-19 pandemic in late 2019, which evolved into a global health crisis, awareness and attitudes toward vaccination among the elderly population have become central topics in academic research.[6] While influenza and pneumococcal vaccines have long been integral components of routine immunization programs for protecting older adults, the rapid development and implementation of COVID-19 vaccines has introduced new dimensions to awareness dynamics in this age group.[7,8]

Despite years of established use of influenza and pneumococcal vaccines, awareness rates for these immunizations demonstrate significant variations across countries and regions, correlating with healthcare policy effectiveness and the reach of public information campaigns.[9] Conversely, the emergency authorization of COVID-19 vaccines and intensive media attention during the pandemic have heightened awareness of these vaccines among elderly individuals, while simultaneously introducing new challenges including vaccine hesitancy and misinformation.[10,11]

Awareness rates extend beyond individual health literacy limitations, being influenced by multifaceted factors including healthcare system accessibility, scope and effectiveness of information campaigns, misinformation disseminated through platforms such as social media, and individual trust perceptions.[12] These factors become more complex within the geriatric population. For instance, age-related cognitive decline, limited access to technological tools, difficulties accessing healthcare services in rural areas, and communication barriers between elderly individuals and healthcare providers can negatively impact vaccine awareness and subsequently vaccination rates.[13]

This study aims to evaluate awareness rates regarding COVID-19, influenza, and pneumococcal vaccination in the geriatric age group, identify reasons for non-vaccination, and examine changes in vaccine acceptance following educational intervention. Additionally, the study investigates the impact of sociodemographic factors on vaccine awareness.

METHODS

Ethics

This study was conducted in accordance with the Declaration of Helsinki and designed within the framework of ethical principles. Ethical approval was obtained from the Balıkesir University Rectorate Health Sciences Non-interventional Researches Ethics Committee (Date: 07.01.2025, Decision No: 2025/16). All data were anonymized to ensure patient privacy and confidentiality. Informed consent forms were obtained from all participants regarding the purpose, content, and possible risks of the study.

Study Design and Setting

This descriptive, cross-sectional study was conducted between January 8, 2025, and February 28, 2025, among patients aged 65 years and older who presented to the hospital for any reason.

Participants and Selection Criteria

Inclusion criteria were defined as: age ≥65 years, intact cognitive function, ability to communicate effectively, and consent to participate in the study. Exclusion criteria included: presence of cognitive impairment, communication difficulties, and refusal to participate in the study.

Data Collection

Demographic characteristics (age, gender, marital status, educational level, occupation) and comorbid conditions were recorded for all participants. During history-taking, patients were questioned about their vaccination status for COVID-19 vaccines, influenza vaccination within the past year, and pneumococcal vaccination after age 65 or within the past five years.

For patients who had not received vaccines, reasons for non-vaccination were classified as follows; lack of awareness of being in a risk group, considering vaccination unnecessary/disbelief in vaccine efficacy, fear of injections, fear of vaccine side effects, and other reasons.

Educational Intervention

Unvaccinated elderly participants received standardized educational information regarding the effects of influenza, pneumococcal, and COVID-19 infections in this age group, complication risks, vaccine efficacy, indications, and side effects. The educational content comprised three main components: epidemiology and clinical significance of the diseases in elderly populations, vaccine mechanisms of action, efficacy and safety, and potential post-vaccination side effects and their management. The educational intervention was delivered to each patient with identical content through approximately 10-minute individual consultations. Following the educational intervention, participants were asked whether they would accept vaccination, and their responses were recorded.

Statistical Analysis

Sample size calculation was performed using G*power (Version 3.1.9.4) software. Effect sizes reported in studies evaluating vaccine awareness and acceptance in geriatric populations were reviewed. Based on studies by Ciblak et al.[14] and Gazibara et al.,[15] which reported influenza vaccination rates of 30-60% in geriatric populations and medium effect sizes (d=0.3-0.7) for the relationship between educational level and vaccine acceptance, a medium effect size (Cohen's d=0.5) was adopted for this study. Statistical significance was evaluated with a 95% confidence interval (standard confidence

level), 5% margin of error (α=0.05, type 1 error probability), and 0.80 power (β=0.20, type 2 error probability), targeting a minimum of 160 participants. With 168 participants included in the study, final power analysis yielded 82.4%.

Data were analyzed using Statistical Package for Social Sciences (IBM SPSS, Armonk, NY, USA) v30.0 software. Descriptive statistical analysis methods were employed. Categorical variables were expressed as numbers and percentages, normally distributed numerical variables as mean±standard deviation, and non-normally distributed numerical variables as median (minimum-maximum). Normality distribution was assessed using the Kolmogorov-Smirnov test. Chi-square test was used to evaluate relationships between categorical variables. Fisher's exact test was applied when the proportion of cells with expected values less than 5 exceeded 20% or when any cell had an expected value less than 1. For comparison of vaccination willingness before and after educational intervention, McNemar's test was used as it involved repeated measurements of the same individuals. Statistical significance level was set at $p<0.05$.

RESULTS

A total of 168 patients aged ≥65 years were included in the study. The mean age was 71.9±6.3 years (range: 65-91), with 88 (52.4%) participants being female. Sociodemographic characteristics of the participants are presented in Table 1. Regarding educational status, 82 (48.8%) participants were primary school graduates, 27 (16.0%) were high school graduates, 27 (16.0%) were university graduates, 22 (13.0%) were middle school graduates, and 10 (5.9%) were illiterate. In terms of marital status, 132 (78.6%) participants were married, 25 (14.9%) were widowed, and 11 (6.5%) were single.

Analysis of comorbid conditions revealed that the most frequently observed comorbidities were hypertension (47.6%, n=80), diabetes mellitus (30.3%, n=51), coronary artery disease (15.5%, n=26), hyperlipidemia (7.1%, n=12), chronic obstructive pulmonary disease (6.0%, n=10), and history of cancer (4.8%, n=8). Other chronic conditions such as renal failure and rheumatoid arthritis were observed at lower rates (3.0%, n=5 each).

Vaccination status analysis demonstrated that 162 (96.4%) patients had received COVID-19 vaccination, 100 (59.5%) had received influenza vaccination within the past year, while only 25 (14.9%) had received pneumococcal vaccination within the past five years (Figure 1).

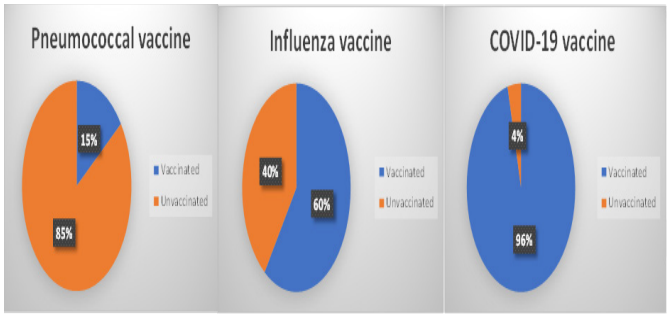

Figure 1. Vaccination rates for pneumococcal, influenza, and COVID-19 vaccines among patients

Table 1. Sociodemographic characteristics, occupational distribution, comorbidities, and educational status of participants (n=168)

Category	Number (n)	Percentage (%)
Age (years)		
Mean±SD	71.9 ± 6.3	
Range	65-91	
Gender		
Female	88	52.4
Male	80	47.6
Marital status		
Married	132	78.5
Single	11	6.5
Divorced/widowed	25	14.8
Occupation		
Housewife	66	39.3
Farmer	17	10.1
Teacher	14	8.3
Civil servant	13	7.7
Tradesperson	11	6.6
Retired	9	5.4
Worker	7	4.2
Other	31	18.4
Comorbidities		
Hypertension	80	47.6
Diabetes mellitus	51	30.3
Coronary artery disease	26	15.5
Hyperlipidemia	12	7.1
COPD	10	6.0
Cancer	8	4.8
Renal failure	5	3.0
Rheumatoid arthritis	5	3.0
Educational level		
Illiterate	10	5.9
Primary school	82	48.8
Middle school	22	13.0
High school	27	16.0
University	27	16.0

SD: Standard deviation, COPD: Chronic obstructive pulmonary disease

When reasons for non-vaccination were examined, the most common reason among patients who had not received influenza vaccination was lack of awareness of being in a risk group (39.4%, n=30), followed by disbelief in vaccine efficacy/considering it unnecessary (30.0%, n=23), and concern about vaccine side effects (9.2%, n=7). For patients who had not received pneumococcal vaccination, the most frequent reason was also lack of awareness of being in a risk group (62.0%, n=92). Other reasons included disbelief in vaccine efficacy/considering it unnecessary (16.9%, n=25), concern about vaccine side effects (7.4%, n=11), and fear of injections (0.6%, n=1). Among patients who had not received COVID-19

vaccination, 85.7% (n=6) cited disbelief in vaccine efficacy/considering it unnecessary, while 14.2% (n=1) cited concern about vaccine side effects (Figure 2). However, it should be noted that the COVID-19 unvaccinated subgroup was extremely small (n=6), which significantly limits the statistical robustness of these findings.

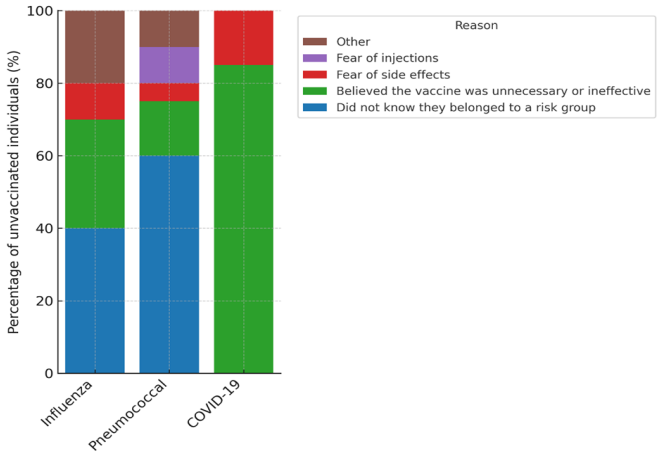

Figure 2. Reasons for not receiving influenza, pneumococcal, and COVID-19 vaccines among unvaccinated participants

Analysis of the relationship between educational level and vaccination status revealed that 55.2% (n=37) of those who had not received influenza vaccination, 47.8% (n=68) of those who had not received pneumococcal vaccination, and 50.0% (n=3) of those who had not received COVID-19 vaccination were primary school graduates (Table 1, 2). A statistically significant association was found between educational level and influenza vaccination (p=0.032). Similarly, a significant relationship was observed between educational level and pneumococcal vaccination (p=0.018). Higher educational levels were associated with increased vaccination rates. The relationship between educational level and COVID-19 vaccination was not statistically significant (p=0.094) (Table 2).

When evaluating the relationship between presence of comorbid conditions and vaccination status, individuals with chronic diseases, particularly those with diabetes (p=0.029) and COPD (p=0.007), had significantly higher rates of influenza vaccination. A similar trend was observed for pneumococcal vaccination, although statistical significance was not reached (p=0.058). Detailed analysis of vaccination status according to comorbidity subgroups is presented in Table 3.

Following provision of information about vaccine effects, indications, and side effects, 64.1% (n=43) of those who had not received influenza vaccination, 74.2% (n=106) of those who had not received pneumococcal vaccination, and 100% (n=6) of those who had not received COVID-19 vaccination expressed willingness to be vaccinated.

Table 2. Association between educational level and vaccination status for influenza, pneumococcal, and COVID-19 vaccines

	Influenza vaccination		Pneumococcal vaccination		COVID-19 vaccination	
Educational level	**Yes n (%)**	**No n (%)**	**Yes n (%)**	**No n (%)**	**Yes n (%)**	**No n (%)**
Illiterate (n: 10)	7 (70.0)	3 (30.0)	0 (0.0)	10 (100)	10 (100)	0 (0.0)
Primary school (n: 82)	45 (54.9)	37 (45.1)	14 (17.1)	68 (82.9)	79 (96.3)	3 (3.7)
Middle school (n: 22)	13 (59.1)	9 (40.9)	7 (31.8)	15 (68.2)	20 (90.9)	2 (9.1)
High school (n: 27)	17 (63.0)	10 (37.0)	2 (7.4)	25 (92.6)	26 (96.3)	1 (3.7)
University (n: 27)	19 (70.4)	8 (29.6)	3 (11.1)	24 (88.9)	27 (100)	0 (0.0)
Total (n: 168)	100 (59.5)	68 (40.5)	25 (14.9)	143 (85.1)	162 (96.4)	6 (3.6)
p-value	p: 0.032*		p: 0.018*		p: 0.094	

*Chi-square test; p<0.05 was considered statistically significant

Table 3. Association between comorbidities and vaccination status for influenza, pneumococcal, and COVID-19 vaccines

	Influenza vaccination		Pneumococcal vaccination		COVID-19 vaccination	
Comorbidity	**Yes n (%)**	**No n (%)**	**Yes n (%)**	**No n (%)**	**Yes n (%)**	**No n (%)**
Hypertension (n: 80)	52 (65.0)	28 (35.0)	14 (17.5)	66 (82.5)	78 (97.5)	2 (2.5)
Diabetes mellitus (n: 51)	35 (68.6)	16 (31.4)	9 (17.6)	42 (82.4)	50 (98.0)	1 (2.0)
Coronary artery disease (n: 26)	18 (69.2)	8 (30.8)	5 (19.2)	21 (80.8)	26 (100)	0 (0.0)
Hyperlipidemia (n: 12)	8 (66.7)	4 (33.3)	2 (16.7)	10 (83.3)	12 (100)	0 (0.0)
COPD (n: 10)	9 (90.0)	1 (10.0)	3 (30.0)	7 (70.0)	10 (100)	0 (0.0)
Cancer (n: 8)	5 (62.5)	3 (37.5)	1 (12.5)	7 (87.5)	8 (100)	0 (0.0)
Renal failure (n: 5)	3 (60.0)	2 (40.0)	1 (20.0)	4 (80.0)	5 (100)	0 (0.0)
Rheumatoid arthritis (n: 5)	4 (80.0)	1 (20.0)	1 (20.0)	4 (80.0)	5 (100)	0 (0.0)
p-value (diabetes mellitus)	p: 0.029*		p: 0.058		p: 0.814	
p-value (COPD)	p: 0.007*		p: 0.112		p: 0.905	

*Chi-square test; p<0.05 was considered statistically significant, COPD: Chronic obstructive pulmonary disease

The difference between pre- and post-educational intervention vaccination willingness was evaluated using McNemar's test and found to be statistically significant for all three vaccine types (influenza: p<0.001; pneumococcal: p<0.001; COVID-19: p=0.031). Statistical analysis of vaccination status and associated factors in the geriatric age group is presented in Table 4. Regarding COVID-19 vaccination specifically, while the small number of unvaccinated patients (n=6) may limit statistical interpretation, the willingness to accept vaccination by all patients following educational intervention represents a clinically meaningful development. However, it is important to note that this study measured vaccination intention rather than actual vaccine uptake, and the intention-behavior gap documented in health behavior literature may result in lower actual vaccination rates.

DISCUSSION

This study presents significant findings regarding awareness levels and reasons for non-vaccination against COVID-19, influenza, and pneumococcal vaccines in the geriatric population. Our results demonstrate substantial differences in awareness and acceptance rates among vaccine types, significant improvement in vaccine acceptance following educational intervention, and the influential role of educational level on vaccination behavior. However, several limitations should be acknowledged. This single-center study focused exclusively on hospital-attending elderly adults, which may limit the generalizability of findings to the broader geriatric population, particularly those without regular healthcare access or residing in rural areas. Additionally, socioeconomic status and urban-rural residence patterns were not assessed, which may have provided additional insights into vaccination disparities.

In our study, COVID-19 vaccination was administered at a remarkably high rate of 96.4%, while influenza vaccination reached 59.5%, and pneumococcal vaccination was administered to only 14.9% of participants. The high COVID-19 vaccination rate can be attributed to the global urgency perception created by the pandemic, intensive media campaigns, government prioritization of vaccination programs, and facilitated vaccine access.[16] Particularly during 2020-2023, the high mortality rate of COVID-19 in the elderly population may have increased this group's motivation toward vaccination.[17] This finding is significant in demonstrating the impact of global health crises on vaccine acceptance and provides lessons for future vaccination campaigns.

The 59.5% influenza vaccination rate suggests that despite long-standing routine recommendations for this vaccine, complete acceptance has not been achieved and awareness gaps persist. Similarly, studies conducted in Turkiye and worldwide report influenza vaccination rates ranging between 30-60% in the geriatric population.[14,15] These rates fall below the World Health Organization's target vaccination rate of 75% for individuals over 65 years.[18] This situation indicates that national health policies should more strongly promote influenza vaccination for the elderly.

The low pneumococcal vaccination rate of 14.9% suggests that this vaccine is not adequately promoted in the geriatric population or is not integrated as a routine practice in healthcare systems. Although literature has proven that pneumococcal polysaccharide and conjugate vaccines are effective in preventing invasive pneumococcal disease and pneumonia in the elderly,[19] the low rates in this study reveal the gap between awareness and implementation. Similarly, a study conducted in Turkiye reported pneumococcal vaccination rates of 7.9% in individuals over 65 years.[20] Pooled PPSV23 effectiveness against VT-IPD was 45% (95% CI: 37%, 51%; I^2=0%) in older adults.[21] The low implementation rate of such an effective preventive measure should be considered a significant public health concern.

Analysis of reasons for non-vaccination reveals different dynamics for each of the three vaccines. While the most common reason for influenza and pneumococcal vaccines was "lack of awareness of being in a risk group" (39.4% and 62%, respectively), the most frequent reason for COVID-19 vaccine was "disbelief in vaccine efficacy/considering it unnecessary" (85.7%). However, this finding for COVID-19 vaccine hesitancy should be interpreted with extreme caution given the very small sample size (n=6) of unvaccinated individuals, which may not be representative of broader hesitancy patterns. This finding may reflect the impact of widespread misinformation and anti-vaccine rhetoric during the pandemic, despite intensive information campaigns about COVID-19 vaccines.[22] MacDonald and colleagues' study emphasized that vaccine hesitancy is a complex phenomenon influenced by contextual, individual, and vaccine-specific factors.[23] Hesitancy toward COVID-19 vaccines may be related to the rapid development process of the vaccine, use of new technologies, and unfounded claims spread on social media.

The high rate of lack of awareness of being in a risk group for pneumococcal vaccination (62%) suggests that this vaccine is not adequately promoted in the geriatric population and is not routinely recommended by healthcare professionals.[24] This finding indicates that healthcare workers should routinely assess vaccination status in geriatric patient evaluations and provide necessary recommendations. Nagata and colleagues' study demonstrated that direct recommendations by healthcare professionals play a decisive role in acceptance of less well-known vaccines, particularly pneumococcal vaccines.[25]

Table 4. Change in willingness to be vaccinated before and after educational intervention

Vaccine type	Vaccinated before intervention n (%)	Willing to be vaccinated after intervention n (%)	Change n (%)	p-value
Influenza	100 (59.5)	143 (85.1)	43 (25.6)	p<0.001*
Pneumococcal	25 (14.9)	131 (78.0)	106 (63.1)	p<0.001*
COVID-19	162 (96.4)	168 (100.0)	6 (3.6)	p=0.031*

*McNemar's test; p<0.05 was considered statistically significant

The relationship between educational level and vaccination status is noteworthy. For all three vaccine types, the majority of those who did not receive vaccination were primary school graduates (influenza: 55.2%, pneumococcal: 47.8%, COVID-19: 50%). This finding emphasizes the impact of low educational level on health literacy and access to healthcare services.[26] As education level increases, individuals' access to health information, comprehension capacity, and critical evaluation skills improve, which can positively influence vaccine acceptance.[27] Lorini and colleagues' systematic review confirmed the effect of health literacy on vaccine acceptance and noted that interventions aimed at improving health literacy could enhance vaccine coverage.[28] This finding highlights the importance of developing comprehensible and accessible educational materials for the geriatric population with low educational levels.

The effect of chronic disease presence on vaccination behavior is particularly noteworthy. Patients diagnosed with diabetes mellitus were found to receive influenza vaccination at significantly higher rates (p=0.029). Similarly, patients with COPD also received influenza vaccination at higher rates compared to other patients (p=0.007). This situation can be explained by more effective education of individuals with chronic diseases, particularly those with respiratory and metabolic diseases, by healthcare professionals regarding vaccination, and these patients' greater awareness of the risks that respiratory tract infections pose to them.[22,24] Interestingly, similar increases in pneumococcal vaccination rates were not observed in the same chronic disease groups, suggesting that education regarding pneumococcal vaccination is more inadequate compared to influenza vaccination. Recent studies have shown that pneumococcal vaccination in COPD patients reduces acute exacerbations and hospitalizations.[29] Therefore, stronger emphasis on pneumococcal vaccination in clinical guidelines prepared for COPD patients and increasing healthcare professionals' awareness on this topic is important.

One of the most striking findings of our study was the significant increase in vaccine acceptance rates following educational intervention. Following education, 64.1% of those who had not received influenza vaccination, 74.2% of those who had not received pneumococcal vaccination, and 100% of those who had not received COVID-19 vaccination decided to accept vaccination. This result demonstrates that accurate and comprehensive education by healthcare professionals is an effective strategy in overcoming vaccine hesitancy.[30] The particularly high conversion rate for pneumococcal vaccination (74.2%) reveals that the low initial acceptance rate for this vaccine largely stems from lack of awareness and that this gap can be closed through targeted education. Jarrett and colleagues' systematic review noted that personalized, face-to-face education is one of the most effective methods for increasing vaccine acceptance.[31] In light of these findings, it is recommended that vaccination programs for the geriatric population prioritize active participation of healthcare professionals and personalized education strategies.

While developing strategies to increase vaccine awareness in the post-pandemic period, lessons can be learned from factors that ensured high acceptance rates for COVID-19 vaccination. Consideration should be given to how elements such as emergency perception, media campaigns, ease of access, and social solidarity can be integrated into other vaccination programs. Particularly, making pneumococcal vaccination part of routine elderly health checkups, developing vaccine reminder systems, and organizing community-based awareness campaigns could be effective strategies.

Special approaches should be developed for elderly individuals with low educational levels. Visual materials, brochures prepared in simple and understandable language, short videos, and education through community leaders may be more effective for this group. Additionally, technological solutions such as mobile health applications and telemedicine services may help reach elderly individuals, particularly those living in rural areas.

The strengths of this study include its comparative evaluation of three different vaccine types, direct measurement of the effect of educational intervention, and examination of the impact of sociodemographic factors on vaccine awareness. Our study is among those that quantitatively evaluate the effect of education on vaccine acceptance using a standardized educational protocol. However, the study has several limitations. First, the study was conducted at a single center, and findings may not be generalizable to the entire geriatric population. Patients included in our study were individuals who already had access to healthcare services and had presented to a healthcare institution; therefore, vaccine awareness and acceptance levels of elderly individuals without access to healthcare services or those who do not present to healthcare facilities may differ. Second, whether the decision to vaccinate following education translated into actual vaccine administration was not followed up. This situation, known as the intention-behavior gap, refers to the difference between individuals stated intentions and actual behaviors, and examination of this relationship is recommended in future studies. Finally, psychological, cultural, and social aspects of vaccine hesitancy were not examined in depth. In future studies, more comprehensive evaluation of these factors using qualitative research methods would be beneficial.

Future studies should be designed to encompass broader and more heterogeneous populations, evaluate long-term effects of interventions aimed at increasing vaccine awareness, and examine the multidimensional nature of vaccine hesitancy more deeply. Additionally, evaluation of the effectiveness of innovative intervention strategies such as mobile health applications, remote education modules, and community-based approaches is recommended. Particularly, increasing awareness of pneumococcal vaccination is critically important in reducing morbidity and mortality due to pneumococcal infections in the geriatric population.

Limitations

In subgroup analyses, particularly for comorbidity groups with small sample sizes (Chronic obstructive pulmonary disease (COPD) n=10, cancer n=8, renal failure n=5, rheumatoid arthritis n=5), analyses were considered exploratory and results were interpreted cautiously. Similarly, due to the small number of patients who had not received COVID-19

vaccination (n=6), statistical analyses for this group may have limited power, and findings were interpreted carefully.

CONCLUSION

This study demonstrates that awareness levels regarding COVID-19, influenza, and pneumococcal vaccination in the geriatric population vary significantly by vaccine type, with the primary reasons for vaccine refusal being lack of knowledge, distrust in vaccine efficacy, and fear of side effects. Patient education emerges as an effective strategy for overcoming these barriers, while low educational level shows a strong association with vaccine hesitancy. However, it is important to note that this study measured vaccination intention rather than actual vaccine uptake, and the single-center design limits generalizability to the broader geriatric population. Based on our findings, the following recommendations can be developed: Routine assessment of vaccination status and recommendation of necessary vaccines during geriatric health checkups should become standard practice for healthcare professionals. Awareness campaigns should be organized specifically for pneumococcal vaccination, emphasizing the importance of this vaccine. Elderly individuals with chronic diseases, particularly those with diabetes and COPD, should receive specialized education regarding regular vaccination. Comprehensible and accessible educational materials should be developed for the geriatric population with low educational levels. In the post-pandemic period, similar strategies should be applied to other vaccination programs by leveraging factors that ensured high acceptance rates for COVID-19 vaccination. In conclusion, efforts to increase vaccine awareness and acceptance intention in the geriatric population should be prioritized in public health initiatives for protecting elderly health and reducing morbidity and mortality from infectious diseases. In this vulnerable population where immunosenescence and comorbidities are prevalent, vaccination should be positioned as an indispensable component of healthcare services, representing the most effective means of protection against vaccine-preventable diseases.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was carried out with the permission of the Balıkesir University Rectorate Health Sciences Non-interventional Researches Ethics Committee (Date: 07.01.2025, Decision No: 2025/16).

Informed Consent

All patients signed and free and informed consent form.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

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

Author Contributions

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

REFERENCES

1. Weinberger B. Vaccines for the elderly: current use and future challenges. *Immun Ageing.* 2018;15:3. doi:10.1186/s12979-017-0107-2
2. World Health Organization. Pneumococcal vaccines: WHO position paper. *Wkly Epidemiol Rec.* 2022;97(8):73-95.
3. Thompson WW, Shay DK, Weintraub E, et al. Mortality associated with influenza and respiratory syncytial virus in the United States. *JAMA.* 2003;289(2):179-186. doi:10.1001/jama.289.2.179
4. Centers for Disease Control and Prevention. Vaccination recommendations for older adults. Updated 2021. Accessed April 20, 2025. https://www.cdc.gov/vaccines/adults/rec-vac/index.html
5. Verger P, Dubé E. Restoring confidence in vaccines in the COVID-19 era. *Expert Rev Vaccines.* 2020;19(11):991-993. doi:10.1080/14760584.2020.1825945
6. Lazarus JV, Ratzan SC, Palayew A, et al. A global survey of potential acceptance of a COVID-19 vaccine. *Nat Med.* 2021;27(2):225-228. doi:10.1038/s41591-020-1124-9
7. World Health Organization. *Immunization in the Context of COVID-19 Pandemic: Practical Considerations.* World Health Organization; 2021.
8. Domnich A, Arata L, Amicizia D, Puig-Barberà J, Gasparini R, Panatto D. Effectiveness of MF59-adjuvanted seasonal influenza vaccine in the elderly: a systematic review and meta-analysis. *Vaccine.* 2017;35(4):513-520. doi:10.1016/j.vaccine.2016.12.011
9. Kwok KO, Li KK, Wei WI, Tang A, Wong SYS, Lee SS. Influenza vaccine uptake, COVID-19 vaccination intention and vaccine hesitancy among nurses: a survey. *Int J Nurs Stud.* 2021;114:103854. doi:10.1016/j.ijnurstu.2020.103854
10. Loomba S, de Figueiredo A, Piatek SJ, de Graaf K, Larson HJ. Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nat Hum Behav.* 2021;5(3):337-348. doi:10.1038/s41562-021-01056-1
11. Forman R, Shah S, Jeurissen P, Jit M, Mossialos E. COVID-19 vaccine challenges: what have we learned so far and what remains to be done? *Health Policy.* 2021;125(5):553-567. doi:10.1016/j.healthpol.2021.03.013
12. Yaqub O, Castle-Clarke S, Sevdalis N, Chataway J. Attitudes to vaccination: a critical review. *Soc Sci Med.* 2014;112:1-11. doi:10.1016/j.socscimed.2014.04.018
13. Porcari DE, Palmer K, Spalletta G, Ciullo V, Banaj N. A survey for examining the effects of COVID-19 and infection control measures in older persons with mild cognitive impairment and dementia and their caregivers. *Front Psychiatry.* 2020;11:599851. doi:10.3389/fpsyt.2020.599851
14. Ciblak MA; Influenza Platform Working Group. Influenza vaccination in Turkiye: prevalence of risk groups, current vaccination status, factors influencing vaccine uptake and steps taken to increase vaccination rate. *Vaccine.* 2013;31(3):518-523. doi:10.1016/j.vaccine.2012.11.022
15. Gazibara T, Kovacevic N, Kisic-Tepavcevic D, et al. Flu vaccination among older persons: study of knowledge and practices. *J Health Popul Nutr.* 2019;38(1):2. doi:10.1186/s41043-018-0159-8
16. Schaffer DeRoo S, Pudalov NJ, Fu LY. Planning for a COVID-19 vaccination program. *JAMA.* 2020;323(24):2458-2459. doi:10.1001/jama.2020.8711
17. Onder G, Rezza G, Brusaferro S. Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. *JAMA.* 2020;323(18):1775-1776. doi:10.1001/jama.2020.4683
18. World Health Organization. Global Influenza Strategy 2019-2030. World Health Organization; 2019.
19. Bonten MJM, Huijts SM, Bolkenbaas M, et al. Polysaccharide conjugate vaccine against pneumococcal pneumonia in adults. *N Engl J Med.* 2015;372(12):1114-1125. doi:10.1056/NEJMoa1408544
20. Dereli F, Uyanik G, Duman JGY, Kundakçi G, Yilmaz M. Determination of immunization status of individuals aged 65 and over: family health center example. *İKÇÜSBFD.* 2022;7(2):291-301.
21. Kobayashi M, Pilishvili T, Farrar JL, et al. Systematic review and meta-analysis of the efficacy and effectiveness of pneumococcal vaccines in adults. *Pathogens.* 2023;12(5):732. doi:10.3390/pathogens12050732

22. Jennings W, Stoker G, Bunting H, et al. Lack of trust, conspiracy beliefs, and social media use predict COVID-19 vaccine hesitancy. *Vaccines (Basel).* 2021;9(6):593. doi:10.3390/vaccines9060593
23. MacDonald NE; SAGE Working Group on Vaccine Hesitancy. Vaccine hesitancy: definition, scope and determinants. *Vaccine.* 2015;33(34): 4161-4164. doi:10.1016/j.vaccine.2015.04.036
24. Eilers R, Krabbe PFM, de Melker HE. Factors affecting the uptake of vaccination by the elderly in Western society. *Prev Med.* 2014;69:224-234. doi:10.1016/j.ypmed.2014.10.017
25. Nagata JM, Hernández-Ramos I, Kurup AS, Albrecht D, Vivas-Torrealba C, Franco-Paredes C. Social determinants of health and seasonal influenza vaccination in adults ≥65 years: a systematic review of qualitative and quantitative data. *BMC Public Health.* 2013;13:388. doi:10.1186/1471-2458-13-388
26. Okan O, Bollweg TM, Berens EM, Hurrelmann K, Bauer U, Schaeffer D. Coronavirus-related health literacy: a cross-sectional study in adults during the COVID-19 infodemic in Germany. *Int J Environ Res Public Health.* 2020;17(15):5503. doi:10.3390/ijerph17155503
27. Castro-Sánchez E, Chang PWS, Vila-Candel R, Escobedo AA, Holmes AH. Health literacy and infectious diseases: why does it matter? *Int J Infect Dis.* 2016;43:103-110. doi:10.1016/j.ijid.2015.12.019
28. Lorini C, Santomauro F, Donzellini M, et al. Health literacy and vaccination: a systematic review. *Hum Vaccin Immunother.* 2018;14(2): 478-488. doi:10.1080/21645515.2017.1392423
29. Kopsaftis Z, Wood-Baker R, Poole P. Influenza vaccine for chronic obstructive pulmonary disease (COPD). *Cochrane Database Syst Rev.* 2018;6(6):CD002733. doi:10.1002/14651858.CD002733.pub3
30. Schmid P, Rauber D, Betsch C, Lidolt G, Denker ML. Barriers of influenza vaccination intention and behavior-a systematic review of influenza vaccine hesitancy, 2005-2016. *PLoS One.* 2017;12(1):e0170550. doi:10.1371/journal.pone.0170550
31. Jarrett C, Wilson R, O'Leary M, Eckersberger E, Larson HJ; SAGE Working Group on Vaccine Hesitancy. Strategies for addressing vaccine hesitancy-a systematic review. *Vaccine.* 2015;33(34):4180-4190. doi:10.1016/j.vaccine.2015.04.040

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1762045
J Health Sci Med. 2025;8(5):934-938

A retrospective multicenter study on solitary fibrous tumors: investigation of clinical features and management approaches

Serkan Yaşar[1], Feride Yılmaz[1], Gözde Yazıcı[2], Alper Kahvecioğlu[2], Yüksel Ürün[3], Hatice Bölek[3], Hilal Karakaş[4], Didem Şener Dede[4], Nuriye Özdemir[5], Volkan Aslan[5], Öztürk Ateş[6], Ahmet Kadıoğlu[6], Çağatay Arslan[7], Furkan Günen[8], Emin Tamer Elkıran[8], Serkan Akın[1], Fatma Alev Türker[1], Ömer Dizdar[1]

[1]Department of Medical Oncology, Cancer Institute, Hacettepe University, Ankara, Turkiye
[2]Department of Radiation Oncology, Cancer Institute, Hacettepe University, Ankara, Turkiye
[3]Division of Medical Oncology, Department of Internal Medicine, Faculty of Medicine, Ankara University, Ankara, Turkiye
[4]Department of Medical Oncology, Ankara Bilkent City Hospital, Ankara, Turkiye
[5]Division of Medical Oncology, Department of Internal Medicine, Faculty of Medicine, Gazi University, Ankara, Turkiye
[6]Department of Medical Oncology, Dr. Abdurrahman Yurtaslan Ankara Oncology Training and Research Hospital, University of Health Sciences, Ankara, Turkiye
[7]Department of Medical Oncology, İzmir University of Economics Medical Point Hospital, İzmir, Turkiye
[8]Division of Medical Oncology, Department of Internal Medicine, Faculty of Medicine, İnonu University, Malatya, Turkiye

Cite this article as: Yaşar S, Yılmaz F, Yazıcı G, et al. A retrospective multicenter study on solitary fibrous tumors: investigation of clinical features and management approaches. *J Health Sci Med.* 2025;8(5):934-938.

Received: 18.08.2025 ◆ **Accepted**: 04.09.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: Solitary fibrous tumors are exceedingly rare neoplasms, for which surgical resection remains the cornerstone of treatment; however, long-term surveillance is imperative due to the significant risk of recurrence. We aimed to evaluate tumor characteristics and treatment responses in this multicenter study.

Methods: Sixty-eight adult patients with histologically confirmed solitary fibrous tumor (SFT) diagnosed between 2010 and 2024 from seven cancer centers in Turkiye were included. Clinical and pathological features along with treatment outcomes were assessed.

Results: The majority of patients (65/68 pts, 95%) presented with localized disease and underwent primary resection, while 3 patients had metastatic disease at the time of diagnosis. Further analyses were performed only in patients with localized disease. Median follow-up time was 41.7 months. All patients underwent primary surgery; R0 resection was achieved in most patients (66.2%). Five -year DFS and OS were 67% and 87%, respectively. DFS of extrameningeal SFTs was further evaluated according to clinicopathological variables. Low Ki-67 Labeling Index, low mitotic rate, and younger age were found to be associated with longer DFS. Response rates to first-line systemic therapy after recurrence was 27% with chemotherapy and while no objective responses were achieved with pazopanib. Two patients experienced disease control for up to 11 months with pazopanib. This study is limited by the relatively small sample size and the heterogeneity arising from the inclusion of both meningeal and extrameningeal tumors located at different anatomical sites.

Conclusion: Localized SFTs usually have favorable prognosis. Patients with extrameningeal SFTs with younger age, low Ki-67 Index and low mitotic rate have longer survival. Once relapsed, repeat surgical resection should be pursued when possible. Response rates to systemic treatment is generally low.

Keywords: Sarcoma, surgery, pazopanib, solitary fibrous tumors, scoring systems

INTRODUCTION

Solitary fibrous tumors (SFT) arise from pericytes located around capillary walls.[1] It is an exceedingly rare tumor that reported incidence rate <1 case/million people/year.[2,3] Median age is 54 years and the incidence ratio of men to women is equal. Most common localization is extremities-trunk (27.9%), followed by intracranial (25%), intrathoracic (22.1%), intraabdominal (14.7%) and extracranial head/neck (10.3%).[4,5]

Various scoring systems were created to determine prognosis and to tailor treatment. Although surgical treatment is the main treatment method, the effects of radiotherapy and

Corresponding Author: Serkan Yaşar, drserkanyasar1522@gmail.com

systemic treatment have also been investigated in selected cases in previous studies. Data on treatment of this rare tumor is limited and usually confined to retrospective series and small size phase 2 trials. In this multicenter retrospective study, we have evaluated patient and tumor characteristics and real-world treatment outcomes in patients with SFTs.

METHODS

The study protocol complied with the tenets of the Declaration of Helsinki. The study was approved by the Hacettepe University Faculty of Medicine Ethics Committee (Date: 24.04.2025, Decision No: 2025/09-30). As this was a retrospective study involving anonymized patient data, the requirement for informed consent was waived by the ethics committee.

This study is a retrospective observational study based on the analysis of previously collected clinical data. Patients with histologically confirmed SFT diagnosed between 2010-2024 from seven different centers were included. Pathological diagnoses were made at each participating center by experienced pathologists, and no central pathology review was performed. Patient and tumor characteristics, local and systemic treatments and outcomes were recorded**.** Overall survival (OS) and disease-free survival (DFS) analyses were carried out with Kaplan-Meier analyses and compared with log-rank test. Predictors of DFS were analyzed and patients were grouped as low- and high risk based on the presence of these factors. All statistical tests were two-tailed, and a p-value <0.05 was considered statistically significant. The statistical analyses were performed with SPSS, version 25.0 (IBM Inc., Armonk, NY, USA).

RESULTS

Sixty-eight patients with SFT were included. Of the patients, 39 (57.4%) were male and 29 (42.6.1%) were female. The median age at diagnosis was 54 years (IQR 42-62). The most common primary sites of SFT were extremities-trunk (27.9%) and intracranial (25%) (**Table 1**).

Table 1. Primary tumor sites

Primary site	n=68	%
Extremities/trunk	19	27.9
Intracranial	17	25.0
Intrathoracic	15	22.1
Intra-abdominal	10	14.7
Extracranial head and neck	7	10.3

At the time of diagnosis, 65 of 68 patients (95%) had localized disease. Further analyses were performed only in patients with localized disease. Median follow-up time was 41.7 months (IQR: 21-71.6).

Median DFS was 95 months (95% CI, 55.07-135.9) (**Figure 1**), with a 5-year DFS rate of 67%, mOS was not reached, and the 5-year OS rate was 87% in patients with localized disease.

Surgery was performed in all these patients. R0 resection was performed in 43 patients (66.2%), R1 resection in 13 patients (20%) and R2 resection in 9 patients (13.8%). Median DFS was 95.5 months in the R0/R1 group, whereas it was 68.3 in the R2 group (p=0.47). 5-year OS of patients with R0/R1 was 89%, while the 5-year survival of the R2 group was 77% (p=NS).

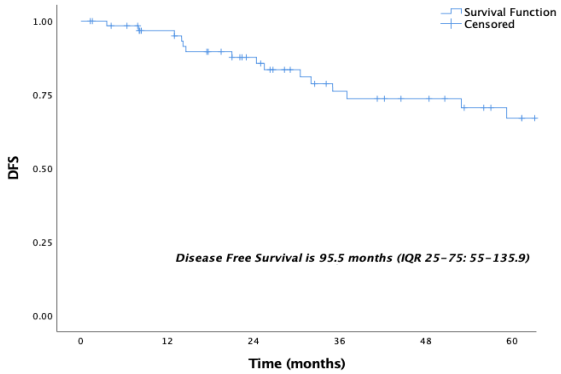

Figure 1. Kaplan-Meier curve for disease-free survival in patients with localized SFTs after definitive treatment

SFTs: Solitary fibrous tumors

28 patients did not receive any adjuvant treatment. 30 patients received adjuvant radiotherapy (RT) at doses between 34-64 gray (Gy). The mDFS was 76.9 months without adjuvant RT and 95.5 months with RT respectively (p=0.75). Similarly, adjuvant RT did not provide any survival benefit in subgroup with R1 resection (5 yr DFS 70 vs. 75%, respectively). All patients in the R2 resection group received RT, and 5-year DFS rate was 70%. Seven patients received adjuvant radiotherapy plus chemotherapy (CT), 3 of which was anthracycline based. No recurrences were seen in this subgroup. However, it should be noted that the number of patients in the adjuvant radiotherapy subgroup was limited, and therefore these results should be interpreted with caution.

On follow-up, recurence occurred in 21 patients, including 11 cases of local recurrence and 10 cases of distant metastasis. Among patients with local recurrence, nine underwent resection, followed by RT in 3 patients and CT in 4 patients. Two patients had unresectable tumors, one received RT and one received CT. All patients with systemic recurrence received systemic treatment, and two underwent palliative surgery prior to treatment

Patients with extrameningeal SFTs were assessed for predictors of DFS and three potentially predictive factors were identified; the group with a low Ki-67 Labeling Index (<10 vs. ≥10, mDFS 109 vs. 65 months, p<0.01), low mitotic rate (<4 vs. ≥4 /HPF, mDFS 159 vs 68.3 months, p=0.013), younger age (<55 vs. ≥55 years, mDFS 109 vs 53 months, p=0.04) had a longer DFS. Subsequently, we classified the patients into two groups based on the presence of these risk factors; group A (low risk group); patients with 0 or 1 risk factor, and group B (high risk group); patients with 2 or 3 risk factors. Patients in group A had significantly longer DFS compared with group B (152 months in group A vs 63 months in group B, p=0.028) (**Figure 2**).

Fifteen patients were administered systemic therapy following the recurrence. Of these patients, 11 received cytotoxic chemotherapy (mainly anthracycline and ifosfamide based therapies) and 4 patients received pazopanib. Treatment

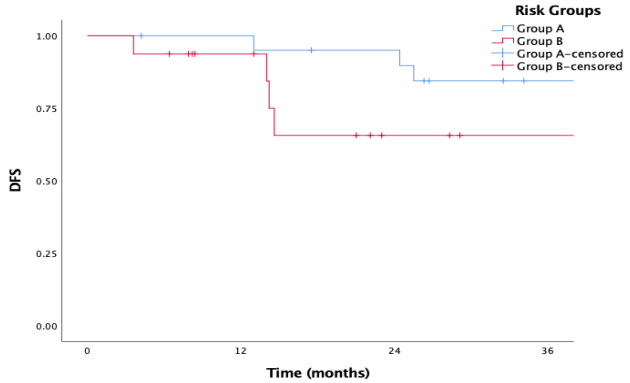

Figure 2. Kaplan Meier curve for disease free survival according to RSS. (group A includes 0 or 1, group B includes 2 or 3 risk factors) (p=0.028)

RSS: Risk scoring systems

responses are shown in Table 2. Overall response rate was 27% (3/11 pts) with chemotherapy and 0% with pazopanib. Median PFS of patients who received first-line treatment was 4.3 months (95% CI 2.27-6.32).

DISCUSSION

The treatment of SFTs, which requires a multidisciplinary approach, includes surgery, radiotherapy and systemic therapy as therapeutic options.[6,7] Surgery is the mainstay of treatment.[5,8] There is no established role for adjuvant systemic treatment. Adjuvant radiotherapy may be considered after R1 resection and in locally recurrent cases. Retrospective series have shown limited activity with standard sarcoma chemotherapy protocols and tyrosine kinase inhibitors in the advanced disease setting.

Due to the unpredictable clinical course of extrameningeal SFTs, various scoring systems incorporating age, mitosis rate, Ki-67 Labeling Index, necrosis, and tumor size have been developed to predict their prognosis, as demonstrated in the studies by Demicco et al.,[9] Salas et al.,[10] and Georgiesh et al.[11] (Table 3).

Table 3. Comparison of risk scoring systems

Demicco et al.[8]	Demicco et al.[9]	Georgiesh et al.[11]	Salas et al.[10]
Extra-meningeal (modified D-score)	**Extra-meningeal**	**Extra-meningeal**	**Extra-meningeal**
Necrosis	Size	Mitotic Index	Mitotic Index
Mitotic Index	Age	Necrosis	Anatomic site
Age	Mitotic Index	Gender	Age
Size			

Demicco model, also known as the 'D-score' or 'MDACC score,' is the most commonly utilized clinical tool that integrates multiple clinicopathologic variables and informs the risk of metastatic relapse[8,12] In their study, 110 patients with non-meningeal SFT, of whom 103 were determined to be eligible for resection. This study demonstrated that age, mitotic index, and tumor size were significant predictors of both time to metastasis and disease-specific mortality. It was also reported that tumors originating in the pleural and abdominal regions were diagnosed at larger sizes, and pleural tumors were detected in older patients. Based on the data obtained, a three-tiered risk stratification model was developed. In this model, neither metastasis nor death occurred in the low-risk group during 50 months of follow-up. However, in the high-risk group, the 5- and 10-year metastasis-free survival rates were 15% and 0%, respectively, while in the intermediate group, these rates were 77% and 64%. A statistically significant difference was observed between the low- and high-risk groups we defined; the mDFS was 152 months versus 63 months (p=0.028).[13] Another prognostic model developed by Tatiana Georgiesh at al. investigated 100 cases of extrameningeal SFT and identified mitotic rate, the presence of necrosis, and gender as independent risk factors for prognosis. The group with a necrosis rate greater than 50% had higher recurrence rates and poorer survival outcomes. Also, male gender was associated with poorer outcomes. Another factor was the mitosis rate in this study and the findings were consistent

Table 2. First line treatment options and responses

Patient	Treatment	Line	Best response	PFS (months)	Primary site
1	Temozolomide-bevacizumab	1	PD	2.5	Extremity and trunk
2	Temozolomide-bevacizumab	1	PR	3.5	Intra-abdominal
3	Temozolomide-bevacizumab	1	SD	3	Intracranial
4	Ifosfamide-adriamycin	1	PR	2	Intracranial
5	Ifosfamide-adriamycin	1	SD	3.4	Extremity and trunk
6	Ifosfamide-adriamycin	1	PD	2.1	Intrathoracic
7	Ifosfamide-adriamycin	1	PR	2	Intra-abdominal
8	Vincristine-temozolomide	1	PD	2.6	Intra-abdominal
9	Cisplatin-etoposide	1	PD	3.2	Head and neck
10	Single agent adriamycin	1	PD	1.9	Extremity and trunk
11	Ifosfamide-etoposide	1	PD	1.7	Extremity and trunk
12	Pazopanib	1	PD	4.3	Intra-abdominal
13	Pazopanib	1	SD	11.2	Extremity and trunk
14	Pazopanib	1	SD	11	Intra-abdominal
15	Pazopanib	1	PD	3	Intrathoracic

PFS: Progression-free survival, PD: Progressive disease, PR: Partial response, SD: Stable disease

with our study. The 10-year relapse-free interval was 77% in the group with mitosis <4 HPF while it was 46% for the group with mitosis ≥4 HPF.[11] Some researchers have investigated the benefit of using the Ki-67 Labeling Index in the scoring system instead of necrosis to better predict prognosis.[14,15] One such study, conducted by Sugita et al.,[16] demonstrated the prognostic importance of the Ki-67 Labeling Index. 5-year metastasis-free survival was 50% in the patient group with a Ki-67 Labeling Index ≥10, while it reached 100% in the group with a Ki-67 Labeling Index <10.

Margin status was also found to associated with recurrence risk in some studies but contradictory studies also exist. A study involving one of the largest cohorts, comprising 303 SFT patients, identified a significantly higher risk of local recurrence (HR=10.0) in patients with positive margins.[9] However, several other studies have reported that local recurrence, metastatic recurrence, and overall survival are not associated with margin status.[10]

We also categorized our patients into low- and high-risk groups based on Ki-67, age, and mitosis, and showed the effect of the risk score on prognosis. In our follow-up after definitive treatment, the high-risk group demonstrated a shorter median disease-free survival (mDFS), as shown in the Kaplan-Meier curve (**Figure 2**). Such a risk stratification can be used for tailoring adjuvant treatment strategies if further assessed in prospective studies.

Postoperative approaches can be discussed based on risk models and according to the surgical margins, Previous studies showed improved local control with RT in high-risk SFTs. However, its contribution to overall survival is not clear.[6,17,18] In a study conducted by Schöffski et al.,[19] the local recurrence rate in the cohort that underwent primary surgery followed by radiotherapy was 7.1%. In the overall population, this rate was 26.7%. Another retrospective study of 549 patients investigating perioperative radiotherapy found that the risk of local relapse was significantly reduced, particularly in the group with higher mitotic rates and positive or inadequate surgical margin. (HR, 0.19; %95 CI, 0.04-0.84; p=0.029).[17] No survival benefit of adjuvant RT was observed; however, these results should be interpreted with caution due to the small number of patients, particularly in the high-risk subgroup with R1 resection.

In case of metastatic disease, systemic chemotherapy is used, but a standard approach to treatment has not yet been established. Available data regarding the response of SFTs to standard chemotherapy are limited. The results obtained from small case series, retrospective studies, present conflicting findings.[6] In many case series, systemic treatment options were extrapolated from soft tissue sarcomas. Anthracycline-based treatments were assessed in first-line therapy by Schoffski et al.,[19] with an ORR of 13.3% and mPFS of 4.8 months. In a study conducted by Levard et al,[7] the similar effectiveness of anthracyclines in the first line was shown (partial response in 2 of 19 patients). Moreover, it was also noted in the study that doxorubicin was more effective than other cytotoxic agents when administered both in combination and alone. The limited effectiveness of conventional chemotherapy has also been demonstrated in subsequent lines. Studies investigating the efficacy of antiangiogenic agents such as sunitinib, sorafenib, pazopanib, and temozolomide-bevacizumab have shown promising results in controlling disease progression and improving outcomes in patients with SFTs. For instance, case reports have demonstrated prolonged disease control for up to 30 months with the use of these agents.[6] Furthermore, retrospective studies and phase I/II trials have provided additional evidence supporting the effectiveness of antiangiogenic therapies in SFT management. Martin-Broto et al.[20] showed partial responses with pazopanib in typical advanced SFTs according to Choi criteria in 58% of patients and stable disease in 39%.[21] We also evaluated the responses of patients who relapsed after definitive treatment and received first-line therapies. Our analysis revealed that 3 of 11 patients (27%) had partial response to chemotherapy Although no objective response was observed with pazopanib, disease control for up to 11 months was achieved in 2 of our patients in the first line. Anti-angiogenic agents including VEGF-TKIs and bevacizumab combinations seem to be more effective and less toxic treatment options.[21]

Limitations

Our study had some limitations. The patient number was limited due to the very low incidence, and patients were heterogeneous in terms of site of origin and other characteristics. In addition, pathological diagnoses were made at each participating center by experienced pathologists, and no central pathology review was performed. Nevertheless, we were able to create a prognostic model and evaluate the activity of various treatment options in both the adjuvant and advanced settings.

CONCLUSION

SFT is an exceedingly rare and highly heterogeneous type of sarcoma, with varied clinical, molecular, and treatment responses. For localized disease, marginal surgery appears important for survival, while postoperative radiotherapy and sequential chemotherapy and radiotherapy may help delay disease recurrence. Antiangiogenic agents show potential, both in combination with chemotherapy and as monotherapy (e.g., VEGF-TKIs), although current evidence remains limited. Prognostic models may help identify high-risk patients who could benefit from additional treatment after surgery to reduce relapse risk, while factors such as low Ki-67 Index, low mitotic rate, and younger age provide useful insights for tailoring follow-up and assessing the need for adjuvant therapy in SFT patients. Given the rarity of this sarcoma subgroup, prospective randomized studies are challenging, and our study should be considered hypothesis-generating, aiming to provide preliminary guidance for clinicians.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was carried out with the permission of the Hacettepe University Faculty of Medicine Ethics Committee (Date: 24.04.2025, Decision No: 2025/09-30).

Informed Consent

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

Referee Evaluation Process
Externally peer-reviewed.

Conflict of Interest Statement
The authors have no conflicts of interest to declare.

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

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

Data Availability
The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

REFERENCES

1. Xiao X, Li Q, Ju Y. Primary hemangiopericytoma in parietal bone: literature review and case report. *Am J Transl Res*. 2023;15(2):1254-1258.
2. Kinslow CJ, Bruce SS, Rae AI, et al. Solitary-fibrous tumor/hemangiopericytoma of the central nervous system: a population-based study. *J Neurooncol*. 2018;138(1):173-182. doi:10.1007/s11060-018-2787-7
3. Kinslow CJ, Wang TJC. Incidence of extrameningeal solitary fibrous tumors. *Cancer*. 2020;126(17):4067. doi:10.1002/cncr.33057
4. Ronchi A, Cozzolino I, Zito Marino F, et al. Extrapleural solitary fibrous tumor: a distinct entity from pleural solitary fibrous tumor. An update on clinical, molecular and diagnostic features. *Ann Diagn Pathol*. 2018;34:142-150. doi:10.1016/j.anndiagpath.2018.01.004
5. Goodlad JR, Fletcher CD. Solitary fibrous tumour arising at unusual sites: analysis of a series. *Histopathology*. 1991;19(6):515-522. doi:10.1111/j.1365-2559.1991.tb01499.x
6. de Bernardi A, Dufresne A, Mishellany F, Blay JY, Ray-Coquard I, Brahmi M. Novel therapeutic options for solitary fibrous tumor: antiangiogenic therapy and beyond. *Cancers (Basel)*. 2022;14(4):1064. doi:10.3390/cancers14041064
7. Levard A, Derbel O, Méeus P, et al. Outcome of patients with advanced solitary fibrous tumors: the Centre Léon Bérard experience. *BMC Cancer*. 2013;13:109. doi:10.1186/1471-2407-13-109
8. Demicco EG, Wagner MJ, Maki RG, et al. Risk assessment in solitary fibrous tumors: validation and refinement of a risk stratification model. *Mod Pathol*. 2017;30(10):1433-1442. doi:10.1038/modpathol.2017.54
9. Demicco EG, Griffin AM, Gladdy RA, et al. Comparison of published risk models for prediction of outcome in patients with extrameningeal solitary fibrous tumour. *Histopathology*. 2019;75(5):723-737. doi:10.1111/his.13940
10. Salas S, Resseguier N, Blay JY, et al. Prediction of local and metastatic recurrence in solitary fibrous tumor: construction of a risk calculator in a multicenter cohort from the French Sarcoma Group (FSG) database. *Ann Oncol*. 2017;28(8):1979-1987. doi:10.1093/annonc/mdx250
11. Georgiesh T, Boye K, Bjerkehagen B. A novel risk score to predict early and late recurrence in solitary fibrous tumour. *Histopathology*. 2020;77(1):123-132. doi:10.1111/his.14078
12. Reisenauer JS, Mneimneh W, Jenkins S, et al. Comparison of risk stratification models to predict recurrence and survival in pleuropulmonary solitary fibrous tumor. *J Thorac Oncol*. 2018;13(9):1349-1362. doi:10.1016/j.jtho.2018.05.040
13. Demicco EG, Park MS, Araujo DM, et al. Solitary fibrous tumor: a clinicopathological study of 110 cases and proposed risk assessment model. *Mod Pathol*. 2012;25(9):1298-1306. doi:10.1038/modpathol.2012.83
14. Rodriguez-Gonzalez M, Novoa NM, Gomez MT, Garcia JL, Ludeña D. Factors influencing malignant evolution and long-term survival in solitary fibrous tumours of the pleura. *Histol Histopathol*. 2014;29(11):1445-1454. doi:10.14670/HH-29.1445
15. Krsková L, Odintsov I, Fabián O, Hroudová P, Mrhalová M. Determination of biological behavior of solitary fibrous tumors: correlation of expression of Ki-67, TPX2 and TERT mRNA subunit level and NAB2-STAT6 fusion compared to morphological aspects of SFTs. *Neoplasma*. 2022;69(1):28-35. doi:10.4149/neo_2021_210511N642
16. Sugita S, Segawa K, Kikuchi N, et al. Prognostic usefulness of a modified risk model for solitary fibrous tumor that includes the Ki-67 Labeling Index. *World J Surg Oncol*. 2022;20(1):29. doi:10.1186/s12957-022-02497-2
17. Haas RL, Walraven I, Lecointe-Artzner E, et al. Extrameningeal solitary fibrous tumors-surgery alone or surgery plus perioperative radiotherapy: a retrospective study from the global solitary fibrous tumor initiative in collaboration with the sarcoma patients EuroNet. *Cancer*. 2020;126(13):3002-3012. doi:10.1002/cncr.32911
18. Bishop AJ, Zagars GK, Demicco EG, Wang WL, Feig BW, Guadagnolo BA. Soft tissue solitary fibrous tumor: combined surgery and radiation therapy results in excellent local control. *Am J Clin Oncol*. 2018;41(1):81-85. doi:10.1097/COC.0000000000000218
19. Schöffski P, Timmermans I, Hompes D, et al. Clinical presentation, natural history, and therapeutic approach in patients with solitary fibrous tumor: a retrospective analysis. *Sarcoma*. 2020;2020:1385978. doi:10.1155/2020/1385978
20. Martin-Broto J, Stacchiotti S, Lopez-Pousa A, et al. Pazopanib for treatment of advanced malignant and dedifferentiated solitary fibrous tumour: a multicentre, single-arm, phase 2 trial. *Lancet Oncol*. 2019;20(1):134-144. doi:10.1016/S1470-2045(18)30676-4
21. Janik AM, Terlecka A, Spałek MJ, et al. Diagnostics and treatment of extrameningeal solitary fibrous tumors. *Cancers (Basel)*. 2023;15(24):5854. doi:10.3390/cancers15245854

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1750794
J Health Sci Med. 2025;8(5):939-945

Ultrasound-derived reference values for neonatal liver and spleen dimensions: a retrospective cohort study

Baran Cengiz Arcagök[1], Akan Yaman[2]

[1]Division of Neonatology, Department of Pediatrics, Faculty of Medicine, Acıbadem University, İstanbul, Turkiye
[2]Department of Child Development, Vocational School of Health Services, Nişantaşı University, İstanbul, Turkiye

Cite this article as: Arcagök BC, Yaman A. Ultrasound-derived reference values for neonatal liver and spleen dimensions: a retrospective cohort study. *J Health Sci Med.* 2025;8(5):939-945.

Received: 25.07.2025 ◆ **Accepted**: 05.09.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: To determine liver and spleen sizes in preterm and term neonates using ultrasonography during the first week of life, and to examine their associations with gestational age, birth weight, length, gender, and postmenstrual week.

Methods: In this retrospective study, 112 neonates (66 preterm, 46 term) admitted to a tertiary NICU between June and December 2020 were included. Only infants with normal liver function and without conditions affecting organ size were analyzed. Liver and spleen dimensions were measured by ultrasonography within the first seven postnatal days. Correlations were assessed using Pearson, Bayesian Kendall's, and Bayesian Pearson tests. Regression analyses and comparisons with published data were also performed.

Results: In preterm infants, liver size correlated strongly with gestational age (r=0.825) and spleen size with birth weight (r=0.777). In term infants, liver size correlated with birth length (r=0.491) and spleen size with birth weight (r=0.495). Each 1-week increase in gestational age was associated with a 1.8 mm increase in liver size in preterm infants, while each 100 g increase in birth weight increased spleen size by 0.8 mm in both groups.

Conclusion: Liver and spleen sizes are closely linked to gestational age, birth weight, and length in neonates. Population-specific percentile references are recommended to improve clinical assessment accuracy.

Keywords: Neonate, ultrasound, visceral organs, birth weight, organ size

INTRODUCTION

Traditionally, neonatal liver and spleen sizes are assessed by palpation and percussion. However, these methods may fail to detect subtle changes reliably.[1,2] Ultrasonography is now the preferred method because it is non-invasive, safe, accurate, and practical for routine clinical use.[3]

Liver and spleen sizes vary according to gestational age, birth weight, length, and gender.[4-6] In addition, infections, congenital conditions, and malignancies can affect these dimensions, highlighting the need for precise measurement for diagnosis and clinical management.[1,2]

Although some reference ranges exist, there is no universally accepted standard. There is also uncertainty regarding which anthropometric parameters best correlate with liver and spleen sizes in neonates. This study aimed to identify the most reliable correlates of liver and spleen dimensions during the early neonatal period and to provide population-specific percentile values for clinical use.

METHODS

Ethics

The study was approved by the Marmara University Faculty of Medicine Clinical Researches Ethics Committee (Date: 07.01.2022, Decision No: 09.2022.23), and was conducted in accordance with the principles of the Declaration of Helsinki. Written informed consent was obtained from the parents of all participants.

This retrospective study included neonates who underwent abdominal ultrasonography in the NICU of a tertiary hospital between June and December 2020. Ultrasonography was initially performed to evaluate urological conditions such as posterior urethral valve, vesicoureteral reflux, and hydronephrosis.

Inclusion and Exclusion Criteria

Neonates were eligible for the study if they met all of the following criteria: gestational age between 24 and 42 weeks, hemodynamically stable without the need for inotropic support, normal liver function tests, and no clinical or laboratory evidence of hepatic or splenic pathology. Written informed consent was obtained from parents or legal guardians.

Exclusion criteria were as follows: congenital anomalies or chromosomal abnormalities affecting liver or spleen size; clinical or laboratory evidence of infection, hemolytic

Corresponding Author: Baran Cengiz Arcagök, baranarcagok@hotmail.com

disease, coagulopathy, or any systemic illness potentially affecting organ dimensions; intrauterine growth restriction beyond ±2 standard deviations according to Fenton growth charts; infants born to consanguineous parents; neonates receiving medications known to affect liver or spleen size (e.g., hepatotoxic drugs); and any condition that could interfere with accurate measurement of liver or spleen dimensions, including severe ascites, abdominal wall defects, or technical difficulties during ultrasonography.

All measurements were performed within the first seven days of life by a single radiologist using a General Electric VIVID 3 Expert device. Liver and spleen dimensions were measured in the craniocaudal plane. Gestational age, postmenstrual week (PMW) at the time of ultrasonography, gender, birth weight, and birth length were recorded for each infant.

We estimated that a minimum of 45 patients would be required for a linear regression model with three potential confounders to detect an effect size of f^2=0.35, with an alpha error rate of 0.05 and a power (1-β) of 0.90.

Statistical Analysis

We presented data as mean±standard deviation or median (interquartile range, IQR), depending on the distribution pattern. Non-normally distributed data were transformed to approximate a normal distribution prior to performing Pearson tests. Generalized linear regression models were created using statistically significant variables, and confounders were assessed using the backward elimination method by including correlated factors as dummy variables. Multicollinearity was also assessed using the variance inflation factor. For regression analysis models, a p-value <0.01 was considered statistically significant. Estimated marginal means tables and graphs were subsequently constructed.

Bayesian Kendall's and Bayesian Pearson tests were used to evaluate the probability of the H1 (association hypothesis) and H0 (independence/no association hypothesis), with a stretched beta prior width set to 1. BF10 between 1-3 represents anecdotal evidence for correlation hypothesis, 3-10 moderate evidence, 10-30 strong evidence, >30 very strong evidence; where between 0.33-1 represents anecdotal evidence for no association, 0.1-0.33 moderate evidence, 0.03-0.1 strong evidence, and <0.03 very strong evidence. We also calculated intraclass correlation coefficients, which resulted in 0.651 for preterm liver and spleen sizes, and 0.674 for term neonates.

To compare our data with findings reported in other publications, we applied the estimated independent mean difference function or the one-sample Wilcoxon test. For these comparisons, a p-value <0.05 was considered statistically significant. All statistical analyses were performed using the JAMOVI 2.3.18 statistical software, including the esci and jsq extensions.

RESULTS

A total of 119 neonates (70 preterm and 49 term) underwent ultrasonographic measurements. Seven infants (four term and three preterm) were excluded based on the predefined criteria (normal liver function, absence of conditions affecting organ size, gestational age and birth weight within ±2 SD of post-conceptual age, and hemodynamic stability), leaving 112 neonates (66 preterm, 46 term) for final analysis. The demographic and anthropometric characteristics of these neonates are summarized in Table 1.

We assessed correlations between liver and spleen sizes and birth week, birth weight, birth length, gender, PMW, and postnatal day across all infant groups (Table 2). In preterm infants, liver and spleen sizes were significantly correlated with birth week, birth weight, birth length, and PMW. Specifically, liver size showed a strong correlation with birth length (Kendall's tau b=0.491, p<0.01, BF_{10}=22.62), while spleen size correlated strongly with birth weight (Kendall's tau b=0.495, p<0.01, BF_{10}=68.04).

In term infants, moderate evidence indicated a correlation between gender and spleen size (BF_{10}=8.08). Other observations included no gender effect on spleen size in preterm infants, no influence of birth week on spleen size in term infants, and no effect of postnatal day on liver or spleen sizes in either group, consistent with the independence hypothesis.

Liver and Spleen Dimensions in Preterm Infants

In preterm infants, liver and spleen dimensions were significantly correlated with birth week, birth weight, birth length, and postmenstrual week (PMW). These factors were included in a multivariate analysis using a generalized linear

Table 1. Comparison of demographic and anthropometric measurements in preterm and term newborns

	Preterm (n=66)	Term (n=46)	All newborns (n=112)
Gender % (n)	51.5% (n=34) male 48.5% (n=32) female	65.2% (n=30) male 34.8% (n=16) female	57.1% (n=64) male 42.9% (n=48) female
PN day (days)	1 (1-2) days	2 (1-4) days	2 (1-3) days
Birth week (weeks)	33.7 (31.0-35.0)	38.4 (38.1-39.9)	34.9 +/- 4.1
PMW (weeks)	33.9 (31.0-35.1)	39.0±1.1	35.2 +/- 4.2
Birth weight (gram)	1866±643	3241±429	2431 +/- 882
Birth length (cm)	43 (41-47)	49.9±2.8	45.2 +/-5.9
Liver dimension (mm)	47.3±7.4	59.2±7.8	52.2 +/- 9.5
Spleen dimension (mm)	34.0±6.6	44.6±6.4	38.4 +/- 8.3

PN day: Postnatal day at the assessment time, PMW: Postmenstrual week at the assessment time. A total of 119 neonates were initially assessed, and 7 infants (4 term, 3 preterm) were excluded based on predefined criteria, leaving 112 neonates for analysis. Data are presented as mean±standard deviation and median (interquartile range).

Table 2. Correlation coefficients for liver and spleen dimensions in preterm and term infants

	Liver dimension				Spleen dimension			
	Preterm		Term		Preterm		Term	
	Correlation co-efficient	BF10	Correlation co-efficient	BF10	Correlation co-efficient	BF10	Correlation co-efficient	BF10
Gender	0.124	0.462^{AI}	0.137	0.463^{AI}	0.056	0.2^{MI}	0.283^{*}	8.08^{M}
Birth week	0.825^{***}	$>1000^{VS}$	0.254	0.747^{AI}	0.743^{***}	$>1000^{VS}$	0.148	0.293^{MI}
PMW	0.803^{***}	$>1000^{VS}$	0.285	1.094^{A}	0.757^{***}	$>1000^{VS}$	0.219	0.517^{AI}
Birth weight	0.791^{***}	$>1000^{VS}$	0.349^{*}	2.828^{A}	0.777^{***}	$>1000^{VS}$	0.495^{***}	68.039^{S}
Birth length	0.736^{***}	$>1000^{VS}$	0.491^{**}	22.62^{S}	0.704^{***}	$>1000^{VS}$	0.188	0.374^{AI}
Postnatal day	0.138	0.155^{MI}	0.042	0.195^{MI}	0.182	2.4^{A}	0.148	0.293^{MI}

PMW: Postmenstrual week. Kendall's Tau-b is used for gender and Pearson's rho test for others. Non-normal data were transformed into normal distribution before being subjected to analysis. *p<0.05, **p<0.01, ***p<0.001. A: Anecdotal evidence for H1 (correlation) hypothesis, M: Moderate evidence for H1 (correlation) hypothesis, S: Strong evidence for H1 (correlation) hypothesis, AI: Anecdotal evidence for H0 (independence) hypothesis, MI: Moderate evidence for H0 (independence) hypothesis, BF: Bayes factor

model, with non-normally distributed data transformed into normal distributions and analyzed as dummy variables. Confounding factors were addressed using the backward elimination method.

The analysis revealed that each 1-week increase in gestational age (GW) was associated with a 1.8 mm increase in liver dimension (95% CI: 1.3-2.2; AIC: 403.1; BIC: 407.5; R^2: 0.691; Chi-squared/DF: 0.371; eta squared from the general linear model: 0.657). Similarly, each 100-gram increase in birth weight corresponded to a 0.8 mm increase in spleen dimension (95% CI: 0.6-1.1; AIC: 404.8; BIC: 409.2; R^2: 0.669; Chi-squared/DF: 0.398; eta squared from the general linear model: 0.603). The estimated marginal means are presented in Table 3, and the corresponding graph is shown in Figure 1 and 2.

Table 3. Estimated marginal means for liver and spleen sizes stratified by birth week and weight

Birth week (weeks)	Mean liver size (mm)	Birth weight (gram)	Mean spleen size (mm)
28	40 (38-42)	1000	27 (24-29)
29	42 (40-44)	1200	34 (33-36)
30	43 (42-45)	1400	34 (33-36)
31	45 (43-47)	1600	32 (30-33)
32	47 (45-48)	1800	33 (32-35)
33	49 (47-50)	2000	35 (34-36)
34	50 (48-52)	2200	37 (35-38)
35	52 (50-54)	2400	38 (36-40)
36	54 (51-56)	2600	40 (38-42)
R^2:0.669 (GLM)		R^2: 0.625 (GLM)	

GLM: Generalized linear model. Results are presented as rounded values for clarity. The 95% confidence intervals are provided in parentheses.

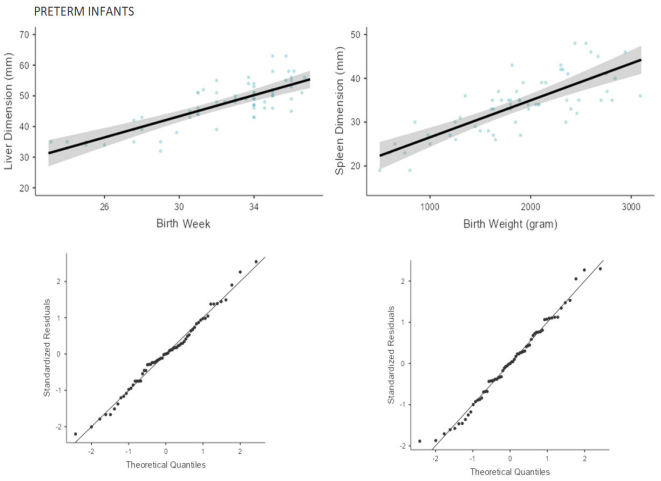

Figure 1. Liver and spleen dimensions in preterm infants and their associations with birth week and birth weight. The x-axis represents liver and spleen sizes, and the y-axis represents gestational age (weeks) or birth weight. The gray shaded area indicates the 95% confidence interval of the mean liver and spleen sizes. Standardized residual plots are also presented

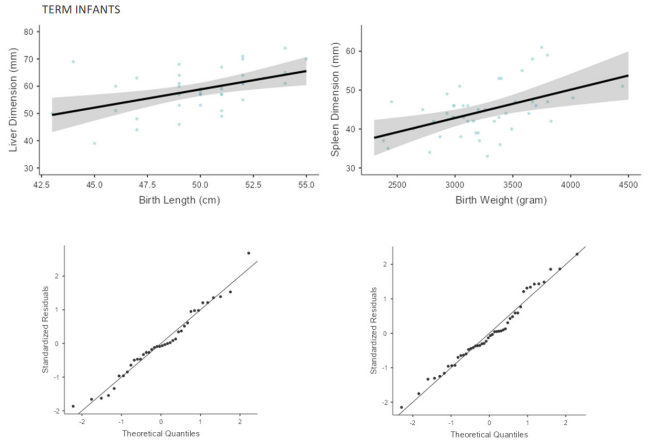

Figure 2. Liver and spleen dimensions in preterm and term infants and their associations with birth weight and birth length. The x-axis represents liver and spleen sizes, and the y-axis represents birth weight or birth length. The gray shaded area indicates the 95% confidence interval of the mean liver and spleen sizes. Standardized residual plots are also presented

Liver and Spleen Dimensions in Term Infants

In our study group, each 1 cm increase in birth length was associated with a 1.3 mm increase in liver dimension (95% CI: 0.5-2.2; AIC: 259.5; BIC: 262.8; R^2: 0.230; Chi-squared/DF: 0.869; eta squared from the general linear model: 0.241), and each 100 g increase in birth weight was associated with a 0.8 mm increase in spleen dimension (95% CI: 0.3-1.2; AIC: 293.2; BIC: 296.8; R^2: 0.246; Chi-squared/DF: 0.690; eta squared from the general linear model: 0.245). The estimated marginal means are provided in Table 4, and the corresponding graphs are shown in Figure 1 and 2.

Table 4. Estimated marginal means for liver and spleen sizes by birth length and weight

Birth length (cm)	Mean liver size (mm)	Birth weight (gram)	Mean spleen size (mm)
46	54 (49-58)	2400	39 (34-43)
47	55 (51-58)	2600	44 (42-46)
48	56 (53-59)	2800	41 (39-44)
49	58 (55-60)	3000	43 (41-45)
50	59 (56-61)	3200	44 (42-46)
51	60 (58-63)	3400	46 (44-48)
52	62 (58-65)	3600	47 (45-50)
53	63 (59-67)	3800	49 (45-52)
	R^2:0.230 (GLM)		R^2:0.246 (GLM)

GLM: Generalized linear model. Results are presented as rounded values for clarity. The 95% confidence intervals are provided in parentheses.

The percentiles and means of liver and spleen dimensions in preterm and term infants are presented in Table 5.

DISCUSSION

The dimensions of visceral organs in newborns vary according to anthropometric measurements[6]. Therefore, when analyzing ultrasonographic results, especially in preterm newborns, it is crucial to consider factors such as gestational age, weight, and length. Previous studies have reported varying correlation coefficients between anthropometric measurements and liver and spleen sizes in different age groups,[3,5,6,8-11] highlighting the need for percentile tables based on the parameters showing the strongest correlation.

In our study, the correlation between liver size and birth week and between spleen size and birth weight was strong in preterm infants, whereas in term infants, liver size correlated more with birth length and spleen size with birth weight, though the associations were less robust. Gender did not significantly affect liver or spleen dimensions in preterm infants, but a moderate correlation between gender and spleen size was observed in term infants.[3,5,6,12] These findings partially align with previous studies; some reported gender effects on spleen size,[3,12] while others did not.[5,6,10,11,13] The impact of gender on organ dimensions in term infants remains inconclusive and may benefit from multicenter studies with larger sample sizes.

Liver Dimensions

Regarding liver sizes, birth length had the strongest correlation in term newborns, whereas birth week was the main determinant in preterm infants. Studies on children aged 0-16 years reported stronger correlations of liver size with length (r=0.81-0.92) than with weight (r=0.74-0.86).[5,8-10] However, studies focusing on newborns, including our own, demonstrated that birth weight had the strongest correlation with liver size in preterm infants[12] and all newborns.[13] Conversely, one study indicated length as the strongest correlate in term infants,[6] which differs from our findings. This discrepancy may reflect better intrauterine growth in term infants, allowing length and visceral organs to continue developing even if weight gain is less optimal.

Our analysis revealed that each 1-week increase in gestational age among preterm infants was associated with a 1.8 mm increase in liver dimension (95% CI: 1.3-2.2), and each 100-gram increase in birth weight corresponded to a 0.8 mm increase in spleen dimension (95% CI: 0.6-1.1). In term infants, each 1 cm increase in birth length was associated with a 1.3 mm increase in liver dimension (95% CI: 0.5-2.2), and each 100 g increase in birth weight with a 0.8 mm increase in spleen dimension (95% CI: 0.3-1.2). These findings suggest that liver and spleen sizes are closely related to birth week, weight, and length in neonates, supporting the use of population-specific percentile references.

Comparisons with published literature revealed both similarities and differences in liver dimensions across populations.[3,5,6,8-15] Liver sizes in term neonates from our study were 4.8 mm smaller than a 3-month postnatal group 5, 6 mm larger than a 2-month group8, and showed minor differences compared with other studies in the 0-3 month age range.[9,14] Upper limits and mean±SD values from studies conducted in Turkiye[12,13] were comparable to our preterm infant results, but notable variability exists, likely influenced by ethnicity, geographic location, patient risk groups, and operator experience (Table 6, 7).[3,6,12,13]

Spleen Dimensions

In both preterm and term neonates, spleen size showed the strongest correlation with birth weight rather than birth length, consistent with previous studies.[6,11,13] In preterm

Table 5. Liver and spleen dimensions in preterm and term infants

	Liver dimensions				Spleen dimensions			
Patient group	Mean	SD	5th p	95th p	Mean	SD	5th p	95th p
450-1500 g (n=17)	38.3	5.6	32.8	46.8	26.6	4.4	19.0	32.3
1501-2000 g (n=22)	48.2	4.0	43.0	54.8	33.5	4.0	27.0	38.9
2001-2501 g (n=16)	51.6	5.5	45.8	63.0	37.4	4.9	31.5	44.3
>2501 g (n=11)	53.5	4.1	48.0	58.0	40.8	4.8	35.0	47.0
24-31 weeks (n=22)	40.2	6.2	33.0	51.0	27.9	4.9	19.0	36.0
32-35 weeks (n=26)	48.8	3.8	43.0	54.8	34.9	4.3	28.0	42.5
35-37 weeks (n=18)	53.8	5.1	45.8	63.0	39.8	5.2	33.9	48.0
All preterm infants (n=66)	47.3	7.4	34.3	58.0	34.0	6.6	23.4	45.8
Term infants (n=46)	59.2	7.8	46.5	71.0	44.6	6.4	35.3	57.3

SD: Standard deviation, g: Grams, p: Percentiles. All measurements are given in milimeters.

Table 6. Comparison of liver dimensions with literature results in term infants

Age group	Liver size	Border percentiles$^{Lowest/highest}$	p^	Ref
Our results (term)	59.2±7.8	46.5/71.0$^{(5th/95th)}$		
1-3 months	64±10.4	48.0/82.0$^{(5th/95th)}$	0.012	Konuş et al.[5]
0-2 months	53.2±5.2	44.0/64.3$^{(3rd/97th)}$	<0.001	Amatya et al.[7]
0-3 months boys	65±12.3	48.0/89.0$^{(3rd/97th)}$	0.062	Dhingra et al.[9]
0-3 months girls	62±6.6	49.0/72.0$^{(3rd/97th)}$	0.277	Dhingra et al.[9]
Term infants	57.2±8.8	32.8/80.2$^{(min/max)}$	0.137	Ayede et al.[14]
Term infants	54.5±8.7	24.0/78.0$^{(min/max)}$	0.001	Soyupak et al.[13]
Term infants	45.8±5.6	32.0/62.0$^{(min/max)}$	<0.001	Chen et al.[15]
Term infants	60.9±4.9	53.0/69.7$^{(5th/95th)}$	0.04	Kahramaner et al.[6]

GW: Gestational week at birth, ^: Estimated independent mean difference

Table 7. Comparison of liver dimensions with literature results in preterm infants

Age group	Liver size	Border percentiles$^{Lowest/highest}$	p^	Ref
Our results (preterm)	47.3±7.4	34.3/58$^{(5th/95th)}$		
34-42 GW	42.4±6.3	33.0/58.0$^{(min/max)}$	<0.001	Chen et al.[15]
Our results 34-42 GW	55.9+/-7.9	44.9/70.1$^{(5th/95th)}$		
24-31 GW	37±6.8	28.0/57.6$^{(min/max)}$	0.139	Soyupak et al.[13]
Our results 24-31	39.8±6.5	32.1/51.0$^{(5th/95th)}$		
32-35 GW	46±7.3	32.0/62.0$^{(min/max)}$	0.105	Soyupak et al.[13]
Our results 32-35	48.6±4.1	41.6/54.7$^{(5th/95th)}$		
36-37 GW	53.6±6.4	35.0/63.0$^{(min/max)}$	0.575	Soyupak et al.[13]
Our results 36-37	52.6±6.2	44.6/63.0$^{(5th/95th)}$		
<1500g (girls&boys)	-	31.7&26.1/58.1&53.1$^{(5th/95th)}$		Kahramaner et al.[12]
Our results <1500g	-	31.8/46.8$^{(5th/95th)}$		
1501-2000g (girls&boys)	-	40.8&42.0/62.2&56.9$^{(5th/95th)}$		Kahramaner et al.[12]
Our results 1501-2000g	-	41.2/54.7$^{(5th/95th)}$		
2001-2500g (girls&boys)	-	45.9&46.0/62.4&71.1$^{(5th/95th)}$		Kahramaner et al.[12]
Our results 2001-2500g	-	45.8/63.0$^{(5th/95th)}$		
>2500g (girls&boys)	-	52.0&47.2/66.1&65.0$^{(5th/95th)}$		Kahramaner et al.[12]
Our results >2500g	-	48.0/58.0$^{(5th/95th)}$		

GW: Gestational week at birth, ^: Estimated independent mean difference

infants, each 100-gram increase in birth weight was associated with a 0.8 mm increase in spleen dimension (95% CI: 0.6-1.1), and in term infants, the corresponding increase was 0.8 mm (95% CI: 0.3-1.2). These findings reinforce the importance of birth weight as a primary determinant of spleen size in the early neonatal period.

Comparisons with published literature revealed notable variability. Two studies involving infants up to three months of age reported larger spleen sizes than our cohort,[5,8] whereas three other studies found no significant differences.[3,9,11] Conversely, two studies reported significantly smaller spleen dimensions, by 10.7 mm[13] and 6.4 mm,[6] respectively, compared with our results. Such discrepancies likely reflect differences in ethnicity, geographic location, study populations, measurement techniques, and age at assessment. **Table 8 and 9** summarizes these comparisons and highlights both the similarities and differences between populations.

Clinically, this variation emphasizes the importance of using locally derived percentile curves for spleen size assessment. Universal cut-off values for splenomegaly may not be appropriate for all populations, as reliance solely on international reference data could lead to over- or under-diagnosis. The locally generated percentile tables in our study provide population-specific reference ranges, which can support accurate early detection of splenomegaly, guide timely investigations, and improve neonatal care outcomes.

Taken together with liver dimension findings, these results highlight that anthropometric parameters-particularly birth weight and length-are essential for interpreting visceral organ sizes in neonates. The percentile tables provided in this study offer clinicians practical tools for assessing liver and spleen enlargement in both preterm and term infants, reinforcing the need for population-specific references rather than generalized international cut-offs.[3,5,6,8-15]

Table 8. Comparison of spleen dimensions with literature results in term infants

Age group	Spleen size	Lowest/highest persentile	p	Ref
Our results (term)	44.6±6.4	35.3/57.3$^{(5th-95th)}$		
1-3 months	53±7.8	40.0/65.0$^{(5th-95th)}$	<0.001^	Konus et al.[5]
0-2 months	53.2±5.2	44.0/64.3$^{(3rd-97th)}$	<0.001^	Amatya et al.[8]
0-3 months boys	49.0±14.4	37.0/87.0$^{(3rd-97th)}$	0.361^	Dhingra et al.[9]
0-3 months girls	44.5±5.3	32.0/52.0$^{(3rd-97th)}$	0.275^	Dhingra et al.[9]
0-3 months girls	44.0±5.7	32.0/55.0$^{(min/max)}$	0.282^	Megremis et al.[3]
0-3 months boys	46.0±8.4	28.0/68.0$^{(min/max)}$	0.956^	Megremis et al.[3]
0-3 month	45$^{(median)}$	33.0/58.0$^{(10th-90th)}$	0.384^{W}	Rosenberg et al.[11]
Term infants	33.9±5.4	18.0/49.0$^{(min/max)}$	<0.001^	Soyupak et al.[5]
Term infants	38.2±4.3	32.0/45.7$^{(5th/95th)}$	<0.001^	Kahramaner et al.[6]

GW: Gestational week at birth, ^: Estimated independent mean difference, W: One sample Wilcoxon rank test

Table 9. Comparison of spleen dimensions with literature results in preterm infants

Age group	Spleen size	Lowest/highest persentile	p	Ref
Our results (preterm)	34.0±6.6	23.4/45.8$^{(5th/95th)}$		
24-31 GW	23.9±3.9	16.0/32.0$^{(min/max)}$	0.006^	Soyupak et al.[13]
Our results 24-31 GW	27.9+/-4.9	19.0/36.0$^{(5th/95th)}$		
32-35 GW	28.2±5.3	17.0/40.0$^{(min/max)}$	<0.001^	Soyupak et al.[13]
Our results 32-35 GW	34.7+/-4.4	28.0/42.4$^{(5th/95th)}$		
36-37 GW	33.3±3.7	26.0/42.0$^{(min/max)}$	<0.001^	Soyupak et al.[13]
Our results 36-37 GW	39.8+/-5.0	33.9/48.0$^{(5th/95th)}$		
<1500g (girls&boys)	-	14.8&17.8/39.3&34.6$^{(5th/95th)}$		Kahramaner et al.[12]
Our results <1500g	-	19.0/36.2		
1501-2000g (girls&boys)	-	22.8&22.8/41.6&42.4$^{(5th/95th)}$		Kahramaner et al.[12]
Our results 1501-2000g	-	27.0/38.8		
2001-2500g (girls&boys)	-	28.8&26.4/45.6&44.4$^{(5th/95th)}$		Kahramaner et al.[12]
Our results 2001-2500g	-	31.5/44.3		
>2500g (girls&boys)	-	33.3&24.7/46.3&53.0$^{(5th/95th)}$		Kahramaner et al.[12]
Our results >2500g	-	35.0/47.0		

GW: Gestational week at birth, ^: Estimated independent mean difference, W: One sample Wilcoxon rank test

Clinical Implications

Our percentile tables for liver and spleen dimensions can assist clinicians in the early detection of hepatomegaly and splenomegaly during the neonatal period. By providing locally derived reference ranges, these tables allow for more accurate assessment of organ sizes in both preterm and term infants, reducing the risk of misdiagnosis. Early identification of abnormal organ enlargement can guide timely investigations and interventions, thereby improving neonatal care outcomes.

Limitations

This was a single-center, retrospective study, which may limit the generalizability of the findings. The relatively small sample size further restricts the strength of the conclusions. In addition, measurement variability due to operator dependency, the lack of external validation, and the absence of certain potential confounders (e.g., maternal diseases, perinatal factors) may have influenced the results. Despite these limitations, our findings provide useful local references and a basis for further multicenter studies with larger populations.

CONCLUSION

As a result, birth week, birth weight, and birth length are the primary determinants of liver and spleen sizes in both preterm and term neonates. While gender may influence spleen size in term infants, its impact on liver dimensions remains unclear. The locally derived percentile tables provided in this study offer practical tools for clinicians to accurately assess liver and spleen enlargement, facilitating early detection of hepatomegaly and splenomegaly. These population-specific references emphasize the need for tailored assessments in neonatal care, rather than relying solely on international standards, ultimately supporting timely interventions and better neonatal outcomes.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was carried out with the permission of the Marmara University Faculty of Medicine Clinical Researches Ethics Committee (Date: 07.01.2022, Decision No: 09.2022.23).

Informed Consent

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

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

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

Author Contributions

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

Availability of Data and Material

The datasets will be shared upon reasonable request

Acknowledgements

We thank Dr. Halil Toprak for performing the ultrasonographic measurements and Dr. Sinem Gülcan Kersin for her contribution to the study idea.

REFERENCES

1. Joshi R, Singh A, Jajoo N, Pai M, Kalantri SP. Accuracy and reliability of palpation and percussion for detecting hepatomegaly: a rural hospital-based study. *Indian J Gastroenterol*. 2004;23(5):171-174.
2. Pelizzo G, Guazzotti M, Klersy C, et al. Spleen size evaluation in children: time to define splenomegaly for pediatric surgeons and pediatricians. *PLoS One*. 2018;13(8):e0202741. doi:10.1371/journal.pone.0202741
3. Megremis SD, Vlachonikolis IG, Tsilimigaki AM. Spleen length in childhood with US: normal values based on age, sex, and somatometric parameters. *Radiology*. 2004;231(1):129-134. doi:10.1148/radiol.2311020963
4. Rousan LA, Fataftah J, Al-Omari M, Hayajneh W, Miqdady M, Khader Y. Sonographic assessment of liver and spleen size based on age, height, and weight: evaluation of Jordanian children. *Minerva Pediatr*. 2019;71(1):28-33. doi:10.23736/S0026-4946.16.04433-9
5. Warnakulasuriya DTD, Peries PPUC, Rathnasekara YAC, Jayawardena KTM, Upasena A, Wickremasinghe AR. Ultrasonographic parameters of the liver, spleen and kidneys among a cohort of school children in Sri Lanka. *BMC Pediatr*. 2017;17(1):192.
6. Kahramaner Z, Erdemir A, Arik B, Bilgili G, Tekin M, Genc Y. Reference ranges of liver and spleen dimensions in term infants: sonographic measurements. *J Med Ultrason (2001)*. 2015;42(1):77-81. doi:10.1007/s10396-014-0578-0
7. Fenton TR, Kim JH. A systematic review and meta-analysis to revise the Fenton growth chart for preterm infants. *BMC Pediatr*. 2013;13:59. doi:10.1186/1471-2431-13-59
8. Amatya P, Shah D, Gupta N, Bhatta NK. Clinical and ultrasonographic measurement of liver size in normal children. *Indian J Pediatr*. 2014;81(5):441-445. doi:10.1007/s12098-013-1288-0
9. Dhingra B, Sharma S, Mishra D, Kumari R, Pandey RM, Aggarwal S. Normal values of liver and spleen size by ultrasonography in Indian children. *Indian Pediatr*. 2010;47(6):487-492. doi:10.1007/s13312-010-0090-6
10. Thapa NB, Shah S, Pradhan A, Rijal K, Pradhan A, Basnet S. Sonographic assessment of the normal dimensions of liver, spleen, and kidney in healthy children at tertiary care hospital. *Kathmandu Univ Med J (KUMJ)*. 2015;13(52):286-291. doi:10.3126/kumj.v13i4.16825
11. Rosenberg HK, Markowitz RI, Kolberg H, Park C, Hubbard A, Bellah RD. Normal splenic size in infants and children: sonographic measurements.*AJR Am J Roentgenol*. 1991;157(1):119-121. doi:10.2214/ajr.157.1.2048509
12. Rosenberg HK, Markowitz RI, Kolberg H, Park C, Hubbard A, Bellah RD. Normal splenic size in infants and children: sonographic measurements.*AJR Am J Roentgenol*. 1991;157(1):119-121. doi:10.2214/ajr.157.1.2048509
13. Soyupak SK, Narli N, Yapicioğlu H, Satar M, Aksungur EH. Sonographic measurements of the liver, spleen and kidney dimensions in the healthy term and preterm newborns. *Eur J Radiol*. 2002;43(1):73-78. doi:10.1016/s0720-048x(01)00466-1
14. Ayede AI, Agunloye AM, Gram L, Omokhodion SI. Normal ultrasonographic dimensions of the liver in neonates in South-West Nigeria. *West Afr J Med*. 2014;33(3):183-188.
15. Chen CM, Wang JJ. Clinical and sonographic assessment of liver size in normal Chinese neonates. *Acta Paediatr*. 1993;82(4):345-347. doi:10.1111/j.1651-2227.1993.tb12693.x

Original Article

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1758040
J Health Sci Med. 2025;8(5):946-951

Assessment of fetal cardiac function and epicardial fat thickness in intrahepatic cholestasis pregnancies

Neval Çayönü Kahraman[1,2], Zeynep Şeyhanlı[2], Gülşan Karabay[2], Özgür Arat[1], Kadriye Yakut Yücel[2], Şevki Çelen[2], Ali Turhan Çağlar[2], Yaprak Engin Üstün[1]

[1]Department of Perinatology, Etlik Zübeyde Hanım Gynecological Diseases Training and Research Hospital, University of Health Sciences, Ankara, Turkiye
[2]Department of Perinatology, Ankara Etlik City Hospital, Ankara, Turkiye

Cite this article as: Çayönü Kahraman N, Şeyhanlı Z, Karabay G, et al. Assessment of fetal cardiac function and epicardial fat thickness in intrahepatic cholestasis pregnancies. *J Health Sci Med.* 2025;8(5):946-951.

Received: 04.08.2025 ◆ **Accepted**: 05.09.2025 ◆ **Published**: 16.09.2025

ABSTRACT

Aims: This study is to evaluate fetal cardiac activity and epicardial fat thickness (EFT) in pregnant women with intrahepatic cholestasis (IHCP) and to investigate the relationship between these measurements and perinatal outcomes.

Methods: This prospective case-control study was conducted between May 2022 and October 2024 at a tertiary perinatology clinic. The study group comprised 38 women with IHCP, and the control group included 39 healthy pregnancies matched for gestational age and maternal characteristics. Fetal cardiac function was evaluated by echocardiographic parameters including the Myocardial Performance Index (MPI), isovolumic contraction time (ICT), ejection time (ET), and PR interval, and EFT was quantified in the four-chamber view. Doppler indices of the umbilical, middle cerebral, and uterine arteries, along with neonatal outcomes such as gestational age, birth weight, APGAR scores, and neonatal intensive care unit (NICU) admission, were recorded.

Results: In the IHCP cohort, bile acid, aspartate aminotransferase (AST), and alanine aminotransferase (ALT) levels were significantly elevated ($p<0.001$). MPI, ICT, ET, and PR interval values were also increased, indicating subclinical effects on fetal cardiac function. No significant differences were detected in EFT between groups. Gestational age at birth was lower in the IHCP group ($p=0.005$), and NICU admission was clinically higher, though not statistically significant.

Conclusion: IHCP may impair fetal cardiac conduction and contractility, with these alterations positively associated with serum bile acid concentrations. EFT seems more closely linked to chronic metabolic stress than to acute conditions. Regular echocardiographic monitoring may aid in safeguarding fetal well-being.

Keywords: Cholestasis, intrahepatic, fetal heart, Myocardial Performance Index, epicardial fat, bile acids

INTRODUCTION

Intrahepatic cholestasis is a prevalent liver disorder observed in pregnancy, with incidence rates between 1.7% and 2.9%, depending on the population studied.[1] This disease usually manifests in the third trimester, marked by pruritus and increased serum bile acid concentrations.[2-4]

The maternal prognosis for intrahepatic cholestasis is typically favorable; nevertheless, it is associated with severe consequences, including premature birth, meconium aspiration, intrauterine growth restriction, and even intrauterine fetal demise regarding fetal outcomes.[5,6]

The transplacental transfer of bile acids may induce harmful effects on the fetal myocardium and other organ systems.[7] This impact may specifically induce rhythm and functional abnormalities in the fetal heart.[7,8] Fetal cardiac function is evaluated and subclinical abnormalities are identified using echocardiographic parameters such as the MPI, ICT, ET, IRT, and PR interval.[9-12] Examination of these indicators, particularly in systemic pregnancy disorders like cholestasis, may facilitate the early identification of fetal risks. In recent years, fetal epicardial adipose tissue has been investigated as a potential marker of fetal metabolic status and cardiac workload.[13,14] The term EFT describes visceral adipose tissue that accumulates on the outer surface of the fetal heart and is thought to be linked with metabolic stress, inflammation, or cardiac dysfunction.[15] Although EFT has traditionally been investigated in chronic maternal conditions such as obesity and gestational diabetes, it may also be influenced by acute intrauterine stress and inflammatory changes.[16] Based on this background, we hypothesized that pregnancies complicated by intrahepatic cholestasis would differ from healthy controls in terms of fetal EFT, and that EFT would be associated with fasting bile acid levels and the MPI. Evaluating MPI and EFT together may therefore provide complementary insights into the cardiovascular effects of intrahepatic cholestasis.

Corresponding Author: Neval Çayönü Kahraman, nevalcayonu@gmail.com

Evaluating fetal cardiac function and epicardial fat thickness in pregnancies complicated by intrahepatic cholestasis is essential for understanding the disease's impact on the fetus. Moreover, elucidating the associations between these parameters, perinatal outcomes, and physiological indicators (particularly fasting bile acid levels) may contribute to the refinement of fetal monitoring strategies.

METHODS

Ethical approval was obtained by the institutional review board from the Etlik Zübeyde Hanım Gynecological Diseases Training and Research Hospital Clinical Researches Ethics Committee (Date: 20.04.2022, Decision No: 2022/56). The study complied with the ethical principles for medical research of the Declaration of Helsinki.

This study is a prospective observational case-control research conducted from May 2022 to October 2024 at a perinatology clinic within a tertiary care hospital's Department of Perinatology clinic. The study group consisted of 38 pregnant women who had been diagnosed with intrahepatic cholestasis, whereas the control group consisted of 39 healthy pregnant women who were matched to the study group in terms of baseline maternal characteristics and gestational age, and who had no systemic diseases or pregnancy complications identified during follow-up. The study group established the inclusion criteria for single pregnant women exhibiting pruritus, with blood total bile acid levels ≥10 μmol/L, and without any other hepatic or systemic disorders. The control group comprised healthy, single pregnant women who were observed at the same gestational week, exhibited no systemic diseases, and experienced no prenatal problems. The following exclusion criteria were implemented; chronic liver illness, hepatitis, preeclampsia, gestational diabetes, fetal structural anomalies, multiple gestation, intrauterine growth restriction (IUGR), polyhydramnios/oligohydramnios, congenital heart disease, and maternal history of hypertension. All instances were documented for maternal age, gravida, parity, body-mass index (BMI), gestational age at diagnosis, aspartate aminotransferase (AST), alanine aminotransferase (ALT), and total serum bile acid concentrations. The established threshold for diagnosing cholestasis is a total bile acid level of ≥10 μmol/L.[8,17,18] Laboratory studies were conducted in the hospital's biochemistry laboratory employing standardized methodologies. A perinatologist performed fetal echocardiogram and Doppler assessments using ultrasound and fetal echocardiography, in accordance with the gestational age, on a Voluson E8 (GE Healthcare, Milwaukee, WI, USA) apparatus with a 2-5 MHz convex transducer.

The fetal cardiac evaluation was conducted using the apical four-chamber view. The ICT, IRT, and ET were quantified. The PR interval was quantified via M-mode imaging, representing the duration from the initiation of atrial activity to the start of the ventricular complex (**Figure 1**). MPI was computed using the formula (ICT + IRT)/ET. The time from when the atrioventricular (AV) valve closes and the semilunar valve opens is known as the ICT, the time from when the semilunar valve closes and the AV valve opens is known as the IRT, and the duration of blood flow when the semilunar valve is open is known as the ET.

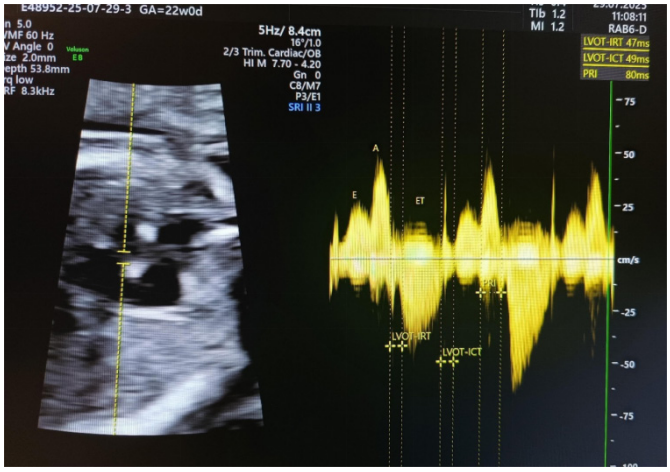

Figure 1. Pulsed-wave Doppler recording obtained from the LVOT in a 22-week fetus. IRT, ICT, ET, and atrioventricular contraction time (PR) are measured to calculate the MPI, also known as the Tei Index. The labels E and A represent early and atrial filling waves, respectively.

LVOT: Left ventricular outflow tract, IRT: Isovolumic relaxation time, ICT: Isovolumic contraction time, ET: Ejection time, PR: The time between the onset of the p wave and the onset of the QRS complex, MPI: Myocardial performance index

The EFT was assessed at its maximum location between the heart's free wall and the visceral pericardium in the four-chamber view (**Figure 2**). Each measurement was conducted a minimum of three times, and the mean values were recorded. The fetal Doppler measurements included the umbilical artery Pulsatility Index (UA-PI), middle cerebral artery Pulsatility Index (MCA-PI), uterine artery average Pulsatility Index (UtA-PI), and the cerebroplacental ratio (CPR), which is calculated as MCA-PI divided by UA-PI. Doppler analyses were conducted during the periods of fetal quiescence, characterized by the absence of respiratory movements, with gonial angles less than 30°. Each Doppler parameter was measured thrice, and the mean values were documented.

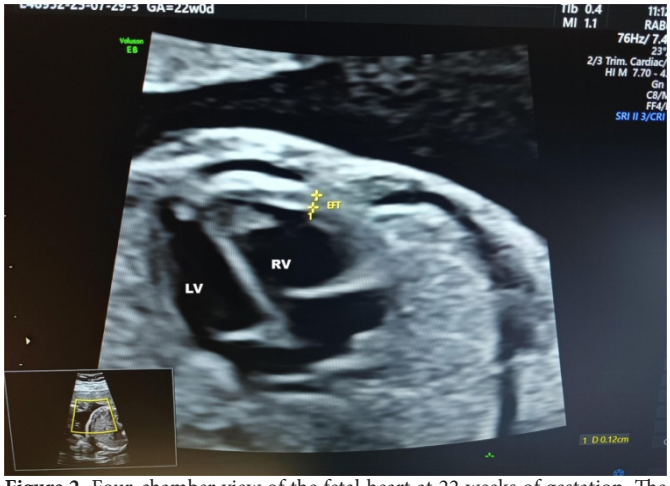

Figure 2. Four-chamber view of the fetal heart at 22 weeks of gestation. The EFT is marked between the right ventricular wall and the pericardium.

EFT: Epicardial fat thickness, LV: Left ventricle, RV: Right ventricle

Gestational age at delivery, birth weight, APGAR scores at 1 and 5 minutes, and the requirement for neonatal intensive care unit (NICU) admission were recorded. NICU admission was defined as the need for respiratory support, management of hypoglycemia, treatment of jaundice, or other medical interventions in the neonate.

Statistical Analysis

Data analyses were performed using IBM SPSS Statistics version 25.0. Normality of the data was assessed with the Shapiro-Wilk test. For non-parametric data, the Mann-Whitney U test was applied, while categorical variables were analyzed using the Chi-square or Fisher's exact test. Correlations were examined with Spearman's rho test. A p-value <0.05 was considered statistically significant. An a priori power analysis was performed using G*power (version 3.1) for a two-sample T test (two-tailed, α=0.05, 1-β=0.80). Assuming a moderate effect size (Cohen's d=0.65) informed by prior literature, the required sample size was 38 participants per group (76 in total). Our actual sample size (38 IHCP vs. 39 controls) therefore met this requirement.[19]

RESULTS

There were no statistically significant differences between the pregnant women in the IHCP and those in the control group regarding maternal age, number of pregnancies (gravida), number of previous births (parity), body-mass index (BMI), and gestational age (p>0.05). Conversely, total bile acid, AST, and ALT levels were significantly elevated in the IHCP group (p<0.001, Table 1). MPI was also higher in the cholestasis group (p=0.024), consistent with fetal cardiac dysfunction. In addition, ICT was significantly prolonged (p<0.001), ET duration was increased (p=0.015, Table 2), and the PR interval was extended (p<0.001), suggesting a potential effect on atrioventricular conduction.

The IRT and MCA-PI values showed borderline significance between the groups (p=0.052 and p=0.079, respectively). UtA-PI was significantly lower in the cholestasis group compared with controls (p=0.017). No significant differences were observed in EFT, UA-PI, or CPR (p>0.05, Table 2). Neonatal outcomes demonstrated a significantly earlier gestational age at birth in the IHCP group (p=0.005, Table 3).

Table 1. Demographic, clinical characteristics and laboratuar parameters

	Intrahepatic cholestasis group (n=38)	Control group (n=39)	p-value
Maternal age (years), median (min-max)	29 (18-39)	30 (21-39)	0.363[a]
Gravida, median (min-max)	2 (1-5)	2 (1-4)	0.623[a]
Parity, median (min-max)	0 (0-3)	1 (0-4)	0.567[a]
BMI (kg/m²), median (min-max)	29 (25.8-39.2)	29.6 (25.4-38.5)	0.501[a]
GW at diagnosis, median (min-max)	33.7 (26-38)	34 (27-38)	0.613[a]
Total bile acid level (µmol/L), median (min-max)	25 (10-139)	4 (2-8.7)	<0.001[a]
AST U/L, median(min-max)	58 (14-731)	30 (15-30)	<0.001[a]
ALT U/L, median (min-max)	81.5 (11-475)	31 (17-124)	<0.001[a]

[a]: Mann-Whitney U test, min: Minimum, max: Maximum, BMI: Body-mass index, GW: Gestational week, AST: Aspartate aminotransferase, ALT: Alanine aminotransferase

Table 2. Fetal ultrasound, Doppler, and cardiac function parameters

	Intrahepatic cholestasis group (n=38)	Control group (n=39)	p-value
EFT (mm), median (min-max)	1.95 (0.8-2.8)	1.90 (0.69-3)	0.208[a]
MPI, median (min-max)	0.45 (0.30-0.72)	0.40 (0.27-0.75)	0.024[a]
ICT (ms), median (min-max)	42 (26-59)	32 (20-49)	<0.001[a]
IRT (ms), median (min-max)	34 (20-56)	31 (20-47)	0.052[a]
ET (ms), median (min-max)	166 (133-221)	156 (120-191)	0.015[a]
PR interval (ms), median(min-max)	123 (67-134)	110 (82-134)	<0.001[a]
UA-PI, median (min-max)	0.95 (0.56-1.76)	0.99 (0.56-1.60)	0.582[a]
MCA- PI, median (min-max)	1.6 (1.06-3.65)	1.8 (1.2-4.2)	0.079[a]
UtA-PI, median (min-max)	0.83 (0.34-1.66)	0.99 (0.63-1.66)	0.017[a]
CPR, median (min-max)	1.9 (0.92-3.99)	1.8 (0.99-5.7)	0.460[a]

[a]: Mann-Whitney U test, EFT: Epicardial fat thickness, min: Minimum, max: Maximum, MPI: Modified Myocardial Performance Index, ICT: Isovolumetric contraction time, IRT: Isovolumetric relaxation time, ET: Ejection time, PR: The time between the onset of the p wave and the onset of the QRS complex, UA: Umbilical artery PI: Pulsalitiy Index, MCA: Middle cerebral artery, UtA: Uterine artery mean, CPR: Cerebroplacental ratio, mm: Millimetre, ms: Milliseconds

Table 3. Comparison of neonatal outcomes between the groups

	Intrahepatic cholestasis group (n=38)	Control group (n=39)	p-value
GW at delivery, median (min-max)	37 (28-39)	38 (30-41)	0.005[a]
Birth weight (grams), median (min-max)	3100 (1200-3900)	3090 (1509-4000)	0.534[a]
APGAR 1-minute score, median (min-max)	9 (6-9)	9 (5-9)	0.971[a]
APGAR 5-minute score, median (min-max)	10 (6-10)	10 (6-10)	0.202[a]
NICU admission (n, %)	7 (18.4)	4 (10.3)	0.347[b]

[a]: Mann Whitney U test, [b]: Chi-square, Fisher exact test, GW: Gestational week, min: Minimum, max: Maximum, NICU: Neonatal intensive care unit

No significant differences were observed between the groups in birth weight, APGAR scores, or NICU admission (p>0.05), although the rate of NICU admission was higher in the cholestasis group (18.4% vs. 10.3%). While not statistically significant, this finding may still have clinical relevance. Spearman correlation analysis demonstrated a moderate positive association between fasting bile acid levels and both ICT (r=0.376, p=0.001) and ET (r=0.317, p=0.005). A weaker but significant positive association was also noted with the PR interval (r=0.273, p=0.016). In contrast, bile acid levels were moderately and inversely correlated with gestational age (r=-0.372, p=0.001; **Table 4**), suggesting that higher bile acid concentrations may contribute to preterm birth.

Table 4. Correlation between total bile acid level and clinical/fetal parameters

	Spearman rho	p-value
ICT (ms)	0.376	0.001
ET (ms)	0.317	0.005
PR interval (ms)	0.273	0.016
GW at delivery	-0.372	0.001

ICT: Isovolumetric contraction time, ET: Ejection time, PR: The time between the onset of the p wave and the onset of the QRS complex, ms: Milliseconds

DISCUSSION

This prospective case–control study evaluated fetal cardiac function and EFT in pregnancies complicated by intrahepatic cholestasis. To our knowledge, this is one of the few studies that has simultaneously evaluated both MPI and fetal EFT in pregnancies complicated by IHCP, thereby providing novel insights into fetal cardiac adaptation in this condition. Our findings suggest that intrahepatic cholestasis may lead to subclinical functional alterations in the fetal cardiac system.

In our study, MPI was significantly higher in the intrahepatic cholestasis group. As a global index of systolic and diastolic performance, MPI plays an important role in the early detection of fetal cardiac dysfunction.[20] Previous research indicates that elevated bile acids may impair myocardial function by disrupting ion channels and contraction=relaxation mechanisms through transplacental transfer.[21-23]

In our study, ICT and ET were significantly prolonged in the IHCP group, suggesting delayed systolic activation and prolonged ventricular emptying. The concurrent extension of the PR interval further indicates possible impairment of atrioventricular conduction. Experimental studies have shown that bile acids may contribute to conduction delays by disrupting the electrical activity of fetal cardiomyocytes.[24-26]

Although no clinically significant arrhythmia was detected, PR interval prolongation may reflect subclinical impairment of the conduction system. EFT did not differ significantly between groups, indicating that acute and transient maternal disorders such as cholestasis may have limited impact on this parameter. In contrast, previous studies have demonstrated increased EFT in chronic metabolic stress conditions, including gestational diabetes and maternal obesity.[26,27] Thus, EFT may remain largely unaffected in short-term inflammatory states such as cholestasis. In contrast, in acute obstetric conditions characterized by inflammatory stress, such as preterm prelabor rupture of membranes (PPROM), an increase in EFT has been observed.[16] Although IHCP is also considered an acute condition, its underlying pathophysiology-primarily hepatobiliary dysfunction with limited systemic inflammation-may explain why EFT did not significantly change in our cohort. This highlights that the impact of acute maternal disorders on fetal EFT may vary depending on their distinct mechanisms.

Interestingly, the UtA-PI was significantly reduced in the IHCP group in our cohort. Although previous studies have not consistently reported changes in uterine artery Doppler parameters in IHCP, this finding may reflect a compensatory vasodilation response in uteroplacental perfusion aimed at counteracting the detrimental effects of elevated bile acids. Alternatively, given the limited sample size, this result should be interpreted cautiously and warrants confirmation in larger cohorts. No significant differences were detected between the groups in MCA-PI and CPR parameters.

Gestational age at delivery was significantly lower in the cholestasis group, consistent with previous reports. Multiple studies have shown that elevated bile acid levels are associated with an increased risk of preterm birth.[29-31] However, as NICU admission is a relatively infrequent outcome, our study was not specifically powered to detect differences in this parameter. Although our actual enrollment matched the a priori power calculation for EFT, larger multicenter cohorts would be required to evaluate differences in rare neonatal outcomes such as NICU admission.

The positive associations between fasting bile acids and ICT, ET, and PR intervals suggest that elevated bile acids may contribute to impaired fetal cardiac function. Conversely, their inverse relationship with gestational age highlights bile acids as an important determinant of neonatal outcomes.

Limitations

This study has several limitations. The relatively small sample size underscores the need for studies with larger populations to confirm these findings. In addition, although all fetal echocardiographic assessments were performed by a single experienced perinatologist, observer variability cannot be completely excluded due to the inherent nature of the measurements. Intra- and inter-observer reproducibility was not formally assessed in our study, which may further limit the robustness of the echocardiographic findings. Because fetal heart rate and function are dynamic, measurements obtained at a single time point may not fully capture temporal variability. Moreover, the absence of standardized reference curves for EFT and the lack of data on additional maternal metabolic factors (e.g., insulin resistance, lipid profile) further limit the interpretation of this parameter.

Long-term neonatal outcomes were not assessed in this study. Future longitudinal research is needed to clarify the postnatal implications of prenatal cardiac alterations.

CONCLUSION

This study demonstrated that intrahepatic cholestasis may impair fetal cardiac function, with significant alterations in parameters such as the MPI, ICT, ET, and PR interval. These findings support the hypothesis that the condition exerts subclinical adverse effects on cardiac conduction and contractility. Furthermore, the positive associations between bile acid concentrations and cardiac indices suggest a direct influence of metabolic disturbance on fetal heart function. The absence of significant differences in fetal EFT suggests that this parameter is more closely associated with chronic metabolic disorders than with acute maternal conditions. The elevated risk of preterm birth and increased need for neonatal intensive care in cholestasis highlight the importance of careful fetal surveillance and timely intervention in affected pregnancies. In conclusion, echocardiographic assessment of fetal cardiac function in intrahepatic cholestasis may enable the early detection of subclinical dysfunction and serve as a valuable tool for optimizing perinatal care strategies.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was carried out with the permission of the Etlik Zübeyde Hanım Gynecological Diseases Training and Research Hospital Clinical Researches Ethics Committee (Date: 20.04.2022, Decision No: 2022/56).

Informed Consent

All patients signed and free and informed consent form.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

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

Author Contributions

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

Data Availability Statement

Data supporting the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

1. Jamshidi Kerachi A, Shahlaee MA, Habibi P, Dehdari Ebrahimi N, Ala M, Sadeghi A. Global and regional incidence of intrahepatic cholestasis of pregnancy: a systematic review and meta-analysis. *BMC Med.* 2025; 23(1):129. doi:10.1186/s12916-025-03935-0
2. Odabaş RK, Sökmen Y, Dünder E, Taşpınar A. The incidence of intrahepatic cholestasis of pregnancy and its maternal, fetal, and neonatal adverse outcomes: a systematic review and meta-analysis. *J Midwife Womens Health.* 2024;69(3):370-382. doi:10.1111/jmwh.13640
3. Hobson SR, Cohen ER, Gandhi S, et al. Guideline no. 452: diagnosis and management of intrahepatic cholestasis of pregnancy. *J Obstet Gynaecol Can.* 2024;46(8):102618. doi:10.1016/j.jogc.2024.102618
4. Demir M, Sertel E. The role of aspartate aminotransferase/Platelet Ratio Index score in predicting intrahepatic cholestasis of pregnancy and its relationship with total bile acid level: a case-control study from a tertiary center. *Gynecol Obstet Reprod Med.* 2024;30(3):167-173. doi:10.21613/GORM.2023.1527
5. Palmer KR, Xiaohua L, Mol BW. Management of intrahepatic cholestasis in pregnancy. *Lancet.* 2019;393(10174):853-854. doi:10.1016/S0140-6736(18)32323-7
6. Arthuis C, Diguisto C, Lorphelin H, et al. Perinatal outcomes of intrahepatic cholestasis during pregnancy: an 8-year case-control study. *PLoS One.* 2020;15(2):e0228213. doi:10.1371/journal.pone.0228213
7. Wang J, Lun W, Shi W. Effects of elevated bile acid levels on fetal myocardium in intrahepatic cholestasis of pregnancy, a retrospective study from a neonatal perspective. *Clin Res Hepatol Gastroenterol.* 2022; 46(8):102013. doi:10.1016/j.clinre.2022.102013
8. Lee RH, Mara Greenberg, Metz TD, Pettker CM. Society for Maternal-Fetal Medicine consult series #53: intrahepatic cholestasis of pregnancy. *Am J Obstet Gynecol.* 2021;224(2):B2-B9. doi: 10.1016/j.ajog.2020.11.002
9. Williamson C, Miragoli M, Sheikh Abdul Kadir S, et al. Bile acid signaling in fetal tissues: implications for intrahepatic cholestasis of pregnancy. *Dig Dis.* 2011;29(1):58-61. doi:10.1159/000324130
10. Oliveira M, Dias JP, Guedes-Martins L. Fetal cardiac function: Myocardial Performance Index. *Curr Cardiol Rev.* 2022;18(4):20-31. doi: 10.2174/1573403X18666211227145856
11. Bligard KH, Doering M, Porat S, Rosenbloom JI. Reference ranges for left Modified Myocardial Performance Index: systematic review and meta-analysis. *Fetal Diagn Ther.* 2024;51(2):159-167. doi:10.1159/000535602
12. Ozturk M, Agaoglu Z, Ozturk FH, et al. Evaluation of fetal myocardial performance index in gestational diabetes mellitus. *Congenital Anomalies.* 2023;63(5):164-169. doi:10.1111/cga.12531
13. Sever B, Bayraktar B, Adıyaman D, et al. Association of increased fetal epicardial fat thickness with maternal pregestational and gestational diabetes. *J Mater Fetal Neonat Med.* 2023;36(1):2183474. doi:10.1080/14767058.2023.2183474
14. Manjunath BST. Fetal epicardial fat thickness as a marker for gestational diabetes mellitus. *IJSR.* 2024;34-36. doi:10.36106/ijsr
15. Nyawo TA, Pheiffer C, Mazibuko-Mbeje SE, et al. Physical exercise potentially targets epicardial adipose tissue to reduce cardiovascular disease risk in patients with metabolic diseases: oxidative stress and inflammation emerge as major therapeutic targets. *Antioxidants.* 2021; 10(11):1758. doi:10.3390/antiox10111758
16. Sakcak B, Farisoğulları N, Denizli R, et al. Evaluation of the Fetal Myocardial Performance Index and epicardial fat thickness in pregnant women with preterm prelabor rupture of membranes. *J Matern Fetal Neonatal Med.* 2023;36(1):2192322. doi:10.1080/14767058.2023.2192322
17. Majsterek M, Wierzchowska-Opoka M, Makosz I, Kreczyńska L, Kimber-Trojnar Ż, Leszczyńska-Gorzelak B. Bile acids in intrahepatic cholestasis of pregnancy. *Diagnostics (Basel).* 2022;12(11):2746. doi:10.3390/diagnostics12112746
18. Ozgen L, Ozgen G, Karasin SS, Bayram F. Fibrinogen levels and total serum bile acids in intrahepatic cholestasis of pregnancy. *J Coll Physicians Surg Pak.* 2022;32(11):1404-1409. doi:10.29271/jcpsp.2022.11.1404
19. Yakut K, Öcal DF, Sanhal Yaşar C, Halıcı Öztürk F, Şanlı C, Çelen Ş. Fetal epicardial fat thickness in fetal growth restriction; effects on fetal heart function and relationship with the severity of disease. *J Matern Fetal Neonatal Med.* 2022;35(25):6946-6952. doi:10.1080/14767058.2021.1931676
20. Dayanan R, Bulan DD, Ozkan MA, Ozkan HD, Caglar AT. Fetal pulmonary-artery Doppler dynamics in preeclampsia: acceleration time, ejection time, PAT/ET ratio, and prediction of adverse neonatal pulmonary outcome. *J Ultrasound Med.* 2025;44(10):1903-1912. doi:10.1002/jum.70031
21. Sheikh Abdul Kadir SH, Miragoli M, Abu-Hayyeh S, et al. Bile acid-induced arrhythmia is mediated by muscarinic M2 receptors in neonatal rat cardiomyocytes. *PLoS One.* 2010;5(3):e9689. doi:10.1371/journal.pone.0009689
22. Schultz F, Hasan A, Alvarez-Laviada A, et al. The protective effect of ursodeoxycholic acid in an in vitro model of the human fetal heart occurs via targeting cardiac fibroblasts. *Prog Biophys Mol Biol.* 2016; 120(1-3):149-163. doi:10.1016/j.pbiomolbio.2016.01.003
23. Ibrahim E, Diakonov I, Arunthavarajah D, et al. Bile acids and their respective conjugates elicit different responses in neonatal cardiomyocytes: role of Gi protein, muscarinic receptors and TGR5. *Sci Rep.* 2018;8(1):7110. doi:10.1038/s41598-018-25569-4

24. Guerra M, Montaño I, Haye MT, Toro V, Maiz N. Fetal PR interval in pregnancies with intrahepatic cholestasis of pregnancy: a case-control study. *Am J Perinatol.* 2024;41(S 01):e2964-e2969. doi:10.1055/ a-2184-0242
25. Rodríguez M, Moreno J, Márquez R, et al. Increased PR interval in fetuses of patients with intrahepatic cholestasis of pregnancy. *Fetal Diagn Ther.* 2016;40(4):298-302. doi:10.1159/000444297
26. Eyisoy ÖG, Demirci O, Taşdemir Ü, Özdemir M, Öcal A, Kahramanoğlu Ö. Effect of maternal ursodeoxycholic acid treatment on fetal atrioventricular conduction in patients with intrahepatic cholestasis of pregnancy. *Fetal Diagn Ther.* 2024;51(6):617-623. doi:10.1159/000540261
27. Anwaier A, Li J, Liu W, Dong L, Ding Y, Yu Z. Influence of maternal diabetes during pregnancy on ultrasound-measured fetal epicardial fat thickness: a meta-analysis. *Biomol Biomed.* 2025;25(6):1245-1258. doi: 10.17305/bb.2025.11909
28. Sweeting A, Wong J, Murphy HR, Ross GP. A clinical update on gestational diabetes mellitus. *Endocr Rev.* 2022;43(5):763-793. doi:10. 1210/endrev/bnac003
29. Zhou Q, Yuan Y, Wang Y, et al. The severity of intrahepatic cholestasis during pregnancy increases risks of adverse outcomes beyond stillbirth: evidence from 15,826 patients. *BMC Pregn Childbirth.* 2024;24(1):476. doi:10.1186/s12884-024-06645-2
30. Ovadia C, Seed PT, Sklavounos A, et al. Association of adverse perinatal outcomes of intrahepatic cholestasis of pregnancy with biochemical markers: results of aggregate and individual patient data meta-analyses. *Lancet.* 2019;393(10174):899-909. doi:10.1016/S0140-6736(18)31877-4
31. You S. Bile acids and premature labor in intrahepatic cholestasis of pregnancy. University of Rhode Island; 2016.

HEALTH SCIENCES
MEDICINE

DOI: 10.32322/jhsm.1784678
J Health Sci Med. 2025;8(5):952

Erratum to "The impact of pre-transplant fibrinogen and D-dimer levels on transplantation outcomes in autologous stem cell transplantation"

Neslihan Mandacı Şanlı[1], Edanur Dilara Mandacı[2]

[1]Department of Hematology and Bone Marrow Transplant Center, Faculty of Medicine, Erciyes University, Kayseri, Turkiye
[2]MSc. Student, Department of Artificial Intelligence and Data Science, Institute of Science, Sivas Cumhuriyet University, Sivas, Turkiye

In the original article, the ORCID of one of the co-authors, Edanur Dilara Mandacı, was incorrectly reported by the corresponding author. At the request of the authors, the ORCID has been corrected to reflect the accurate information. This correction does not affect the content or conclusions of the article. The authors apologize for this oversight.

DOI of original article: https://doi.org/10.32322/jhsm.1711682

Corresponding Author: Neslihan Mandacı Şanlı, ortoforia@hotmail.com