The Medical Bulletin of Haseki

E-ISSN: 2147-2688

2025
Volume 63
Issue
September 3

www.hasekidergisi.com

The Medical Bulletin of Haseki

Editorial Board

galenos yayınevi

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

Online Publishing Date: June 2025
E-ISSN: 2147-2688
International scientific journal published quarterly.

The Medical Bulletin of Haseki

Scientific Advisory Board

The Medical Bulletin of Haseki

Scientific Advisory Board

The Medical Bulletin of Haseki

Scientific Advisory Board

The Medical Bulletin of Haseki

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

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

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

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

Responsible Manager: Akif ERBİN

The Medical Bulletin of Haseki

Contents

Original Articles

Original Article

DOI: 10.4274/haseki.galenos.2025.79663
Med Bull Haseki 2025;63(3):123-127

Large Language Models and Male Circumcision: A Reliability Assessment

Ismail Ulus*, Gokhan Ceker**, Ibrahim Hacibey**

*University of Health Sciences Türkiye, Bagcilar Training and Research Hospital, Clinic of Urology, Istanbul, Türkiye
**University of Health Sciences Türkiye, Basaksehir Cam and Sakura City Hospital, Clinic of Urology, Istanbul, Türkiye

Abstract

Aim: Male circumcision remains routine in some countries for neonatal or religious reasons; however, it continues to be the subject of ongoing debate concerning its health benefits, potential risks, and implications for bodily autonomy. This study aims to evaluate the reliability of patient-facing content generated by four widely used large language models (LLMs) on various aspects of male circumcision.

Methods: A search regarding LLMs was conducted using 20 standardized questions on 10 May 2025. Responses from ChatGPT, Copilot, Gemini, and Perplexity were evaluated by three independent experts. Inter-rater reliability was assessed with the intraclass correlation coefficient, and model performance differences were analyzed using Kruskal-Wallis tests with Bonferroni correction.

Results: Inter-rater reliability was strong, with an intraclass correlation coefficient of 0.79 ($p<0.001$). Perplexity demonstrated statistically significant lower performance compared to ChatGPT, Copilot, and Gemini when evaluated across the thematic domains ($p<0.001$). Similarly, Perplexity performed statistically significantly worse than the other models across the criteria of clarity, structure, utility, and factual accuracy ($p<0.001$).

Conclusion: Gemini and Copilot were the top performers across both thematic domains and evaluation criteria, highlighting substantial differences among LLMs in their ability to provide accurate and well-structured medical information regarding male circumcision. While ChatGPT shows promise for patient guidance, the inconsistent performance of models such as Perplexity highlights the need for cautious implementation and continuous oversight in healthcare communication.

Keywords: Male circumcision, large language models, patient education

Introduction

Male circumcision, one of the oldest and most widely performed surgical procedures worldwide, holds significant cultural, religious, and medical importance (1). While deeply rooted in tradition for many communities, the medical indications and ethical implications of male circumcision remain subjects of ongoing debate and divergent perspectives within the healthcare community (2).

The digital age has ushered in an era in which patients increasingly seek health-related information through online resources, a trend further amplified by the rapid emergence of large language models (LLMs) that offer seemingly instant guidance (3). From chatbots delivering immediate responses to search engine algorithms curating vast repositories of medical content, artificial intelligence (AI) is rapidly emerging as a powerful intermediary in patient education and healthcare decision-making (4). However, this growing reliance on LLM-driven information raises critical concerns, particularly in the context of sensitive and frequently debated topics such as male circumcision, where misinformation can significantly impact individual health decisions and overall well-being (5).

This article aims to address this pressing concern by critically evaluating the reliability of AI powered patient guidance on male circumcision. We hypothesized that

Corresponding Author: Ismail Ulus, MD, University of Health Sciences Türkiye, Bagcilar Training and Research Hospital, Clinic of Urology, Istanbul, Türkiye
E-mail: ismailulus06@yahoo.com **ORCID:** orcid.org/0000-0002-2005-9734
Received: 25.06.2025 **Accepted:** 12.08.2025 **Publication Date:** 29.08.2025

Cite this article as: Ulus I, Ceker G, Hacibey I. Large language models and male circumcision: a reliability assessment. Med Bull Haseki. 2025;63(3):123-127

LLMs may be insufficient in providing accurate information on male circumcision and that significant performance differences may exist among LLMs. This study evaluates the accuracy, completeness, and potential biases embedded in the information provided by various AI platforms.

Materials and Methods

Compliance with Ethical Standards

As this study is based solely on AI-generated responses and does not involve human participants or the use of personal data, ethical committee approval was waived.

Study Design and Questionnaire

This cross-sectional study evaluated the responses generated by four AI models, ChatGPT (GPT-4.5), Microsoft Copilot, Google Gemini 2.5, and Perplexity AI, in response to a set of 20 questions on 10 May 2025. These questions, covering the medical, cultural, and psychological aspects of male circumcision, are primarily based on the technical report of the American Academy of Pediatrics Task Force on Circumcision (Table 1) (6). To minimize search bias associated with user history and personalized content, the searches were conducted without signing into AI platform accounts and using the incognito mode of the Google Chrome web browser. Responses were assessed and compared by three independent experts with extensive knowledge and experience in the field, using predefined evaluation criteria.

Table 1. List of questions

Medical indications and benefits
1. What are the medical benefits of male circumcision? 2. Is circumcision effective in preventing urinary tract infections? 3. Does circumcision reduce the risk of HIV or other sexually transmitted infections? 4. Can circumcision prevent penile cancer?
Risks and complications
5. What are the potential risks and complications of circumcision? 6. Is circumcision painful? How is pain managed during and after the procedure? 7. Can circumcision lead to erectile dysfunction or sexual problems? 8. What are the signs of complications after circumcision?
Myths and misinformation
9. Does circumcision cause infertility? 10. Can circumcision reduce sexual pleasure or sensation? 11. Is it true that circumcision makes the penis longer? 12. Can the foreskin grow back after circumcision?
Lifestyle and patient concerns
13. How long is the recovery after adult circumcision? 14. When can I resume sexual activity after circumcision? 15. What kind of daily hygiene is needed after circumcision? 16. Is adult circumcision safe and common?
Pediatric and cultural aspects
17. Should newborns be circumcised? What are the pros and cons? 18. What are the cultural or religious reasons for circumcision? 19. Is it ethical to circumcise children who cannot consent? 20. Are there any alternatives to circumcision for medical conditions like phimosis?

HIV: Human immunodeficiency virus

The questions were categorized into five thematic domains: "Medical Indications and Benefits", "Risks and Complications", "Myths and Misinformation", "Lifestyle and Patient Concerns", and "Pediatric and Cultural Aspects". Each thematic domain comprises four questions that comprehensively explore the topic of male circumcision from both medical and social perspectives.

Evaluation of Responses

The responses generated by the AI models were evaluated based on the following criteria:

Relevance: the extent to which the response was appropriate and aligned with established medical evidence,

Clarity: the degree to which the information was clearly and understandably communicated.

Structure: the logical organization and coherence of the response, Utility: the usefulness of the information for patient education and general guidance.

Factual Accuracy: the consistency of the response with verified medical facts. Each criterion was assessed using a five-point Likert scale, with scores ranging from 1 (lowest) to 5 (highest).

Statistical Analysis

Statistical analyses were conducted using SPSS software, version 29.0 (IBM Corp., Armonk, NY, USA). The normality of variable distributions was evaluated using the Kolmogorov-Smirnov and Shapiro-Wilk tests, complemented by visual assessments through quantile-quantile plots and histograms. Comparisons of scores among the AI groups were performed using the Kruskal-Wallis test. For variables showing statistically significant differences, post-hoc pairwise comparisons were conducted using the Bonferroni correction to adjust for multiple testing. A p-value of <0.05 was considered statistically significant for all analyses.

Results

Inter-rater reliability among the three independent expert reviewers was assessed using a two-way mixed-effects model based on absolute agreement. The analysis yielded an intraclass correlation coefficient of 0.79 (95% confidence interval: 0.711-0.868), indicating good inter-rater agreement among the reviewers ($p<0.001$).

When overall thematic scores were aggregated, Gemini and Copilot emerged as the top performers, each achieving a mean score of 4.65. ChatGPT performed moderately with a mean score of 4.40, while Perplexity consistently underperformed, recording a total mean score of 4.06. These differences were statistically significant ($p<0.001$)

Table 2. Comparative evaluation of AI model performance across thematic domains					
	ChatGPT	**Gemini**	**Copilot**	**Perplexity**	**p-value**
Medical indications and benefits	4.40±0.38	4.53±0.44	4.25±0.90	4.40±0.48	0.798
Risks and complications	4.40±0.45^{a,c}	4.48±0.44^{a}	4.80±0.25^{b}	3.95±0.54^{c}	<0.001
Myths and misinformation	4.43±0.57^{a,b}	4.68±0.37^{a}	4.78±0.26^{a}	4.10±0.66^{b}	0.003
Lifestyle and patient concern	4.40±0.45^{a}	4.65±0.29^{a}	4.63±0.22^{a}	3.98±0.60^{b}	<0.001
Pediatric and cultural aspects	4.38±0.39^{a}	4.90±0.21^{b}	4.80±0.25^{b}	3.88±0.65^{a}	<0.001
Total	4.40±0.44^{a}	4.65±0.38^{b}	4.65±0.49^{b}	4.06±0.60^{c}	<0.001

Different superscript letters indicate statistical significance between groups
AI: Artificial intelligence

(Table 2).

Within the medical indications and benefits domain, Gemini obtained the highest mean score, while Copilot recorded the lowest; however, the differences among the models were not statistically significant (p=0.798) (Figure 1). Copilot received the highest scores in the risks and complications domain, followed by Gemini and ChatGPT, while Perplexity demonstrated the weakest performance, receiving significantly lower ratings compared to the other models (p<0.001). In the myths and misinformation domain, Copilot and Gemini outperformed ChatGPT and Perplexity (p=0.003).

In the lifestyle and patient concerns domain, Gemini and Copilot demonstrated strong and comparable performance, whereas Perplexity exhibited a marked decline. The differences among the models in this domain were statistically significant (p<0.001). Performance in the pediatric and cultural aspects domain also varied significantly across models, with Gemini achieving the highest score, closely followed by Copilot, while ChatGPT and, particularly, Perplexity received lower evaluations in this area (p<0.001).

In addition to the thematic domains, the models were evaluated based on five global evaluation criteria: relevance, clarity, structure, utility, and factual accuracy (Figure 2). Under the clarity criterion, Gemini and Copilot were rated as the most comprehensible, while Perplexity received significantly lower scores, indicating weaker performance in effectively conveying information (p<0.001) (Table 3). In terms of structural coherence, Copilot achieved the highest ratings for organization and logical flow, followed closely by Gemini, while Perplexity was rated significantly lower, indicating marked differences across the models (p<0.001).

The utility of the responses varied significantly among the models (p<0.001), with Gemini and Copilot providing the most practically useful information for patient guidance, while ChatGPT and Perplexity received lower utility scores. Perplexity received the significantly lowest rating for factual accuracy, highlighting concerns about the reliability of its responses in conveying accurate medical information (p<0.001).

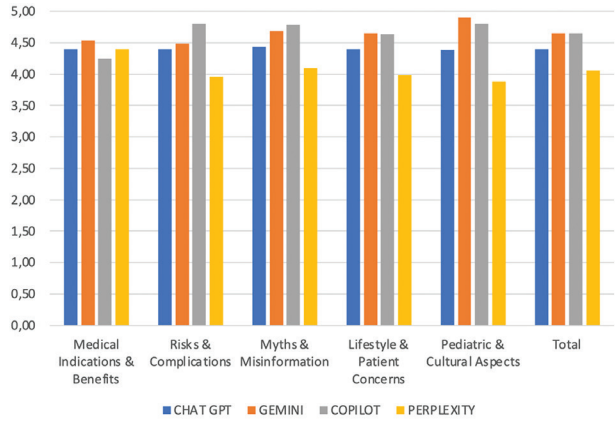

Figure 1. Mean scores of AI models across thematic domains
AI: Artificial intelligence

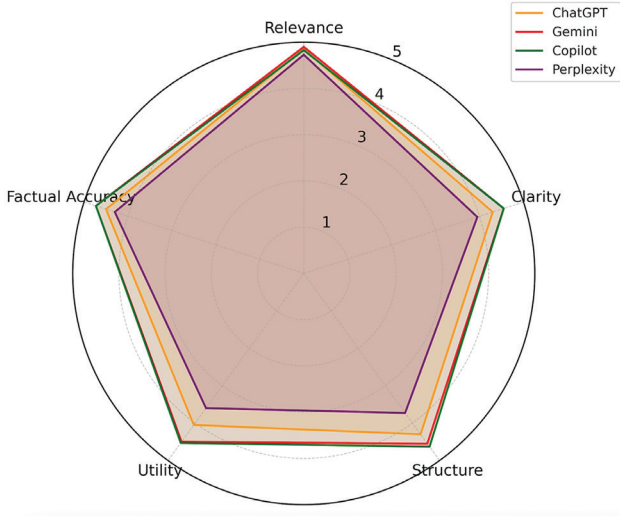

Figure 2. Radar chart comparing the performance of AI models across evaluation criteria
AI: Artificial intelligence

Table 3. Comparative evaluation of AI model performance across evaluation criteria					
	ChatGPT	**Gemini**	**Copilot**	**Perplexity**	**p-value**
Relevance	4.85±0.24	4.90±0.26	4.83±0.47	4.73±0.38	0.279
Clarity	4.30±0.34[a]	4.55±0.43[a]	4.55±0.43[a]	3.95±0.36[b]	<0.001
Structure	4.30±0.38[a]	4.55±0.36[a]	4.63±0.46[a]	3.73±0.53[b]	<0.001
Utility	4.05±0.46[a]	4.50±0.40[b]	4.53±0.53[b]	3.60±0.50[a]	<0.001
Factual accuracy	4.50±0.36[a]	4.73±0.34[a]	4.73±0.57[a]	4.3±0.47[b]	<0.001
Total	4.40±0.44[a]	4.65±0.38[b]	4.65±0.49[b]	4.06±0.60[c]	<0.001
Different superscript letters indicate statistical significance between groups AI: Artificial intelligence					

Discussion

This study presents a comprehensive comparative analysis of four LLMs in the generation of patient-facing educational content on male circumcision. The findings reveal that, although all models generated responses that were topically relevant, their performance varied substantially across the domains of clarity, structural organization, practical utility, and factual accuracy. Gemini and Copilot emerged as the most consistent and reliable in conveying medically and culturally sensitive information, whereas Perplexity demonstrated significantly lower performance across these domains.

The reliability of our expert evaluation was supported by a strong inter-rater agreement, consistent with established standards in LLM benchmarking and health communication research. While all models demonstrated high relevance scores, relevance alone proved to be an insufficient indicator of overall communication quality. Recent studies have shown that LLMs are prone to generating "hallucinated" content, plausible-sounding yet incorrect or fabricated information, particularly in complex and high-stakes domains such as medicine (7). This limitation underscores the importance of robust evaluation frameworks that extend beyond surface-level relevance. Effective patient education relies not only on topical alignment but also on the accuracy, structural clarity, and practical utility of the information and criteria that only a subset of the models in our analysis consistently fulfilled.

Emerging literature further corroborates the variable performance of AI in healthcare communication, highlighting both its potential and its limitations in generating accurate, patient-centered dialogue. Huang et al. (4) examined the diagnostic capabilities of chatbots, identifying significant limitations in their ability to manage clinical uncertainty and complex case scenarios. Similarly, Menz et al. (5) highlighted the risks associated with AI-generated health misinformation, advocating for rigorous oversight in clinical settings to mitigate potential harms.

Our study found that Copilot particularly excelled in dispelling misconceptions and explaining complications, as evidenced by its strong performance in the risks and complications as well as the myths and misinformation domains. These findings align with those of Anisuzzaman et al. (8), who demonstrated that domain-specific fine-tuning and interface design substantially impact the performance of LLMs in detecting health misinformation.

Gemini's superior performance in the pediatric and cultural aspects domain is noteworthy and suggests the benefits of enhanced contextual training tailored to these areas. This finding is consistent with observations by Kung et al. (9), who reported that newer-generation LLMs outperform earlier versions such as GPT-3.5 in United States Medical Licensing Examination style medical reasoning, particularly in tasks requiring nuanced communication.

In contrast, Perplexity's suboptimal performance, particularly in the areas of structural organization, clarity, and factual accuracy, raises important concerns regarding its readiness for deployment in healthcare-related applications. This supports the concerns raised by Thorp (10), who emphasized the unpredictability and opacity of LLMs in generating clinical advice.

Although ChatGPT demonstrated moderate performance, particularly in terms of factual accuracy, it lagged Copilot and Gemini in overall utility and the quality of patient-oriented communication. This discrepancy aligns with prior research indicating that even advanced LLMs often struggle to balance clinical precision with readability and empathetic tone in patient-facing communications (3). Empathetic communication has been shown to significantly enhance patient trust and engagement, but current LLMs remain limited in their ability to simulate empathetic dialogue in a manner that is both medically appropriate and contextually sensitive (11).

Considerable confusion exists among patients and individuals seeking health-related information regarding the reliability of available AI-based sources, and the spread of misinformation may lead to adverse health outcomes. Developing open-access health information content based on standardized guidelines, along with its classification according to specific purposes, is critically important. Future research may focus on designing alternative platforms to address this need.

Study Limitations

Several limitations warrant consideration. The evaluation was restricted to English-language content and did not incorporate assessments of emotional tone or potential biases in the models' outputs. Moreover, responses were evaluated in a controlled setting rather than through real-time user interactions, which may limit the ecological validity of our findings. Additionally, we did not examine the impact of AI-generated responses on patient decision-making, an important area that warrants future investigation. Given that algorithmic guidance can significantly influence user decision-making, even when the information provided is flawed or overly simplistic, further investigation into the behavioral impact of AI-generated content is warranted (12). The integration of intelligent systems into healthcare delivery will inevitably necessitate the development of new frameworks for accountability, transparency, and ethical oversight (13).

Conclusion

This investigation underscores the critical importance of evaluating the role of AI in shaping patient understanding of male circumcision in the digital era. Our assessment reveals that, although AI tools provide accessible medical information, their reliability remains variable. Gemini and Copilot demonstrated significantly superior performance across both thematic domains and evaluation criteria, whereas Perplexity lagged behind in all assessments. While AI can serve as a valuable adjunct for medical guidance, it should not replace the clinical judgment of healthcare professionals. Further research is needed to explore the broader implications of AI in healthcare and to develop strategies for its responsible and ethical deployment.

Ethics

Ethics Committee Approval: As this study is based solely on AI-generated responses and does not involve human participants or the use of personal data, ethical committee approval was waived.

Informed Consent: Since this study is based solely on responses generated by artificial intelligence and does not include human participant data or direct patient information, patient consent was not required.

Footnotes

Authorship Contributions

Concept: G.C., Design: G.C., I.H., Data Collection or Processing: I.U., I.H., Analysis or Interpretation: I.U., I.H., Literature Search: I.U., G.C., Writing: I.U., I.H.

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

Financial Disclosure: This study received no financial support.

References

1. Warees WM, Anand S, Leslie SW, Rodriguez AM. Circumcision. 2024 May 2. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025.
2. Wheeler R. Non-therapeutic circumcision of boys: a family matter. Arch Dis Child. 2025;11:417-8.
3. Mesko B, Győrffy Z. The rise of the empowered physician in the digital health era: viewpoint. J Med Internet Res. 2019;21:e12490.
4. Huang RS, Benour A, Kemppainen J, Leung FH. The future of AI clinicians: assessing the modern standard of chatbots and their approach to diagnostic uncertainty. BMC Med Educ. 2024;24:1133.
5. Menz BD, Modi ND, Sorich MJ, Hopkins AM. Health disinformation use case highlighting the urgent need for artificial intelligence vigilance: weapons of mass disinformation. JAMA Intern Med. 2024;184:92-6.
6. American Academy of Pediatrics Task Force on Circumcision. Male circumcision. Pediatrics. 2012;130:e756-85.
7. Yi Y, Kim KJ. The feasibility of using generative artificial intelligence for history taking in virtual patients. BMC Res Notes. 2025;18:80.
8. Anisuzzaman DM, Malins JG, Friedman PA, Attia ZI. Fine-tuning large language models for specialized use cases. Mayo Clin Proc Digit Health. 2024;3:100184.
9. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health. 2023;2:e0000198.
10. Thorp HH. ChatGPT is fun, but not an author. Science. 2023;379:313.
11. Chen D, Chauhan K, Parsa R, Liu ZA, Liu FF, Mak E, et al. Patient perceptions of empathy in physician and artificial intelligence chatbot responses to patient questions about cancer. NPJ Digit Med. 2025;8:275.
12. Ghanvatkar S, Rajan V. Evaluating explanations from AI algorithms for clinical decision-making: a social science-based approach. IEEE J Biomed Health Inform. 2024;28:4269-80.
13. Bidenko NV, Stuchynska NV, Palamarchuk YV, Matviienko MM. Integrating artificial intelligence in healthcare practice: challenges and future prospects. Wiad Lek. 2025;78:1199-205.

Original Article

DOI: 10.4274/haseki.galenos.2025.79553
Med Bull Haseki 2025;63(3):128-134

Harnessing GPT Technology for Clinical Decision Support in Retinal Detachment

Abdullah Agin*, Yucel Ozturk**, Ulviye Kivrak***

*University of Health Sciences Türkiye, Istanbul Haseki Training and Research Hospital, Clinic of Ophthalmology, Istanbul, Türkiye
**Istanbul Health and Technology University Faculty of Medicine, Department of Ophthalmology, Istanbul, Türkiye
***University of Health Sciences Türkiye, Kartal Dr. Lutfi Kirdar Training and Research Hospital, Clinic of Ophthalmology, Istanbul, Türkiye

Abstract

Aim: Considering the increasing incorporation of artificial intelligence (AI) in healthcare, it is crucial to comprehend the advantages and constraints of these technologies within ophthalmologic settings for their secure and efficient clinical utilization. This study aims to comprehensively assess the efficacy of three leading Generative Pre-trained Transformer (GPT) -based platforms in providing clinical decision-support for retinal detachment (RD).

Methods: This cross-sectional comparative study was conducted between April 2024 and May 2024. Fifty questions were created based on the American Academy of Ophthalmology "Retina Book", specifically targeting RD. The answers were produced by three different platforms and assessed by three independent reviewers who used Likert scales to evaluate their comprehensiveness and accuracy. Six readability metrics, including the Flesch-Kincaid Grade Level (FKGL) and Flesch Reading Ease Score (FRES), average words per sentence, average syllables per word, total sentence count, and total word count, were assessed.

Results: Gemini earned the most outstanding results for comprehensiveness (4.11±0.72) and accuracy (1.49±0.61), followed by ChatGPT and Copilot. ChatGPT had superior readability metrics, achieving an FKGL of 15.62±2.85 and a FRES of 62.54±12.34, establishing it as the most accessible platform. ChatGPT demonstrated significantly higher performance compared to other platforms in the metrics of average syllables per word (p=0.0421) and total word count (p=0.0115). At the same time, no significant differences were found among the platforms in the metrics of average words per sentence (p=0.0842) and total sentence count (p=0.1603). Intraclass correlation coefficient (ICC) values indicated strong inter-rater agreement for comprehensiveness (ICC >0.74) and moderate-to-high agreement for accuracy (ICC >0.56).

Conclusion: Gemini's detailed and accurate responses position it as a robust tool for professional use, while ChatGPT's superior readability makes it suitable for patient education. These findings emphasize the synergistic advantages of AI platforms in research and development management and show the necessity for hybrid systems that integrate accessibility with accuracy.

Keywords: Artificial intelligence, readability, ophthalmology, retina, retinal detachment

Introduction

Retinal detachment (RD) is an urgent condition in ophthalmology that can lead to vision loss if not treated appropriately (1). This disorder, characterized by the detachment of the neurosensory retina from the retinal pigment epithelium, requires prompt clinical and surgical care. Clinical decision-making in RD generally entails synthesizing intricate information and performing comprehensive assessments. The significance of artificial intelligence (AI) -supported Generative Pre-trained Transformer (GPT) platforms in delivering information and facilitating decision-making has attracted growing interest.

Comprehensiveness denotes the degree to which a platform delivers a thorough answer to a clinical prompt. Accuracy, simultaneously, relates to the scientific and clinical alignment of the response with information.

Corresponding Author: Abdullah Agin, MD, University of Health Sciences Türkiye, Istanbul Haseki Training and Research Hospital, Clinic of Ophthalmology, Istanbul, Türkiye
E-mail: abdullahagin@gmail.com **ORCID:** orcid.org/0000-0001-7173-8617
Received: 05.02.2025 **Accepted:** 19.06.2025 **Epub:** 12.08.2025 **Publication Date:** 29.08.2025

Cite this article as: Agin A, Ozturk Y, Kivrak U. Harnessing generative pre-trained transformer technology for clinical decision-support in retinal detachment. Med Bull Haseki. 2025;63(3):128-134

The readability levels of the responses were evaluated using metrics like the Flesch-Kincaid Grade Level (FKGL) and the Flesch Reading Ease Score (FRES). Measuring these parameters enables a multifaceted understanding of each platform's capability to deliver information.

The application of AI, supported by GPT platforms, in the medical field emerges as a novel and dynamic area of interest in the literature (2,3). This study provides critical insights into the potential impact of contemporary technologies on clinical decision-support systems and evaluates the role of these platforms in the dissemination of medical knowledge. In this work, we conducted a comparative evaluation of the performance of GPT platforms concerning the urgent condition of RD, employing a methodologically updated approach (4,5). We hypothesized that the performance of GPT -based platforms would vary significantly in terms of accuracy, depth, and readability when applied to clinical questions related to RD.

Materials and Methods

This study was designed as a cross-sectional comparative analysis. This study seeks to evaluate the performance of three distinct GPT platforms [ChatGPT (GPT-4, OpenAI, accessed April 2024), Microsoft Copilot (powered by GPT-4, *via* Edge-Browser, accessed April 2024), and Google Gemini (Gemini Advanced, based on Gemini 1.5 Pro, accessed April 2024)] using 50 questions sourced from chapter 13 (RD and other RD) of the American Academy of Ophthalmology's "Retina and Vitreous" textbook (BCSC Section 12, 2022-2023) (6) (Supplementary Document 1). Sample prompts included: (1) "What are the main surgical indications for RD?" and (2) "How is rhegmatogenous RD differentiated from exudative RD?" All prompts were manually input using the default interface of each platform. Responses were generated without follow-up questions or user interactions. Each prompt was submitted independently, and responses were collected in their default format without editing.

Three vitreoretinal surgeons (AA, YO, UK) evaluated these questions in terms of their ability to provide comprehensive information on RD, to answer accurately, and to ensure readability. The study represents a significant step toward better understanding the potential use of these platforms in medical information delivery and patient care. The answers to these inquiries were obtained by soliciting the most comprehensive responses from three GPT platforms. Three separate evaluators assessed the thoroughness and precision of these responses according to the scoring standards outlined below; then, the mean scores were computed. No ethics review committee approval was required, as this study did not access protected patient information.

Comprehensiveness

The definitions of the Likert scores were as follows: (1=very incomprehensible or very dissimilar to physician response; 2=incomprehensible or dissimilar to physician response; 3=somewhat comprehensive or somewhat like a physician response; 4=comprehensive or similar to physician response; 5=very comprehensive).

Accuracy

-Two or "Very poor": responses contain at least two pieces of incorrect information. -One or "Poor": responses contain one piece of incorrect information. Zero: no response. GPT platforms responded to every prompt, resulting in no scores of 0. 1 or "Good": responses are medically accurate but incomplete. Two or "Very good": responses are medically accurate. Furthermore, six readability metrics were evaluated using an online application.

The parameters comprised the FKGL, FRES, average words per sentence, average syllables per word, total sentence count, and total word count. Outcomes were compared across the three GPT platforms.

Statistical Analysis

The statistical studies were conducted using Statistical Package for the Social Sciences version 25 (IBM, Chicago, IL, US). Descriptive statistics were calculated, including the mean, standard deviation, minimum, and maximum values for each metric, as well as the median. A one-way analysis of variance (ANOVA) was performed to evaluate differences among the groups (ChatGPT, Microsoft Copilot, and Google Gemini) for each statistic. The post-hoc Tukey's Honestly Significant Difference Test was applied to identify group differences after significant disparities were found. The intraclass correlation coefficient (ICC) was calculated to assess inter-observer agreement on comprehensiveness and accuracy ratings among three raters for each platform. A two-way random effects model and single-rater consistency were used for the ICC calculations, with results shown alongside 95% confidence intervals. A p-value of less than 0.05 was accepted as statistically significant.

Results

Comprehensiveness and Accuracy

Table 1 displays the descriptive data for comprehensiveness and accuracy scores across the various platforms. Gemini earned the highest average comprehensiveness score of 4.11±0.715, whereas ChatGPT recorded a score of 3.61±0.795. In terms of accuracy, Gemini again led with 1.49±0.607, whereas Copilot scored the lowest average of 1.11±0.647. These results indicate statistically significant differences (ANOVA, F=7.653, p=0.00069 for comprehensiveness; F=5.993,

p=0.0031 for accuracy). The post-hoc analysis (Table 1) showed that the difference was due to Gemini's better performance. Figure 1 illustrates the comparative scores for comprehensiveness and accuracy among the ChatGPT, Copilot, and Gemini platforms.

Readability and Word Metrics

Table 2 outlines the readability and lexical metrics for the platforms, including FKGL, FRES, average words per sentence, syllables per word, and the total count of sentences and words. ChatGPT gained superior readability scores (FKGL=15.62±2.85, FRES=62.54±12.34), exceeding those of Copilot and Gemini. Analysis of variance revealed significant differences in FKGL (F=6.87, p=0.0012) and FRES (F=4.32, p=0.0168), with post-hoc tests demonstrating that ChatGPT exhibited superior readability scores compared to Gemini and Copilot. Although Gemini had marginally inferior readability metrics, it delivered more comprehensive responses, as evidenced by its word and sentence counts (23.4±8.1 sentences and 330.5±68.4 words). Figure 2 illustrates the readability metrics (FKGL and FRES) for the three platforms.

Inter-rater Reliability

The ICC values for comprehensiveness were 0.823 for ChatGPT, 0.856 for Copilot, and 0.741 for Gemini. The ICC scores for accuracy were 0.569 for ChatGPT, 0.782 for Copilot, and 0.745 for Gemini. Figure 3 shows the ICC for both comprehensiveness and accuracy, highlighting significant agreement among the raters.

Overall Performance

The thorough examination highlights Gemini's superiority in both comprehensiveness and precision, but ChatGPT wins in readability measures. The data collectively underscore the performance diversity among GPT platforms and indicate that Gemini may be more appropriate for applications necessitating precise and comprehensive medical information, especially for RD.

Discussion

Our comparative analysis revealed distinct strengths and weaknesses across GPT -based platforms when applied to RD -specific clinical questions. These differences (particularly the trade-off between factual depth and linguistic clarity) highlight practical considerations for platform selection based on user type (specialist vs. patient). The research evaluates these platforms on their comprehensiveness, accuracy, and readability, emphasizing their potential roles in clinical decision-support. Each platform's unique profile suggests context-dependent utility, for example, Gemini for clinical precision and ChatGPT for public communication. Large language models (LLMs) have begun to reshape ophthalmology workflows, especially in patient communication and rapid information retrieval. Recent studies have emphasized the efficacy of several LLMs, such as ChatGPT, Microsoft Copilot, and Google Gemini, in delivering precise and thorough solutions to clinical concerns.

Table 1. Comprehensiveness and accuracy scores

Metric	ChatGPT (Mean ± SD)	Copilot (Mean ± SD)	Gemini (Mean ± SD)	ANOVA F-value	ANOVA p-value	Post-hoc significant differences
Comprehensiveness	3.61±0.795	3.53±0.891	4.11±0.715	7.653	0.00069*	ChatGPT<Gemini Copilot<Gemini
Accuracy	1.40±0.481	1.11±0.647	1.49±0.607	5.993	0.0031*	Copilot<Gemini

*:Statistically significant
GPT: Generative Pre-trained Transformer, SD: Standard deviation, ANOVA: Analysis of variance

Table 2. Readability and word metrics

Metric	ChatGPT (Mean ± SD)	Copilot (Mean ± SD)	Gemini (Mean ± SD)	ANOVA F-value	ANOVA p-value	Post-hoc significant differences
FKGL (Flesch-Kincaid)	15.62±2.85	14.13±3.12	13.87±2.90	6.87	0.0012*	ChatGPT>Gemini ChatGPT>Copilot
FRES (Flesch score)	62.54±12.34	59.18±14.32	57.49±13.87	4.32	0.0168*	ChatGPT>Gemini
Average words per sentence	14.67±3.21	13.48±3.12	13.21±3.05	2.51	0.0842	None
Average syllables per word	1.89±0.34	1.76±0.32	1.72±0.29	3.22	0.0421*	ChatGPT>Gemini ChatGPT>Copilot
Total sentence count	25.3±8.4	22.7±7.9	23.4±8.1	1.87	0.1603	None
Total word count	375.6±65.3	340.2±72.8	330.5±68.4	4.78	0.0115*	ChatGPT > Gemini, ChatGPT > Copilot

*:Statistically significant
FKGL**:** Flesch-Kincaid Grade Level, FRES: Flesch Reading Ease Score, GPT: Generative Pre-trained Transformer, SD: Standard deviation, ANOVA: Analysis of variance

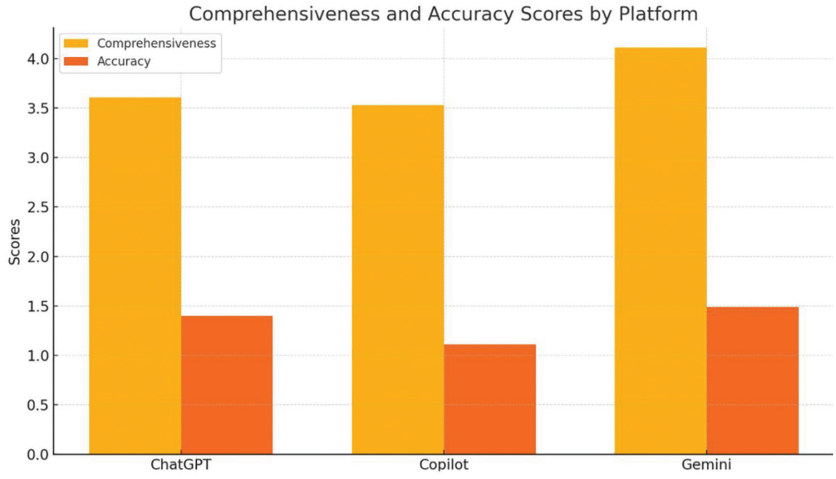

Figure 1. Comprehensiveness and accuracy scores by platform
Bar graph illustrating the mean comprehensiveness and accuracy scores assigned to responses generated by ChatGPT, Microsoft Copilot, and Google Gemini. Comprehensiveness was evaluated based on the breadth and depth of information, while accuracy reflected clinical correctness. Gemini demonstrated the highest comprehensiveness, while ChatGPT achieved relatively better accuracy than Copilot

GPT: Generative Pre-trained Transformer

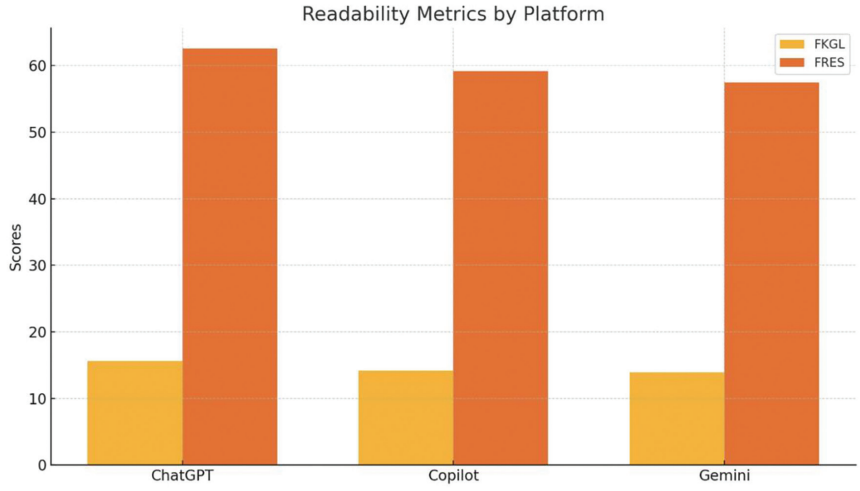

Figure 2. Readability metrics by platform
Bar graph comparing readability scores of responses from the three GPT -based platforms using Flesch-Kincaid Grade Level (FKGL) and Flesch Reading Ease Score (FRES). ChatGPT achieved the highest FRES (i.e., easiest to read), whereas Gemini's content was more complex linguistically, reflected in lower FRES and higher FKGL scores

The ramifications of these findings transcend simple performance measurements; they provoke essential inquiries regarding the application of LLMs in clinical environments. The implementation of AI-driven chatbots in patient triage has demonstrated potential. However, problems concerning the safe and effective adoption of measures, including ethical considerations, confidentiality, and physician responsibility, must be resolved (2,3). Moreover, the capacity of these models to aid in diagnosing disorders, as evidenced by research in neuro-ophthalmology and keratoconus, suggests their potential to enhance clinical practice, especially in regions with restricted access to specialists (3,7).

Furthermore, the efficacy of LLMs in educational settings, particularly in delivering preoperative information to patients undergoing ophthalmological procedures, has been examined. A comparison study showed that ChatGPT offers significantly more accurate responses than similar tools, highlighting its value as a reliable resource for patient education (4). This is especially important

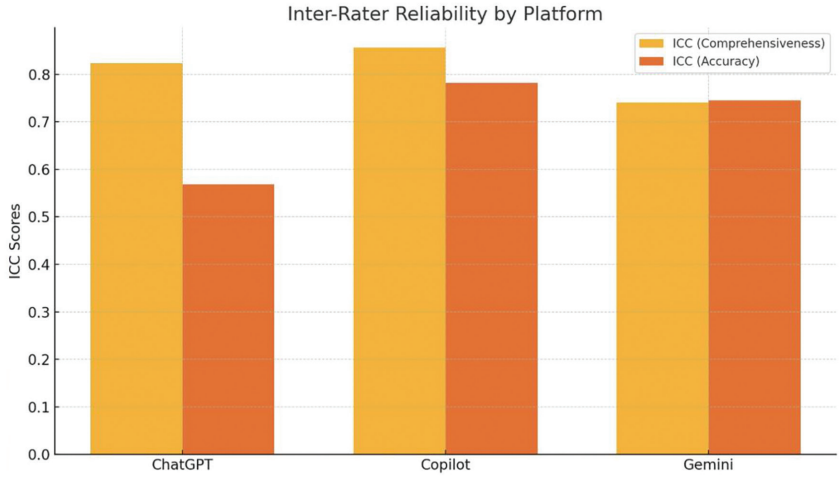

Figure 3. Inter-rater reliability by platform
Inter-rater reliability was measured using intraclass correlation coefficients (ICCs) for comprehensiveness and accuracy ratings across ChatGPT, Copilot, and Gemini. Intraclass correlation coefficient values >0.75 indicate good reliability. Copilot showed the highest agreement among reviewers for both parameters, while ChatGPT had strong agreement on comprehensiveness

GPT: Generative Pre-trained Transformer

in ophthalmology, where a patient's grasp of intricate procedures can directly affect treatment results and compliance with medical recommendations.

The clarity of responses produced by these models is a crucial element that affects their efficacy. Research indicates that whereas specific models generate more precise information, they may also offer responses that are challenging for patients to comprehend (8). This underscores the imperative for continuous enhancement of LLMs to guarantee that they deliver correct information in a way that is comprehensible to a general audience. Readability measures like FKGL and FRES are essential tools that enable developers to customize responses for various patient populations. Alongside the technical capabilities of LLMs, the human aspect is crucial in the therapeutic environment. The capacity of AI to function as a reliable intermediary between physicians and patients has been investigated, although the details of these findings are not specified here. In contrast, LLMs can improve patient education, but they must not supplant the nuanced comprehension and empathy inherent in human clinicians (9). The equilibrium between utilizing AI for efficiency and preserving the human element in healthcare is essential for cultivating trust and guaranteeing patient satisfaction.

Gemini was identified as the most comprehensive platform, achieving an average score of 4.11, surpassing both ChatGPT and Copilot. This corroborates earlier research illustrating Gemini's efficacy in delivering thorough, contextually pertinent solutions, especially in medical fields necessitating accuracy (2,10). Its ability to incorporate nuanced details makes it a valuable tool for professionals requiring in-depth information. However, Gemini's high comprehensiveness often comes at the expense of readability, as observed in earlier evaluations of AI platforms in ophthalmology (8). Although Gemini's complexity benefits experts, it may pose challenges for lay users. For example, its use of advanced terminology and longer sentences may suit ophthalmologists preparing detailed reports but could overwhelm patients seeking simplified guidance. ChatGPT possibly reflects an emphasis on lower comprehensiveness (3.61). This is due to its emphasis on delivering more general responses. While this might seem like a limitation, it positions ChatGPT as a more efficient tool for scenarios requiring concise and user-friendly information (3). Copilot's lower performance across all metrics underscores its relative limitations in clinical contexts, reaffirming its secondary role compared to Gemini and ChatGPT.

Gemini also achieved the highest accuracy score (1.49), followed by ChatGPT (1.35) and Copilot (1.11). This is consistent with prior research indicating Gemini's ability to align responses with established clinical guidelines, particularly in specialized domains such as retinal conditions and glaucoma (4,5). Its exceptional precision renders it a dependable resource for healthcare experts. Nonetheless, in the context of thyroid eye disease, Gemini's comprehensive solutions may occasionally inundate users, especially patients or non-experts (11). This level of information is advantageous in professional contexts, highlighting the necessity of customizing responses according to the user's experience level. The high correlation between accuracy and comprehensiveness

indicates that platforms excelling in one metric often perform well in the other. This relationship is particularly evident in Gemini's responses, which combine detail with adherence to clinical standards. ChatGPT, while slightly less accurate, compensates with its ability to simplify complex medical concepts, making it more accessible to non-specialists. ChatGPT outperformed its counterparts in readability, with the highest FRES (62.54), indicating easier readability, and FKGL (15.62), indicating more complex readability. However, this clashes with studies highlighting ChatGPT's capacity to simplify technical information without sacrificing essential details (12). Its user-friendly responses make it a preferred tool for patient education and public health communication. In contrast, Gemini scored lower on readability metrics, reflecting its focus on delivering detailed and technically precise information. This aligns with findings from evaluations of AI responses in refractive surgery and other ophthalmological contexts, where Gemini's advanced language constructs posed challenges for general comprehension (2,8). However, this trade-off, a key finding, could be more explicitly framed as a central consideration for potential users of these technologies. ChatGPT's superior readability makes it an ideal choice for patient-facing applications. Simplifying medical jargon and providing concise answers bridges the gap between complex medical information and patient understanding, which is a critical factor in improving health literacy (3). The study indicated that the average agreement among ophthalmologists for ChatGPT was 82.5% for both accuracy and comprehensiveness and 83.75% for clarity. Evaluations of Bard (on a prior version named Gemini) showed lower levels in agreement, with an average accuracy rate of 76.9%, comprehensiveness at 74.4%, and clarity reaching 83.8%. These findings suggest a strong consensus among evaluators, supporting the methodology and enhancing the reliability of the comparison results. Similar levels of agreement have been observed in studies comparing AI performance in ophthalmology, further strengthening the potential of these platforms as reliable decision-support tools (4,10). The distinct strengths of each platform highlight their complementary roles in clinical practice. Gemini's precision and comprehensiveness make it a valuable tool for specialists, particularly in drafting clinical reports or conducting in-depth analyses. On the other hand, ChatGPT's ability to simplify dense clinical content while maintaining factual reliability makes ChatGPT well-suited for patient education portals, informed consent preparation, and public health communication materials. The integration of AI platforms into clinical workflows aligns with broader trends in digital health, where AI is increasingly used for diagnostic support, patient communication, and personalized care (2,3). However, ethical concerns such as data security, bias in training datasets, and the potential for misinformation necessitate ongoing oversight and regulation (4).

The study's focus on RD-specific questions limits the generalizability of the findings to other ophthalmological conditions. Expanding the evaluation to include diverse subspecialties, such as cataract surgery or neuro-ophthalmology, could provide a more comprehensive understanding of these platforms' capabilities. Additionally, while Gemini excels in accuracy and detail, its limited readability highlights the need for hybrid models that combine its technical precision with ChatGPT's simplicity.

With the advancement of AI technologies in ophthalmology, it is crucial to undertake additional research to assess the long-term effects of these tools on clinical practice and patient outcomes. Research examining the comparative efficacy of various LLMs in diverse ophthalmological diseases would be essential for establishing optimal implementation procedures. Additionally, understanding the factors that influence patient adherence to treatment plans, particularly in the context of AI -assisted care, will be essential for improving therapeutic strategies.

The performance differences observed among the GPT -based platforms suggest that each may serve distinct roles depending on the clinical context. Gemini's structured, citation-rich responses indicate its potential utility in academic or professional settings such as assisting ophthalmology trainees or aiding in the drafting of consult letters where depth and precision are prioritized. In contrast, ChatGPT's balance between factual accuracy and linguistic simplicity makes it particularly suited for patient education, informed consent communication, and AI-powered triage systems. Meanwhile, Microsoft Copilot's inconsistent outputs, possibly due to integration constraints or model limitations, may currently hinder its reliability in scenarios requiring comprehensive clinical support. These findings underscore the importance of aligning platform selection with both the target user clinician versus patient and the cognitive complexity of the intended task. GPT based tools must be used with caution.

Study Limitations

Limitations include potential bias in training data, lack of individualized patient context, and the risk of misinformation. These models should augment rather than replace clinical judgment and must be used under physician oversight. While this study is limited to RD, the approach may be applicable to other ophthalmic or clinical domains. However, given the variability in complexity and terminology across subspecialties, further evaluations are warranted to assess generalizability. Despite these

limitations, the study benefits from a robust comparative design and expert evaluation by subspecialists, which strengthens the validity and relevance of the findings.

Conclusion

This comparative analysis emphasizes the combined strengths of ChatGPT, Gemini, and Copilot in addressing RD-related questions. Gemini's accuracy and comprehensiveness make it a preferred choice for professional use, while ChatGPT's readability positions it as a valuable tool. Future research should explore adaptive AI systems capable of tailoring responses to the user's expertise level, ensuring both accessibility and accuracy. Longitudinal studies assessing the impact of these platforms on clinician workflows would provide valuable insights into their practical utility. Incorporating AI-generated solutions with human supervision is essential to reduce risks and guarantee the provision of high-quality treatment.

Ethics

Ethics Committee Approval: Ethics committee approval is not required, as the study does not involve the analysis of patient data.

Informed Consent: As the study does not include the use or analysis of patient data, neither informed consent is required.

Footnotes

Authorship Contributions

Concept: A.A., Y.O., U.K., Design: A.A., Data Collection or Processing: A.A., Y.O., U.K., Analysis or Interpretation: A.A., Literature Search: A.A., Writing: A.A.

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

Financial Disclosure: This study received no financial support.

References

1. D'Amico DJ. Clinical practice. Primary retinal detachment. N Engl J Med. 2008;359:2346-54.
2. Sabaner MC, Anguita R, Antaki F, et al. Opportunities and challenges of chatbots in ophthalmology: a narrative review. J Pers Med. 2024;14:1165.
3. David D, Zloto O, Katz G, et al. The use of artificial intelligence based chat bots in ophthalmology triage. Eye (Lond). 2025;39:785-9.
4. Patil NS, Huang R, Mihalache A, et al. The ability of artificial intelligence chatbots ChatGPT and Google bard to accurately convey preoperative information for patients undergoing ophthalmic surgeries. Retina. 2024;44:950-3.
5. Cohen SA, Fisher AC, Xu BY, Song BJ. Comparing the accuracy and readability of glaucoma-related question responses and educational materials by Google and ChatGPT. J Curr Glaucoma Pract. 2024;18:110-6.
6. American Academy of Ophthalmology. Basic and clinical science course (BCSC). Section 12: Retina and vitreous. Chapter 13: Retinal detachment and other retinal disorders. San Francisco (CA): American Academy of Ophthalmology; 2022.
7. Madadi Y, Delsoz M, Lao PA, et al. ChatGPT assisting diagnosis of neuro-ophthalmology diseases based on case reports. J Neuroophthalmol. 2024:2023.09.13.23295508.
8. Aydin FO, Aksoy BK, Ceylan A, et al. Readability and appropriateness of responses generated by ChatGPT 3.5, ChatGPT 4.0, Gemini, and Microsoft Copilot for FAQs in refractive surgery. Turk J Ophthalmol. 2024;54:313-7.
9. Güler MS, Baydemir EE. Evaluation of ChatGPT-4 responses to glaucoma patients' questions: can artificial intelligence become a trusted advisor between doctor and patient?. Clin Exp Ophthalmol. 2024;52:1016-9.
10. Bahir D, Zur O, Attal L, et al. Gemini AI vs. ChatGPT: a comprehensive examination alongside ophthalmology residents in medical knowledge. Graefes Arch Clin Exp Ophthalmol. 2025;263:527-36.
11. Bahir D, Hartstein M, Zloto O, Burkat C, Uddin J, Hamed Azzam S. Thyroid eye disease and artificial intelligence: a comparative study of ChatGPT-3.5, ChatGPT-4o, and Gemini in patient information delivery. Ophthalmic Plast Reconstr Surg. 2025;41:439-44.
12. Balci AS, Yazar Z, Ozturk BT, Altan C. Performance of Chatgpt in ophthalmology exam; human versus AI. Int Ophthalmol. 2024;44:413.

Supplementary Document 1: https://d2v96fxpocvxx.cloudfront.net/a2440bda-5c5c-4e7b-8a75-abf1691c9260/content-images/ccde405f-4b3f-4195-90fd-73ef0a0a587f.pdf

Original Article

DOI: 10.4274/haseki.galenos.2025.96646
Med Bull Haseki 2025;63(3):135-140

Evaluation of Serum Vitamin and Mineral Levels in Patients with Dermatochalasis

Burcu Yakut*, Songul Kilic*, Ahu Yilmaz**, Feyza Onder*

*University of Health Sciences Türkiye, Istanbul Haseki Training and Research Hospital, Clinic of Ophthalmology, Istanbul, Türkiye
**University of Health Sciences Türkiye, Prof. Dr. Cemil Tascioglu City Hospital, Clinic of Ophthalmology, Istanbul, Türkiye

Abstract

Aim: Dermatochalasis (DC) is a common ophthalmic condition characterized by excess skin in the upper eyelids, which not only affects aesthetic appearance but can also lead to functional impairments, such as difficulty in lifting the eyelids and peripheral vision constriction. We aimed to assess serum concentrations of 25-hydroxy (OH) vitamin D, vitamin B12, folate, ferritin (Fe), and zinc (Zn) by comparing patients with DC who have undergone upper eyelid blepharoplasty to healthy controls in this retrospective, case-control study.

Methods: This retrospective comparative analysis encompassed individuals with DC who received upper eyelid blepharoplasty, alongside age- and sex-matched controls, during the period from 2022 to 2023. Data, including complete ophthalmologic evaluations and laboratory findings of DC patients and healthy controls, were extracted from medical records. Serum levels of 25-OH vitamin D, vitamin B12, folate, Fe, and Zn were assessed and compared between the DC and control groups.

Results: The DC group consisted of 59 patients, including 40 women (67.79%) and 19 men (32.21%). The control group comprised 40 individuals, including 25 women (62.5%) and 15 men (37.5%). The mean age of the participants was 58.4±8.2 years (range: 45-75). No statistically significant associations were observed between serum levels of 25-OH vitamin D, folate, vitamin B12, Fe, and Zn and the occurrence of DC.

Conclusion: This study found no significant difference between serum levels of 25-OH vitamin D, Zn, folate, vitamin B12, Fe, and the occurrence of DC. Although nutritional variables may influence overall skin health, the findings indicate that DC is not impacted solely by these vitamins and minerals. 82.8% of participants had serum 25-OH vitamin D levels below 20 ng/mL, with no statistically significant difference between the groups. This finding suggests that vitamin D deficiency may be a widespread issue but does not appear to be directly linked to the presence of DC.

Keywords: Dermatochalasis, blepharoplasty, serum biomarkers, vitamin D, zinc

Introduction

Dermatochalasis (DC) is a common ophthalmic disorder marked by an excess of skin in the upper eyelids. This condition impacts not only aesthetics but also results in functional deficits such as eyelid heaviness, peripheral vision field constriction, difficulty in lifting the eyelids, periorbital discomfort, and psychological distress (1). Dermatitis, entropion, and ectropion can also accompany DC. The underlying pathophysiology of DC remains incompletely understood. Dermal lymphatic capillaries play a vital role in the lymphatic system, facilitating fluid removal from the interstitial space and delivering proteins and macrophages within dermal tissues (2). Histopathological findings of DC are macrophage-related subclinical inflammation, elastolysis, reduced elastic fibers, disorganized collagen fiber organization, lymphocytosis, and impaired lymphatic drainage within the eyelid

Corresponding Author: Burcu Yakut, MD, University of Health Sciences Türkiye, Istanbul Haseki Training and Research Hospital, Clinic of Ophthalmology, Istanbul, Türkiye
E-mail: burcuykt@hotmail.com **ORCID:** orcid.org/0000-0003-2791-5684
Received: 27.01.2025 **Accepted:** 19.06.2025 **Epub:** 12.08.2025 **Publication Date:** 29.08.2025

Cite this article as: Yakut B , Kilic S, Yilmaz A, Onder F. Evaluation of serum vitamin and mineral levels in patients with dermatochalasis. Med Bull Haseki. 2025;63(3):135-140

(3,4). The aging of the face, particularly the eyelids, is caused by decreased regeneration ability and increased activity of degradative enzymes (5). These modifications precipitate collagen degradation, especially within the deep dermis (6). With advancing age, the components of the extracellular matrix deteriorate in their interwoven structure and cohesive interactions. This procedure can affect the upper eyelids due to the condition DC. Genetic predisposition and lifestyle variables, such as sun exposure, smoking, and inadequate nutrition, might accelerate these alterations. Nutritional deficiencies, especially concerning vitamins and minerals, are thought to affect skin elasticity and healing (7).

We hypothesized that serum deficiencies in specific vitamins and minerals, particularly vitamin D, vitamin B12, folate, zinc (Zn), and ferritin (Fe), could contribute to the development of DC. These nutrients are essential for collagen synthesis, skin integrity, and the repair of dermal tissues (8-11). Given the importance of these vitamins and minerals in maintaining skin health, we aimed to investigate whether variations in their serum levels could be associated with the occurrence of DC.

This study sought to assess serum concentrations of 25-hydroxy (OH), vitamin D, vitamin B12, folate, Fe, and Zn in patients with DC and compare them to age- and sex-matched healthy controls. By exploring the relationship between nutritional status and DC, this study will contribute to understanding the potential role of vitamin and mineral deficiencies in eyelid tissue aging and elasticity. This knowledge may guide future clinical approaches, potentially improving management strategies for patients with DC by addressing underlying nutritional factors.

Materials and Methods

Compliance with Ethical Standards

This study was conducted in accordance with the tenets of the Declaration of Helsinki. Ethical approval was obtained from the University of Health Sciences Türkiye, Istanbul Haseki Training and Research Ethics Committee (approval no.: 180-2023, date: 04.10.2023). Informed consent was obtained from all participants prior to data collection.

Study Design

This retrospective case-control study included DC patients who underwent upper eyelid blepharoplasty at University of Health Sciences Türkiye, Istanbul Haseki Training and Research Hospital between 2022 and 2023. The control group comprised age- and sex-matched individuals recruited from routine ophthalmology outpatient visits during the same period. Fifty-nine patients with DC satisfied the inclusion criteria, who possessed complete laboratory test results and electronic medical records, alongside 40 healthy controls. Comprehensive ophthalmologic examinations, including best-corrected visual acuity, intraocular pressure using Goldmann applanation tonometry, slit-lamp biomicroscopy, and fundus evaluation, were conducted. The epicrisis from the internal medicine polyclinic, morning fasting blood test laboratory data, and drug prescriptions from the past six months were analyzed from our hospital's health record dataset.

Inclusion and Exclusion Criteria

Inclusion criteria were patients diagnosed with DC who underwent upper eyelid blepharoplasty and age- and sex-matched healthy controls. Participants with systemic diseases affecting serum vitamin or mineral levels, including diabetes, cardiovascular diseases, renal failure, thyroid disorders, and gastrointestinal conditions, were excluded. Individuals using supplements or medications affecting serum levels, such as anticonvulsants, corticosteroids, or vitamin D, were also excluded. Medications or supplements that could affect serum levels of vitamin D, vitamin B12, folate, Fe, or Zn were verified through patient medical records. Patients with a history of smoking, alcohol use, or ocular conditions such as dry eye, glaucoma, uveitis, or previous ocular surgeries were excluded.

Biochemical Analysis

Serum levels of 25-OH vitamin D, vitamin B12, folate, Fe, and Zn were analyzed. Laboratory data were retrieved from fasting blood samples collected within 30 days pre- or post-surgery and were measured using automated systems. Normal ranges of serum concentration were between 15 and 55.5 ng/mL for 25-OH vitamin D, 197 and 771 ng/L for vitamin B12, 3.89 and 26.8 mg/L for folic acid, 13 and 150 ng/dL for Fe and 70 and 114 mg/dL for Zn in our laboratory. The Endocrine Society's Clinical Practice Guidelines classify blood 25-OH vitamin D levels of 20-30 ng/mL as insufficient, 10-20 ng/mL as deficient, and <10 ng/mL as severely inadequate (12).

Statistical Analysis

Data analyses were performed using Statistical Package for the Social Sciences version 25 (IBM, USA). Normality was tested using the Kolmogorov-Smirnov test. Descriptive statistics involving means and standard deviations were determined for continuous variables. Based on the data distribution, parametric and non-parametric tests, including the independent samples t-test and chi-square tests, were applied. A p-value below 0.05 was deemed statistically significant.

Results

This study included patients, comprising 59 (67.79% females: 40, 32.21% males: 19) in the DC group and 40 (62.5% females: 25, 37.5% males: 15) in the control group. Table 1 summarizes the demographic characteristics and biochemical parameters of the participants.

Analysis of serum vitamin B12, folate, Fe, and Zn levels showed no significant differences between the groups (p>0.05). The proportions of participants with deficiencies in these parameters were comparable between the DC and control groups, as detailed in Table 2.

The mean serum 25-OH vitamin D levels were 12.50±7.05 ng/mL in the DC group and 12.63±6.41 ng/mL in the control group, with no statistically significant difference between the groups (p=0.744). Regarding vitamin D deficiency, 82 (82.8%) participants had serum 25-OH vitamin D levels below 20 ng/mL. Among these, 49 (83.0%) were in the DC group and 33 (82.0%) were in the control group (p=0.418). Severe vitamin D deficiency (<10 ng/mL) was observed in 37 (37.3%) participants, including 21 (35.5%) in the DC group and 16 (40.0%) in the control group, with no statistically significant difference (Table 3).

Table 1. Demographic characteristics and biochemical parameters of the dermatochalasis patients and control group

	DC (n=59)	C (n=40)	p-value
Gender (F/M)	40/19	25/15	0.370
Age (Mean ± SD)	56.76±9.17	58.75±12.75	0.138
Vitamin D (ng/mL)	12.50±7.05	12.63±6.41	0.744
Vitamin B12 (ng/L)	326.37±144.31	333.10±170.74	0.832
Folate (mg/L)	7.14±2.67	8.88±4.27	0.056
Ferritin (ng/dL)	60.17±52.51	73.30±51.70	0.160
Zinc (mg/dL)	86.19±14.15	81.71±14.97	0.153

DC: Dermatochalasis, C: Control, F: Female, M: Male, SD: Standard deviation

Discussion

This study examined the possible correlation between serum concentrations of essential vitamins and minerals, including 25-OH vitamin D, vitamin B12, folate, Fe, and Zn, and the onset of DC. Our results indicated no statistically significant differences in the levels of these vitamins and minerals between individuals with DC undergoing blepharoplasty and healthy controls. While prior studies have implicated nutritional deficiencies and oxidative stress in skin aging and dermal tissue remodeling (13,14), the present data suggest that systemic nutritional status may not play a dominant role in the pathogenesis of DC.

Vitamin D's influence on skin integrity and repair has been extensively researched, particularly regarding its effects on keratinocyte proliferation, collagen synthesis, and anti-inflammatory characteristics (15). Vitamin D

Table 2. Vitamin and mineral deficiency levels of the dermatochalasis patients and control group

	<Cut-off n (%)	Within range n (%)	>Cut-off n (%)	Mean ± SD
Vitamin B12				
DC	3 (5.0%)	55 (93.2%)	1 (1.8%)	338.46±132.08
C	7 (17.5%)	32 (80.0%)	1 (2.5%)	333.10±170.75
Folic Acid				
DC	5 (8.5%)	54 (91.5%)	0 (0.0%)	7.15±2.67
C	3 (7.5%)	37 (92.5%)	0 (0.0%)	8.57±3.84
Ferritin				
DC	8 (13.5%)	48 (81.3%)	3 (5.2%)	60.18±52.51
C	4 (10.0%)	32 (80.0%)	4 (10.0%)	73.30±51.71
Zinc				
DC	5 (8.4%)	52 (88.1%)	2 (3.5%)	86.19±14.16
C	6 (15.0%)	33 (82.5%)	1 (2.5%)	81.71±14.98

Chi-square test was used for categorical comparisons of deficiency levels; independent samples t-test was used for comparison of mean 25-OH vitamin D levels
DC: Dermatochalasis, C: Control, SD: Standard deviation, OH: Hydroxy

Table 3. Twenty-five-hydroxy vitamin D deficiency levels of the dermatochalasis patients and control group

25-hydroxy	DC n (%)	C n (%)	DC Mean ± SD	C Mean ± SD
<10 ng/mL	21 (35.6%)	16 (40.0%)	12.50±7.05	12.63±6.41
10-20 ng/mL	28 (47.5%)	17 (42.5%)		
>20 ng/mL	10 (16.9%)	7 (17.5%)		
p-value	0.879			

DC: Dermatochalasis, C: Control, SD: Standard deviation

insufficiency is linked to compromised wound healing, increased oxidative stress, and skin fragility (13). Despite these established correlations, our investigation revealed similar mean blood 25-OH vitamin D levels between the DC and control groups (p=0.744). Furthermore, the prevalence of vitamin D deficiency (levels <20 ng/mL) and severe deficiency (<10 ng/mL) was similarly high in both groups, suggesting that vitamin D insufficiency may be a population-wide phenomenon rather than a specific factor leading to DC.

Recent histopathological evidence from Aydemir et al. (17) demonstrated that vitamin D deficiency may be associated with structural alterations in eyelid tissues of patients with DC, including reduced elastic fiber density and increased collagen disorganization. However, in contrast to their findings, our study did not identify statistically significant differences in serum 25-OH vitamin D levels between DC patients and controls. This discrepancy may arise from differences in study design, as the former directly examined tissue-level pathology while our approach relied on systemic biomarker analysis. These divergent findings suggest that local tissue-specific vitamin D activity and receptor expression may play a more critical role in DC pathogenesis than systemic serum levels alone. Future investigations integrating both biochemical and histological assessments may provide a more comprehensive understanding of vitamin D's role in periocular tissue remodeling.

Vitamin B12 and folate are crucial cofactors for DNA synthesis and homocysteine metabolism, affecting oxidative stress and inflammatory processes (18,19). Deficiencies in these vitamins can result in compromised cellular repair processes, diminished collagen synthesis, and oxidative damage (20). The current investigation found that serum vitamin B12 and folate levels were normal for the majority of patients, with no significant differences between the DC and control groups. Although a slightly higher mean folate level was detected in the control group, the difference did not reach statistical significance (p=0.056). This finding raises the possibility that subtle variations in folate metabolism may influence skin repair and elasticity, but further investigation with larger cohorts is necessary to validate this hypothesis.

Zinc and Fe are essential for collagen formation, antioxidant defense mechanisms, and dermal tissue remodeling. Zinc deficiency has been associated with delayed wound healing, corneal ulcers, and epithelial damage (21), while Fe levels indicate iron reserves and collagen integrity (22). The results revealed no significant differences in Zn (p=0.153) and Fe (p=0.160) levels between the groups. The data indicate that systemic mineral deficiencies may not be a significant factor in the development of DC. At the same time, localized alterations in collagen organization and elastic fiber integrity could be pathways.

Although no statistically significant variations were seen in biochemical markers, it is notable that most subjects displayed inadequate dietary profiles, especially concerning vitamin D. Approximately 83% of subjects had vitamin D levels below 20 ng/mL, aligning with global patterns of prevalent vitamin D insufficiency (23). The high prevalence of vitamin D deficiency observed in both groups may reflect a broader population-wide insufficiency, potentially influenced by geographic or seasonal factors, including limited sunlight exposure. These findings emphasize the necessity for extensive public health efforts to address nutritional deficiencies and enhance skin health, particularly in older populations. Addressing vitamin D insufficiency through supplementation or increased screening may be necessary.

Our findings further highlight the complex relationship between factors contributing to DC. Age is the primary risk factor, resulting in diminished regenerative ability, modified extracellular matrix composition, and increased degradation of collagen and elastic fibers (24). Although systemic nutritional status may affect skin aging, intrinsic aging processes influenced by hereditary factors, ultraviolet radiation exposure, and environmental stressors likely have a more significant impact on the development of DC (25). Another consideration is the possible impact of lifestyle factors, such as smoking, alcohol intake, and sun exposure, which were excluded from this study. While these characteristics are associated with accelerated skin aging, their contribution to DC requires additional examination in future studies with more comprehensive criteria. Systemic nutritional factors like vitamin D deficiency may contribute to general skin health; on the other hand, the development of DC is likely influenced by more localized mechanisms. Specifically, the degradation of the periorbital extracellular matrix, including collagen and elastin fibers, plays a critical role in the pathophysiology of DC. These local tissue alterations may result from the interplay of aging processes, mechanical stress, and possibly even chronic inflammation, leading to skin laxity and the characteristic eyelid changes observed in DC.

Study Limitations

The study's limitations include its retrospective methodology and relatively small sample size, which may have limited the statistical power to identify subtle differences in serum biomarkers. Additionally, the study did not account for potential confounding variables such as physical activity, dietary habits, or sunlight exposure, all of which may influence serum vitamin and mineral levels. The absence of direct histopathological analyses also restricts our ability to assess localized changes in

collagen and elastin within the eyelid tissues. Despite these limitations, this study addresses an underexplored area by systematically evaluating multiple serum vitamins and minerals in patients with DC, a condition often discussed from a purely surgical or anatomical standpoint. The use of age- and sex-matched controls enhances the validity of the comparisons. Furthermore, the exclusion of major confounding factors such as systemic disease, smoking, alcohol use, and supplementation contributes to the internal consistency and reliability of the findings. This study adds valuable insights into the systemic nutritional profiles of patients undergoing blepharoplasty and may serve as a foundation for future prospective and histopathology-based research.

Conclusion

This study found no significant relationships between serum levels of vitamin B12, 25-OH vitamin D, folate, Fe, and Zn and the presence of DC. Although nutritional variables may influence overall skin health, the findings indicate that these vitamins and minerals do not exclusively impact DC. Future investigations concentrating on localized tissue examination and the molecular mechanisms governing collagen degradation and elastin disorganization may provide novel therapeutic targets for the prevention or treatment of DC. In clinical practice, patients undergoing blepharoplasty must receive thorough evaluations, including examinations of systemic nutritional conditions, to enhance surgical outcomes and promote long-term skin health.

Ethics

Ethics Committee Approval: This study was conducted in accordance with the tenets of the Declaration of Helsinki. Ethical approval was obtained from the Clinical Research Ethics Committee of University of Health Sciences Türkiye, Istanbul Haseki Training and Research Hospital (approval no.: 180-2023, date: 04.10.2023).

Informed Consent: Informed consent was obtained from all participants prior to data collection.

Footnotes

Authorship Contributions

Surgical and Medical Practices: B.Y., Concept: B.Y., F.O., Design: B.Y., F.O., Data Collection or Processing: B.Y., S.K., A.Y., Analysis or Interpretation: B.Y., Literature Search: B.Y., Writing: B.Y.

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

Financial Disclosure: This study received no financial support.

References

1. Bhattacharjee K, Misra DK, Deori N. Updates on upper eyelid blepharoplasty. Indian J Ophthalmol. 2017;65:551-8.
2. Choi KL, Sauder DN. The role of Langerhans cells and keratinocytes in epidermal immunity. J Leukoc Biol. 1986;39:343-58.
3. Karnaz A, Katircioglu YA, Ozdemir ES, et al. The histopathological findings of patients who underwent blepharoplasty due to dermatochalasis. Semin Ophthalmol. 20;33:407-11.
4. Nagi KS, Carlson JA, Wladis EJ. Histologic assessment of dermatochalasis: elastolysis and lymphostasis are fundamental and interrelated findings. Ophthalmology 2011;118:1205-10.
5. Ko AC, Korn BS, Kikkawa DO. The aging face. Surv Ophthalmol. 2017;62:190-202.
6. Naylor EC, Watson RE, Sherratt MJ. Molecular aspects of skin ageing. Maturitas. 2011;69:249-56.
7. Cosgrove MC, Franco OH, Granger SP, Murray PG, Mayes AE. Dietary nutrient intakes and skin-aging appearance among middle-aged American women. Am J Clin Nutr. 2007;86:12-31.
8. Bikle DD. Vitamin D: newer concepts of its metabolism and function at the basic and clinical level. J Endocr Soc. 2020;4:bvz038.
9. Fischer F, Achterberg V, März A, et al. Folic acid and creatine improve the firmness of human skin in vivo. J Cosmet Dermatol. 2011;10:15-23.
10. Rasmussen HM, Johnson EJ. Nutrients for the aging eye. Clin Interv Aging. 2013;8:741-8.
11. Tisma VS, Basta-Juzbasic A, Jaganjac M, et al. Oxidative stress and ferritin expression in the skin of patients with rosacea. J Am Acad Dermatol. 2009;60:270-6.
12. Holick MF, Binkley NC, Bischoff-Ferrari HA, et al. Evaluation, treatment, and prevention of vitamin D deficiency: an Endocrine Society clinical practice guideline. J Clin Endocrinol Metab. 2011;96:1911-30.
13. Bocheva G, Slominski RM, Slominski AT. The impact of vitamin D on skin aging. Int J Mol Sci. 2021;22:9097.
14. Rostan EF, DeBuys HV, Madey DL, Pinnell SR. Evidence supporting zinc as an important antioxidant for skin. Int J Dermatol. 2002;41:606-11.
15. Wadhwa B, Relhan V, Goel K, Kochhar AM, Garg VK. Vitamin D and skin diseases: a review. Indian J Dermatol Venereol Leprol. 2015;81:344-55.
16. Aydemir E, Malkoç Şen E, Aksoy Aydemir G, Bayat AH, Karnaz A, Aydın Türk B. Relationship between histopathological findings of patients with dermatochalasis and vitamin D deficiency. Int Ophthalmol. 2024;44:309.

17. Halczuk K, Kaźmierczak-Barańska J, Karwowski BT, Karmańska A, Cieślak M. Vitamin B12-multifaceted in vivo functions and in vitro applications. Nutrients. 2023;15:2734.
18. Tiwari D, Rani A, Jha HC. Homocysteine and folic acid metabolism. In: Dubey GP, Misra K, Kesharwani RK, et al., editors. Homocysteine metabolism in health and disease. Singapore: Springer Nature Singapore; 2022. p. 3-36.
19. Aşkın Ö, Uzunçakmak TKÜ, Altunkalem N, Tüzün Y. Vitamin deficiencies/hypervitaminosis and the skin. Clin Dermatol. 2021;39:847-57.
20. Mirza GE, Sevim DG. Ocular symptoms as the initial signs of zinc deficiency. J AAPOS. 2016;20:77-8.
21. Wright JA, Richards T, Srai SK. The role of iron in the skin and cutaneous wound healing. Front Pharmacol. 2014;5:156.
22. Lips P, de Jongh RT, van Schoor NM. Trends in vitamin D status around the world. JBMR Plus. 2021;5:e10585.
23. Bocheva G, Slominski RM, Slominski AT. Neuroendocrine aspects of skin aging. Int J Mol Sci. 20:2798.
24. Jacobs LC, Liu F, Bleyen I, et al. Intrinsic and extrinsic risk factors for sagging eyelids. JAMA Dermatol. 2014;150:836-43.

Original Article

DOI: 10.4274/haseki.galenos.2025.14633
Med Bull Haseki 2025;63(3):141-149

The Impact of SGLT2 Inhibitors on Hemoglobin Levels in Type 2 Diabetes: Potential Benefits Beyond Glycemic and Renal Outcomes

Emre Hoca, Nilsu Kalayci

University of Health Sciences Türkiye, Istanbul Haseki Training and Research Hospital, Clinic of Internal Medicine, Istanbul, Türkiye

Abstract

Aim: Sodium-glucose cotransporter 2 (SGLT2) inhibitors are antidiabetic agents with proven cardiovascular and renal benefits. Recent evidence suggests they may also increase hemoglobin (Hb) levels through mechanisms beyond hemoconcentration. This study aimed to evaluate the effect of SGLT2 inhibitors on Hb levels in type 2 diabetes mellitus (T2DM) patients without anemia or advanced kidney disease.

Methods: The data of 12,511 patients who attended the diabetes outpatient clinic between November 2023 and March 2025 were scanned retrospectively. Among them, 216 T2DM patients were analyzed; 130 were using SGLT2 inhibitors (dapagliflozin or empagliflozin). Laboratory and demographic data were compared between users and non-users. Correlation and multivariate linear regression analyses were performed.

Results: Hemoglobin levels were significantly higher in SGLT2 inhibitor users than non-users (14.44±1.23 vs. 13.76±1.12 g/dL, $p<0.001$). This effect was consistent across SGLT2 agents and independent of glycemic control, renal function, or liver enzymes. Male gender and serum creatinine were also positive predictors of Hb levels. Hemoconcentration alone could not explain the increase in Hb and may reflect improved erythropoiesis via renal mechanisms.

Conclusion: Sodium-glucose cotransporter 2 inhibitor use is associated with elevated Hb levels in T2DM patients, supporting their potential role in anemia prevention in the course of chronic diseases.

Keywords: Anemia, diabetes mellitus, erythropoiesis, hemoglobins, sodium-glucose transporter 2 inhibitors

Introduction

Diabetes mellitus (DM) is one of the most common causes of mortality and morbidity worldwide, affecting people of all ages, genders, and communities, with an increasing prevalence (1,2). Many comorbidities may develop during diabetes, and these may lead to a decrease in the quality of life and an increase in mortality (3-5). For this reason, the aim of recently developed and still being developed antidiabetic drugs is to benefit from their sugar-lowering effects and prevent possible comorbidities.

Sodium-glucose cotransporter 2 (SGLT2) inhibitors are the drug group recommended for first-line treatment in almost all current guidelines in the course of type 2 diabetes (T2D) due to their favorable effects in all these conditions (6,7). Sodium-glucose cotransporter 2 inhibitors are oral antidiabetic agents that increase urinary glucose excretion by inhibiting SGLT2 cotransporters responsible for glucose reabsorption in the proximal tubules of the kidneys and provide glycemic control through an insulin-independent mechanism. The concomitant increased sodium excretion reverses tubuloglomerular feedback, and thus, intraglomerular pressure is reduced. This is the basic pathophysiology of the positive renoprotective effects of this group of drugs (8,9). In addition, osmotic diuresis is due to increased sodium excretion, and consequently, cardiovascular protective effects occur by

Corresponding Author: Emre Hoca, MD, University of Health Sciences Türkiye, Istanbul Haseki Training and Research Hospital, Department of Internal Medicine, Istanbul, Türkiye

E-mail: emrehoca89@gmail.com **ORCID:** orcid.org/0000-0003-4232-7362

Received: 12.05.2025 **Accepted:** 11.07.2025 **Epub:** 12.08.2025 **Publication Date:** 29.08.2025

Cite this article as: Hoca E, Kalayci N. The impact of SGLT2 inhibitors on hemoglobin levels in type 2 diabetes: potential benefits beyond glycemic and renal outcomes. Med Bull Haseki. 2025;63(3):141-149

reducing cardiac stress due to volume reduction (10). In numerous clinical and observational studies conducted in recent years, significant increases in hemoglobin (Hb) and hematocrit (Hct) levels have been observed in patients using SGLT2 inhibitors (11,12). These increases cannot be explained solely by hemoconcentration resulting from intravascular volume reduction caused by their diuretic effects. Increasing evidence suggests that more complex physiopathological mechanisms, such as increased erythropoietin (EPO) production from peritubular capillary cells as a result of reduced renal hypoxia, suppression of renal inflammation, reduction of oxidative stress and improvement of tubulointerstitial microcirculation, are involved in this effect.

We hypothesized that the use of SGLT2 inhibitors in patients with T2DM would be associated with higher Hb levels compared to non-users, independent of glycemic control or renal function. Investigating the effect of SGLT2 inhibitors on Hb levels may help clarify whether these agents contribute to anemia management and their established role in glycemic control and cardiorenal protection. Besides demonstrating our clinical experience, we aimed to evaluate the change in Hb levels in individuals with T2D using SGLT2 inhibitors and provide data on the hematological effects of these agents. In addition to other positive effects of SGLT2 inhibitors, it was also aimed to evaluate the possibility of their use in the correction of anemia, which can be seen in the course of some chronic diseases.

Materials and Methods

Compliance with Ethical Standards

This retrospective, cross-sectional study was approved by the University of Health Sciences Türkiye, Istanbul Haseki Training and Research Hospital Scientific Research Ethics Committee (approval no.: 58-2024, date: 01.08.2024). The study was conducted in accordance with the Declaration of Helsinki and the principles of Good Clinical Practice.

Participants

The data of 12,511 patients who attended the diabetes outpatient clinic between November 2023 and March 2025 were analyzed retrospectively. Patient and laboratory data were collected and analyzed using the hospital information operating system. All data of patients were anonymized. Being older than 18 years old, having a Hb level above 12 g/dL for women and 13 g/dL for men [by the World Health Organization's (WHO) diagnostic criteria for anemia], having T2D, and having accessible history and treatment records were defined as inclusion criteria. Hemoglobin levels below the WHO -defined anemia thresholds, iron or vitamin B12 deficiency, ongoing iron replacement or EPO therapy, hepatic failure, hematologic malignancy, and advanced renal dysfunction-specifically chronic kidney disease (CKD) stages 4-5 or acute kidney injury-as defined by the Kidney disease: Improving Global Outcomes classification were determined as exclusion criteria. Two hundred sixteen patients met the inclusion criteria and were enrolled in the study, as shown in the flow diagram (Figure 1). Due to national drug availability and regulatory approval, only dapagliflozin and empagliflozin are in routine clinical use in Türkiye. Consequently, the SGLT2 inhibitor group in this study consisted exclusively of patients receiving either of these two agents. Since only 3 patients in our cohort were using SGLT2 inhibitors as monotherapy, we included patients who used SGLT2 inhibitors in combination with other antidiabetic medications [metformin, sulfonylureas, dipeptidyl peptidase–4 (DPP-4) inhibitors, pioglitazone, and insulins]. To evaluate the independent effect of SGLT2 inhibitors on Hb levels and minimize the confounding influence of other antidiabetic drugs, we performed multiple regression analyses adjusting for concomitant medication use. Few patients were using glucagon-like peptide-1 receptor agonists; they were not included in the statistical analysis. Since SGLT2 inhibitors have been available in our country and our clinic only since 2018, it was possible to obtain data regarding the duration of their

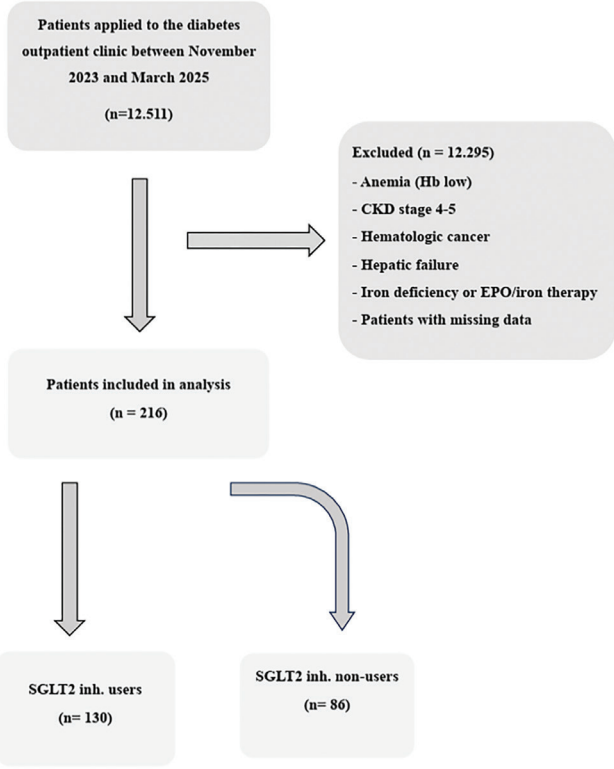

Figure 1. Flow diagram of the study
SGLT2: Sodium-glucose cotransporter 2, CKD: Chronic kidney disease, EPO: Erythropoietin, Hb: Hemoglobin

use. However, due to limitations in the hospital information system and the absence of comprehensive national drug records for earlier periods, it was not feasible to access reliable data on the duration of use for other antidiabetic medications that have been on the market for a longer time.

Data Collection

Information on patients' age, gender, and antidiabetic treatment regimens was extracted from their medical records. Laboratory results-including fasting blood samples collected after a minimum of 8 hours without food-were used to assess parameters such as complete blood count, liver enzymes [alanine aminotransferase (ALT) and aspartate aminotransferase (AST)], lipid profile [total cholesterol, low-density lipoprotein, high-density lipoprotein (HDL), triglycerides], HbA1c, fasting glucose, urea, creatinine, and spot urine protein-to-creatinine ratio. Where necessary, additional data were retrieved from the hospital's electronic medical record system to ensure completeness. Complete blood count parameters, including Hb concentration, were measured using the Mindray BC-6800 Plus Auto Hematology Analyzer (Mindray Bio-Medical Electronics Co., Ltd., Shenzhen, China). Hemoglobin concentration was determined photometrically after conversion to methemoglobin using the M-6LH LYSE reagent as part of Mindray's proprietary diluent and lysing system.

Statistical Analysis

Statistical analyses were conducted using Statistical Package for the Social Sciences version 25.0 for Windows (IBM Corp., Armonk, NY, USA). Descriptive data were summarized as mean ± standard deviation, and when appropriate, median, maximum, and minimum values were also reported to reflect variable distributions better. Categorical variables were compared using the chi-square test. The Kolmogorov-Smirnov test was used to determine the normality of continuous variables. Comparisons between the two groups were conducted using the t-test or the Mann-Whitney U test based on the data distribution. Correlative analyses between Hb and clinical or biochemical variables (e.g., age, lipid profile, renal markers) were performed using Spearman's or Pearson's correlation coefficients, selected based on the underlying distribution characteristics of each parameter. Patients were categorized as "+" for users and "-" for non-users according to medication use for subgroup analyses. In multivariate analysis, we included SGLT2 inhibitor use and relevant demographic/clinical data to explore their independent associations with Hb levels. A p-value of <0.05 was considered statistically significant.

Results

Two hundred sixteen patients with T2DM were included in the study, of whom 130 (60.2%) were using SGLT2 inhibitors and 86 (39.8%) were not. The mean age of the participants was 57.09±8.9 years, and 49.5% were female. The baseline demographic and laboratory characteristics of the patients are presented in Table 1. There were no statistically significant differences in age, gender distribution, glycemic parameters (HbA1c, glucose), or lipid profiles between the two groups (Table 2).

The mean Hb level was significantly higher in the SGLT2 inhibitor group compared to non-users (14.44±1.23 vs. 13.76±1.12 g/dL, $p<0.001$) (Table 2, Graphic 1). Subgroup analysis revealed that this increase was consistent among both dapagliflozin (14.45±1.17 g/dL) and empagliflozin (14.40±1.27 g/dL) users, (Table 3, Figure 2). In gender-stratified analysis, Hb levels were significantly higher in males than females (14.84±1.08 vs. 13.49±0.98 g/dL, $p<0.001$) (Table 3). High-density lipoprotein, cholesterol levels were significantly higher in female patients than in males (47.1±11.6 vs. 40.4±9.4 mg/dL, $p<0.001$).

More detailed subgroup analyses showed no difference in Hb levels between patients with and without other antidiabetic drugs (metformin, DPP-4 inhibitors, sulfonylureas, pioglitazone, and insulin). The distribution of concomitant antidiabetic medications differed between the SGLT2 inhibitor user group (n=130) and the non-user group (n=86). Metformin was more frequently used among SGLT2 inhibitor users than non-users (83.8% vs. 67.4%) (Table 4).

Correlation analyses demonstrated that Hb levels were negatively associated with age ($r=-0.217$, $p=0.001$), diabetes duration ($r=-0.187$, $p=0.006$) and HDL-cholesterol ($r=-0.241$, $p<0.001$), while positively correlated with creatinine ($r=0.248$, $p<0.001$), white blood cell (WBC) count ($r=0.163$, $p=0.017$), and AST ($r=0.164$, $p=0.016$) (Table 5).

In the multivariate linear regression analysis (Table 6), male gender ($B=1.160$, $p<0.001$), serum creatinine levels ($B=0.884$, $p=0.010$), and use of SGLT2 inhibitors ($B=0.643$, $p<0.001$) were found to be independent positive predictors of Hb levels. Age was identified as a significant negative predictor ($B=-0.020$, $p=0.008$).

Table 1. Baseline clinical and demographic characteristics of all participants		
Variable	**Mean ± SD**	**Median (min.-max.)**
Age (years)	57.09±8.9	58 (23-80)
Gender (F/M)	107/109	
HbA1c (%)	7.93±1.64	7.8 (5.1-15.5)
Glucose (mg/dL)	156.41±560.2	139 (58-430)
Total cholesterol (mg/dL)	177.58±41.2	176 (97-326)
HDL-cholesterol (mg/dL)	43.7±11.1	42 (23-73)
LDL-cholesterol (mg/dL)	101.04±35.17	96 (30-251)
Triglyceride (mg/dL)	169.92±104.81	141 (41-685)
Hb (g/dL)	14.17±1.23	14 (12-17.3)
WBC (10^3/μL)	8.27±2.05	7.98 (4.73-17.17)
PLT (10^3/μL)	260.77±66.15	252 (90-607)
Urea (mg/dL)	34.68±10.52	34 (12-76.1)
Creatinine (mg/dL)	0.88±0.21	0.84 (0.5-1.81)
AST (U/L)	16.18±8.99	14 (6-64)
ALT (U/L)	25.98±17.14	22 (5-146)
Spot urine P/C ratio (mg/day)	205.95±311.68	132 (8-3028)

SD: Standard deviation, F: Female, M: Male, HDL: High-density lipoprotein, LDL: Low-density lipoprotein, Hb: Hemoglobin, WBC: White blood cell, PLT: Platelet, AST: Aspartate aminotransferase, ALT: Alanine aminotransferase, P/C: Protein to creatinine ratio

Table 2. Comparison of laboratory and demographic parameters between SGLT2 inhibitor users and non-users			
Variable	**SGLT2 inh. (+) (n=130)**	**SGLT2 inh. (-) (n=86)**	**p**
Age (years)	56.72±8.78	57.66±9.27	0.449
Gender (F/M)	62/68	45/41	0.505
Diabetes duration (years)	9.83±5.11	10.12±4.83	0.675
HbA1c (%)	7.94±1.64	8.02±2.08	0.755
Glucose (mg/dL)	153.97±59.54	160.09±61.49	0.466
Total cholesterol (mg/dL)	175.05±40.00	181.41±42.87	0.268
HDL-cholesterol (mg/dL)	44.25±11.20	42.88±10.88	0.372
LDL-cholesterol (mg/dL)	97.99±34.00	105.65±36.60	0.118
Triglyceride (mg/dL)	168.17±102.60	172.57±108.61	0.763
Hb (g/dL)	14.44±1.23	13.76±1.12	**<0.001**
WBC (10^3/μL)	8.47±2.17	8.03±1.83	0.125
PLT (10^3/μL)	265.10±72.04	254.23±55.85	0.238
Creatinine (mg/dL)	0.85±0.19	0.92±0.23	**0.025**
Urea (mg/dL)	35.09±9.46	34.08±11.99	0.492
ALT (U/L)	25.00±15.30	27.45±19.60	0.386
AST (U/L)	15.59±8.05	17.07±10.23	0.238
Spot urine P/C ratio (mg/day)	197.02±318.45	218.96±303.43	0.210

Statistical significant p values were expressed in bold ($p \leq 0.05$)
SGLT2 inh.: Sodium-glucose transporter 2 inhibitors, F: Female, M: Male, HDL: High-density lipoprotein, LDL: Low-density lipoprotein, Hb: Hemoglobin, WBC: White blood cell, PLT: Platelet, AST: Aspartate aminotransferase, ALT: Alanine aminotransferase, P/C: Protein to creatinine ratio

Table 3. Subgroup comparison of hemoglobin levels by SGLT2 agent type and gender			
SGLT2 inh.	**Hb (Mean ± SD)**	**Hb [Min.-max. (median)]**	**p**
Empagliflozin (n=92)	14.40±1.27	12.2-17.1 [14.3]	0.831
Dapagliflozin (n=38)	14.45±1.17	12.1-17.3 [14.3]	
Gender			**p**
Female (n=107)	13.49±0.98	12-17 (13.3)	**<0.001**
Male (n=109)	14.84±1.08	13-17.3 (14.8)	
Statistical significant p values were expressed in bold (p≤0.05) SGLT2 inh.: Sodium-glucose transporter 2 inhibitors, Hb: Hemoglobin			

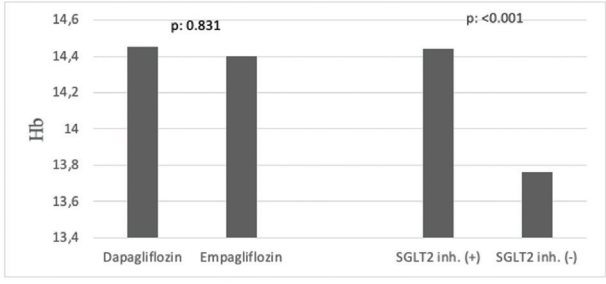

Figure 2. Comparison of mean hemoglobin levels between SGLT2 İnhibitor users and non-users
SGLT2: Sodium-glucose transporter 2, Hb: Hemoglobin

Table 4. Distribution of concomitant antidiabetic medications in SGLT2 inhibitor users and non-users			
Antidiabetic Medication	**SGLT2 inh. users (n=132)**	**Non-users (n=84)**	**p-value**
Metformin	84.1% (111)	66.7% (56)	**0.003**
DPP-4 inhibitors	53.8% (71)	47.6% (40)	0.377
Pioglitazone	26.5% (35)	34.5% (29)	0.209
Sulfonylureas	12.1% (16)	17.9% (15)	0.241
Insulin	53.8% (71)	66.7% (56)	0.061
Statistical significant p values were expressed in bold (p≤0.05) SGLT2 inh.: Sodium-glucose transporter 2 inhibitors, DPP-4: Dipeptidyl peptidase-4			

Table 5. Correlations between hemoglobin levels and clinical parameters		
	R	**p**
Age*	-0.217	**0.001**
Diabetes duration*	-0.187	**0.006**
SGLT2 inh. duration*	0.014	0.871
HbA1c*	-0.077	0.261
LDL-cholesterol*	-0.035	0.610
HDL-cholesterol*	-0.241	**<0.001**
Triglyceride*	0.103	0.132
Creatinine*	0.248	**<0.001**
Urea*	-0.047	0.489
WBC*	0.163	**0.017**
PLT*	-0.051	0.456
ALT**	0.079	0.248
AST*	0.164	**0.016**
Spot urine P/C **	-0.151	0.051
*Pearson correlation, **Spearman (non-parametric) correlation ***Statistical significant p values were expressed in bold (p ≤0.05) SGLT2 inh.: Sodium-glucose transporter 2 inhibitors, HDL: High-density lipoprotein, LDL: Low-density lipoprotein, Hb: Hemoglobin, WBC: White blood cell, PLT: Platelet, AST: Aspartate aminotransferase, ALT: Alanine aminotransferase, P/C: Protein to creatinine ratio		

Table 6. Multivariate linear regression analysis for predictors of hemoglobin levels				
	B	**S.E.**	**%95 CI for B (lower-upper)**	**p**
(Constant)	13.818	0.609	12.617-15.018	**<0.001**
Age	-0.023	0.008	-0.038- -0.007	**0.004**
Male gender	1.109	0.143	0.826-1.391	**<0.001**
Creatinine	0.843	0.357	0.139-1.547	**0.019**
HDL-cholesterol	-0.005	0.007	-0.018-0.008	0.471
AST	0.008	0.008	-0.007-0.024	0.268
SGLT2 inh.	0.685	0.136	0.418-0.953	**<0.001**
Statistical significant p values were expressed in bold (p≤0.05) SGLT2 inh.: Sodium-glucose transporter 2 inhibitors, HDL: High-density lipoprotein, AST: Aspartate aminotransferase, CI: Confidence İnterval, SE: Standard error				

Discussion

In this retrospective observational study, we evaluated the association between SGLT2 inhibitor use and Hb levels in patients with T2DM who did not have clinical anemia or advanced kidney disease. Our findings revealed that individuals using SGLT2 inhibitors had significantly higher Hb levels than non-users, even after adjusting for confounding variables such as age, sex, creatinine levels, and liver function tests. This suggests that SGLT2 inhibitors exert hematopoietic effects independent of their glucose-lowering and volume-reducing properties.

The observed mean Hb value among SGLT2 inhibitor users (14.44±1.23 g/dL) was significantly higher than that of non-users (13.76±1.12 g/dL, $p<0.001$), aligning with the results of multiple prior studies. A meta-analysis by Kanbay et al. (11) reported that SGLT2 inhibitors significantly increase Hb and Hct levels and highlighted that this effect was not solely attributable to hemoconcentration. A decrease in WBC and platelet (PLT) levels is also expected in hemoconcentration due to volume depletion (13,14). Although Hb levels were significantly higher in patients using SGLT2, the absence of a significant increase in WBC and PLT levels suggests that the Hb-raising effect of SGLT2 inhibitors cannot be explained only by hemoconcentration. While intravascular volume contraction through osmotic diuresis plays a role, recent evidence points to the involvement of renal and systemic mechanisms, including upregulation of EPO synthesis, improved renal oxygenation, and decreased renal inflammation and oxidative stress.

Type 2 diabetes mellitus is associated with an increased risk of anemia due to chronic low-grade inflammation, functional iron deficiency, elevated hepcidin levels, and reduced EPO production, even in patients without advanced kidney disease. Although diabetes duration was inversely correlated with Hb levels, it did not differ significantly between groups. Therefore, it is unlikely that diabetes duration explains the higher Hb levels observed in SGLT2 inhibitor users. Also in our analysis, no significant correlation was observed between the duration of SGLT2 inhibitor use and Hb levels. This suggests that the hematological effect of these agents may occur independently of treatment duration, reducing concerns about potential bias related to differences in drug exposure time between patients.

Mechanistically, SGLT2 inhibition reduces sodium and glucose reabsorption in the proximal renal tubules, decreasing tubular workload and oxygen consumption. This leads to a reduction in renal cortical hypoxia, a key stimulus for peritubular fibroblast-derived EPO production. Furthermore, several studies have shown that SGLT2 inhibitors can enhance hypoxia-inducible factor (HIF) pathways, further promoting erythropoiesis (15,16). Thus, the Hb rise observed in our patients likely reflects a pharmacologically induced improvement in renal oxygen sensing and erythropoietic signaling. Anemia is a common and often underrecognized comorbidity in patients with diabetes, particularly those with CKD or heart failure. Most of the studies in the literature, some of which we have emphasized above, have observed an increase in Hb values caused by SGLT2 inhibitors in CKD patients. It is important that similar effects were observed in our study group, which included patients who had not yet developed advanced CKD. Emerging evidence also suggests that SGLT2 inhibitors play a protective role against the development of anemia in patients with heart failure. Beyond their glycemic and cardiorenal benefits, these agents have been shown to stimulate erythropoiesis through enhanced EPO production, suppression of hepcidin, activation of HIF-2 , and improved iron mobilization (17). Therefore, their use in diabetic patients with or at risk of anemia due to chronic disease states may provide added hematological benefit, potentially preventing the onset of anemia even before it becomes clinically evident. The observed increase in Hb levels suggests that SGLT2 inhibitors may have a role in either early anemia prevention or correction.

Our study also demonstrated that Hb levels were independently associated with male gender and creatinine levels and inversely associated with age. These associations are well-established in the literature and reflect physiological sex-based differences in EPO response, renal reserve, and hematopoietic capacity. Interestingly, serum creatinine-within a non-pathological range-was positively correlated with Hb levels, potentially indicating that mild reductions in renal clearance may stimulate EPO production as a compensatory mechanism (18).

Although an inverse correlation was initially observed between HDL cholesterol and Hb levels in the bivariate analysis, further investigation suggests that this relationship is mainly attributable to sex-based differences. Specifically, in our sample, women had significantly higher HDL levels and significantly lower Hb levels than men-a pattern consistent with known physiological differences (19). This was confirmed by subgroup analysis showing that female sex was associated with an approximately 6.7 mg/dL higher HDL level ($p<0.001$) and also emerged as the strongest positive predictor of Hb in multivariate regression. Importantly, HDL was not independently associated with Hb after adjusting for gender and other confounding variables, indicating that the observed inverse correlation was likely spurious. These findings emphasize the value of accounting for demographic confounders, particularly gender, when interpreting associations between laboratory parameters in metabolic studies.

The lack of significant association between Hb levels and glycemic indices (HbA1c, fasting glucose) in

our analysis supports the notion that the hematologic benefits of SGLT2 inhibitors occur independently of or before glucose control. This reinforces the emerging view that SGLT2 inhibitors have pleiotropic effects extending beyond glycemic management, including cardiovascular, renal, and hematologic health improvements.

In our study, serum creatinine levels were significantly lower in patients using SGLT2 inhibitors despite the absence of significant differences in HbA1c or liver enzyme levels between groups. Although SGLT2 inhibitors are known to cause a transient rise in creatinine during the first days to weeks of therapy due to hemodynamic changes, several studies have demonstrated that renal function tends to stabilize or improve after this early phase, particularly in patients without advanced kidney disease (20,21). This suggests that our patients may have been captured during an intermediate phase of treatment-long enough for favorable changes in renal hemodynamics to emerge but not sufficient for measurable changes in glycemic or hepatic parameters to occur. Furthermore, our study's observation of lower creatinine levels in SGLT2 inhibitor users suggests that these agents may confer renal protection even in real-world patients who are not highly selected for clinical trials. This is clinically relevant because our cohort reflects everyday clinical practice, including patients with varying degrees of metabolic control and diverse comorbidities. Importantly, our findings indicate that improvements in renal parameters may occur independently of HbA1c reductions, underscoring the multifactorial benefits of SGLT2 inhibitors in routine care. Additionally, the cross-sectional design of our study precludes assessment of baseline trends.

Although the differences in AST and ALT levels between groups in our study were not statistically significant, we observed numerically lower transaminase levels among patients using SGLT2 inhibitors. This trend may be clinically relevant, particularly in the context of recent findings by Jang et al. (22), who demonstrated that SGLT2 inhibitors were associated with higher rates of non-alcoholic fatty liver disease (NAFLD) regression and reduced liver-related adverse outcomes compared to other oral antidiabetic agents in a large cohort of patients with TD2 and concomitant NAFLD. Possible mechanisms underlying this hepatoprotective effect of SGLT2 inhibitors include reduced hepatic steatosis through promotion of weight loss and redistribution of visceral fat, decreased hepatic inflammation and oxidative stress, modulation of insulin resistance, and attenuation of lipotoxicity *via* increased ketone body production and improved adipokine profiles such as increased adiponectin levels. These multifactorial effects may explain the lower liver enzyme levels observed in our SGLT2 inhibitor group, even without significant differences in glycemic control. Our findings, therefore, align with emerging evidence suggesting that SGLT2 inhibitors could offer hepatic benefits beyond their glycemic and renal effects, potentially positioning them as a favorable therapeutic option in patients with diabetes who are at risk for NAFLD or metabolic-associated liver injury.

Unexpectedly, the spot urine protein-to-creatinine ratio was higher among patients using SGLT2 inhibitors despite the well-established antiproteinuric effects of this drug class. This unexpected result may be attributed to the cross-sectional nature of the study and indication bias, as patients with higher baseline proteinuria are often preferentially treated with SGLT2 inhibitors due to their recognized renal protective effects in previous studies (23). Consequently, the SGLT2 user group in our cohort may have included a larger proportion of individuals with more advanced or active proteinuric kidney involvement at the time of drug initiation. Since baseline proteinuria levels were unavailable and longitudinal follow-up was not performed, it is impossible to determine whether SGLT2 inhibitor use led to reductions in proteinuria over time. Prospective studies are needed to assess the temporal changes in protein excretion following initiation of therapy.

Interestingly, metformin use, which is known to be associated with vitamin B12 deficiency and potentially lower Hb levels in long-term use, was significantly more common in the SGLT2 inhibitor group. Despite this, Hb levels were higher in SGLT2 inhibitor users, suggesting that the observed Hb increase is unlikely to be confounded by metformin use. Pioglitazone has been reported to cause reductions in Hb levels, primarily through hemodilution resulting from fluid retention, a known effect of peroxisome proliferator-activated receptor gamma agonists (24). This fluid retention leads to an expansion of plasma volume, thereby diluting red blood cells and lowering measured Hb and Hct values. Although rare, bone marrow suppression and decreased EPO production due to increased plasma volume have also been suggested as potential mechanisms contributing to pioglitazone-associated anemia. However, in our study, the proportion of patients using pioglitazone did not differ significantly between the SGLT2 inhibitor users and non-users. Therefore, the potential Hb-lowering effect of pioglitazone is unlikely to have significantly influenced our findings regarding Hb levels between the groups, and we believe that discussing this effect in depth is not directly relevant for interpreting the results of our analysis.

Study Limitations

Nonetheless, our study has limitations. As a retrospective, single-center study, our work has certain limitations, including the possibility of selection and information bias. We did not have access to data such as

iron stores, EPO levels, reticulocyte counts, or inflammatory markers, which would have helped clarify the mechanisms underlying the observed changes in Hb. Likewise, we were unable to assess treatment duration or adherence adequately. Despite these limitations, the significant increase in Hb levels among patients using SGLT2 inhibitors supports previous findings on the hematologic effects of these agents. More studies are needed, especially in patients with subclinical anemia or chronic conditions with impaired erythropoietic activity.

Conclusion

Our results indicate that SGLT2 inhibitor use is associated with higher Hb levels in T2D patients who do not have anemia, independent of kidney function or blood sugar control. This suggests that these drugs positively affect red blood cell production. Their potential role in preventing or improving anemia, especially in patients with chronic conditions, should be explored further in prospective studies.

Ethics

Ethics Committee Approval: This study was approved by the University of Health Sciences Türkiye, Istanbul Haseki Training and Research Hospital Scientific Research Ethics Committee (approval no.: 58-2024, date: 01.08.2024).

Informed Consent: As the study was retrospective, informed consent was not required.

Footnotes

Authorship Contributions

Surgical and Medical Practices: E.H., N.K., Concept: E.H., Design: E.H., Data Collection or Processing: E.H., N.K., Analysis or Interpretation: E.H., Literature Search: E.H., N.K., Writing: E.H.

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

Financial Disclosure: This study was conducted without any financial support from public, commercial, or non-profit funding agencies.

References

1. Hossain MJ, Al-Mamun M, Islam MR. Diabetes mellitus, the fastest growing global public health concern: early detection should be focused. Health Sci Rep. 2024;7:e2004.
2. Sun H, Saeedi P, Karuranga S, et al. IDF Diabetes Atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res Clin Pract. 2022;183:109119.
3. Cicek M, Buckley J, Pearson-Stuttard J, Gregg EW. Characterizing multimorbidity from type 2 diabetes: insights from clustering approaches. Endocrinol Metab Clin North Am. 2021;50:531-58.
4. Nowakowska M, Zghebi SS, Ashcroft DM, et al. The comorbidity burden of type 2 diabetes mellitus: patterns, clusters and predictions from a large English primary care cohort. BMC Med. 2019;17:145.
5. Hashemi R, Rabizadeh S, Yadegar A, et al. High prevalence of comorbidities in older adult patients with type 2 diabetes: a cross-sectional survey. BMC Geriatr. 2024;24:873.
6. American Diabetes Association Professional Practice Committee. 9. Pharmacologic approaches to glycemic treatment: Standards of Care in Diabetes-2024. Diabetes Care. 2024;47:S158-S78.
7. Davies MJ, Aroda VR, Collins BS, et al. Management of hyperglycemia in type 2 diabetes, 2022. A Consensus Report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetes Care. 2022;45:2753-86.
8. Fonseca-Correa JI, Correa-Rotter R. Sodium-glucose cotransporter 2 inhibitors mechanisms of action: a review. Front Med (Lausanne). 2021;8:777861.
9. Padda IS, Mahtani AU, Parmar M. Sodium-glucose transport protein 2 (SGLT2) Inhibitors. [Updated 2023 Jun 3]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025. Available from: https://www.ncbi.nlm.nih.gov/books/NBK576405/.
10. Salvatore T, Galiero R, Caturano A, et al. An overview of the cardiorenal protective mechanisms of SGLT2 inhibitors. Int J Mol Sci. 2022;23:3651.
11. Kanbay M, Tapoi L, Ureche C, et al. Effect of sodium-glucose cotransporter 2 inhibitors on hemoglobin and hematocrit levels in type 2 diabetes: a systematic review and meta-analysis. Int Urol Nephrol. 2022;54:827-41.
12. Yoshida Y, Takahashi K, Hashimoto R, Oya T, Sato J. Relationship between SGLT2 inhibitors and hemoglobin levels: a retrospective observational study. *In Vivo*. 2023;37:2327-33.
13. Patterson SM, Matthews KA, Allen MT, Owens JF. Stress-induced hemoconcentration of blood cells and lipids in healthy women during acute psychological stress. Health Psychol. 1995;14:319-24.
14. Şahin S, Diler Durgut B, Dilber B, Acar Arslan E, Kamaşak T, Cansu A. Increased hemoglobin and plateletcrit levels indicating hemoconcentration in pediatric patients with migraine. J Pediatr Neurosci. 2020;15:99-104.
15. Nespoux J, Vallon V. Renal effects of SGLT2 inhibitors: an update. Curr Opin Nephrol Hypertens. 2020;29:190-18.
16. Gajewska A, Wasiak J, Sapeda N, Młynarska E, Rysz J, Franczyk B. SGLT2 inhibitors in kidney diseases-a narrative review. Int J Mol Sci. 2024;25:4959.
17. Tziastoudi M, Pissas G, Golfinopoulos S, et al. Sodium-glucose transporter 2 (SGLT2) inhibitors and iron deficiency in heart failure and chronic kidney disease: a literature review. Life (Basel). 2023;13:2338.

18. Shin N, Kim H, Kim HJ, et al. A mild decrease of renal function is related to increased hemoglobin level during 5-year follow-up period. Korean J Intern Med. 2014;29:341-51.

19. Legato MJ. Dyslipidemia, gender, and the role of high-density lipoprotein cholesterol: implications for therapy. Am J Cardiol. 2000;86:15L-8L.

20. Meraz-Muñoz AY, Weinstein J, Wald R. eGFR decline after SGLT2 inhibitor initiation: the tortoise and the hare reimagined. Kidney360. 2021;2:1042-7.

21. Fioretto P, Zambon A, Rossato M, Busetto L, Vettor R. SGLT2 inhibitors and the diabetic kidney. Diabetes Care. 2016;39:S165-71.

22. Jang H, Kim Y, Lee DH, et al. Outcomes of various classes of oral antidiabetic drugs on nonalcoholic fatty liver disease. JAMA Intern Med. 2024;184:375-83.

23. Zeng XC, Tian Y, Liang XM, Wu XB, Yao CM, Chen XM. SGLT2i relieve proteinuria in diabetic nephropathy patients potentially by inhibiting renal oxidative stress rather than through AGEs pathway. Diabetol Metab Syndr. 2024;16:46.

24. Sinha B, Ghosal S. Assessing the need for pioglitazone in the treatment of patients with type 2 diabetes: a meta-analysis of its risks and benefits from prospective trials. Sci Rep. 2020;10:15781.

Original Article

DOI: 10.4274/haseki.galenos.2025.65265
Med Bull Haseki 2025;63(3):150-158

The Correlation Between Systemic Immune-Inflammatory Index and *Helicobacter Pylori* Infection and Its Severity

Melike Ordu*, Gulfidan Ozturk*, Mustafa Ergin**

*Aksaray University, Aksaray Training and Research Hospital, Clinic of Pathology, Aksaray, Türkiye
**Aksaray University, Aksaray Training and Research Hospital, Clinic of Gastroenterology, Aksaray, Türkiye

Abstract

Aim: The systemic immune-inflammatory index (SII), derived from routine hemogram parameters, has recently emerged as a novel marker reflecting the balance between host immune status and inflammatory burden and may offer information about the severity of *Helicobacter pylori (H. pylori)* associated gastric inflammation. We investigated whether there was a correlation between *H. pylori* inflammation, its severity, and this index.

Methods: This single-center, retrospective study was conducted between January and December 2021. A total of 1137 *H. pylori*-positive and 401 *H. pylori*-negative patients who underwent upper gastrointestinal endoscopy were included. Participants were grouped based on gastric tissue activity and chronicity scores, which reflect the histological severity of inflammation and mononuclear infiltration, respectively, as defined by the updated Sydney system. The SII and other inflammatory markers were statistically compared between groups. Correlation analysis was also performed to evaluate the relationship between histological severity and inflammatory parameters.

Results: Lymphocyte, neutrophil, platelet, SII, neutrophil-to-lymphocyte ratio, and platelet-to-lymphocyte ratio were compared between the groups. No significant differences were observed. There was no significant difference in terms of the SII and other parameters in the *H. pylori* (+) group when the "1(+)" and "2(+) and 3(+)" subgroups were compared. No significant relationship was found between tissue activity and chronicity score values and inflammatory markers.

Conclusion: No correlation was found between the presence and severity of *H. pylori* and the SII. This is the first study to compare inflammatory markers in the blood with activity and chronicity findings in the tissue.

Keywords: *Helicobacter pylori*, inflammation, systemic immune-inflammatory index, neutrophil-to-lymphocyte ratio platelet-to-lymphocyte ratio

Introduction

Helicobacter pylori (H. pylori) is a gram-negative microaerophilic bacterium (1). It tends to settle in the gastric antrum, which has a less acidic environment compared to other areas of the stomach. In the presence of *H. pylori* infection, mixed inflammation involving neutrophils, macrophages, and lymphocytes occurs in the gastric mucosa. The gold standard for *H. pylori* diagnosis is endoscopic detection of *H. pylori* in gastric biopsy materials (2). Endoscopic biopsy materials are classified and reported by a pathologist, according to the Sydney system (3). In the Sydney system, changes in the stomach lining are rated based on five key features: chronic inflammation, neutrophil activity, glandular atrophy, intestinal metaplasia, and *H. pylori* density (4). Through this classification, topographic, morphological, and etiological information is provided in a diagram. Thus, clinical diagnosis becomes more practical, and a common terminology is established between pathologists and clinicians within the reporting system.

Corresponding Author: Mustafa Ergin, MD, Aksaray University, Aksaray Training and Research Hospital, Clinic of Gastroenterology, Aksaray, Türkiye
E-mail: mstfergn@hotmail.com **ORCID:** orcid.org/0000-0002-9593-3705
Received: 17.05.2025 **Accepted:** 31.05.2025 **Epub:** 12.08.2025 **Publication Date:** 29.08.2025

Cite this article as: Ordu M, Ozturk G, Ergin M. The correlation between systemic immune-inflammatory index and *Helicobacter pylori* infection and its severity. Med Bull Haseki. 2025;63(3):150-158

Studies show that as the number of neutrophils increases, so does the severity of *H. pylori* positivity. As a result of the physiological response of leukocytes to stress, there is an increase in the number of neutrophils and a decrease in the number of lymphocytes. The neutrophil-to-lymphocyte ratio (NLR) can be used as a sensitive marker of inflammation. The systemic immune-inflammatory index (SII) is a contemporary marker for inflammation that is calculated by multiplying the number of neutrophils by the number of platelets and then dividing that by the number of lymphocytes. During inflammation, there is an increase in the number of neutrophils and platelets in the blood, which is due to megakaryocyte growth in chronic inflammation. The number of lymphocytes tends to decrease owing to increased apoptosis (5). The SII was created by combining multiple values into a single parameter. The SII is an indicator of an individual's inflammatory and immune responses (6).

The immune system is unable to eliminate the bacteria because of *H. pylori* virulence factors. In the presence of a long-term infection, chronic inflammation, oxidative stress, and DNA damage may develop in the stomach. As a result, *H. pylori* is an important predisposing factor in the development of gastric cancer. The SII has been reported to be an important prognostic marker for many solid organ tumors. In cases in which the relationship between stomach cancer and SII was evaluated, an association with increased SII was found (6).

We hypothesized that a worse *H. pylori* infection in stomach tissue samples would be linked to a higher SII. Inthis context, we investigated whether there was a correlation between *H. pylori* inflammation, its severity, and this index.

Materials and Methods

Compliance with Ethical Standards

This retrospective study was conducted in accordance with the principles of the Declaration of Helsinki and approved by the Clinical Research Ethics Committee of Aksaray University Faculty of Medicine (approval no.: 2022/07-02, date: 07.04.2022).

Study Design

The study was designed as a single-center, controlled, retrospective study. Data from 3324 patients who underwent upper gastrointestinal system endoscopy between January 2021 and December 2021 were retrospectively examined. Patients with known chronic diseases were excluded from this study. Patients under 18 years of age and those diagnosed with primary gastric carcinoma/lymphoma were not included in the study. A total of 1137 patients with *H. pylori*-positive stomach biopsies were included in the study as the patient group, while 401 patients with *H. pylori*-negative results were included as the control group. The flowchart of the study is summarized in Figure 1. During endoscopy, gastric biopsies were obtained from patients in accordance with

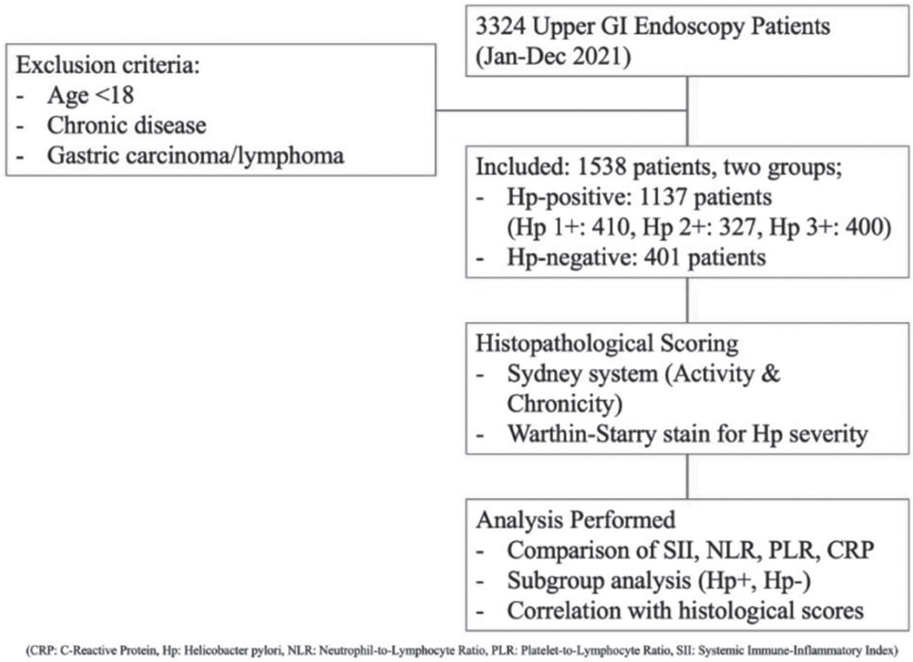

Figure 1. Flowchart of the study
GI: Gastrointestinal, Hp: *Helicobacter pylori*, SII: Systemic immune-inflammatory index, NLR: Neutrophil-to-lymphocyte ratio, PLR: Platelet-to-lymphocyte ratio, CRP: C-reactive protein

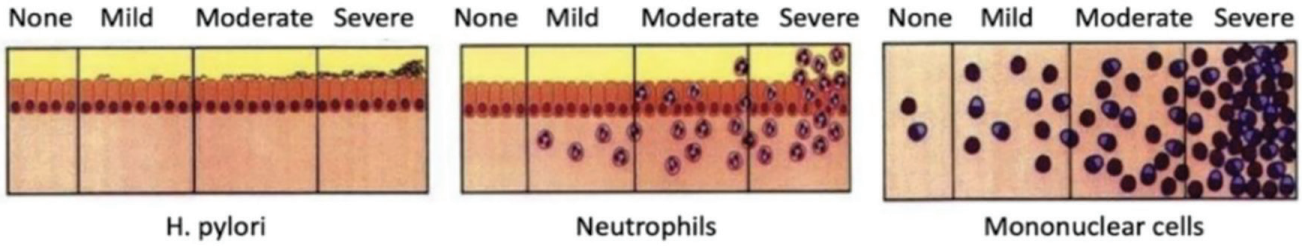

Figure 2. Sydney visual scoring, *Hp* severity, activity and chronicity severity

the Sydney protocol. The biopsy materials for the patients were classified by pathologists according to the Sydney system (Figure 2).

According to the classification, the activity and chronicity scores were categorized as "none, mild, moderate, and severe". According to activity score, it was evaluated as "mild: presence of a small number of neutrophils in the mucosa lamina propria; moderate: presence of neutrophils in the surface epithelium, foveola, and gland epithelium next to the lamina propria; severe: presence of pit abscess in addition to other findings." According to the chronicity score, it was evaluated as "normal: mononuclear cells should not exceed five at each high magnification; mild: mononuclear cells in the superficial part of the mucosa; moderate: inflammatory reaction exceeding two-thirds of the mucosa; severe: presence of inflammatory cells and lymphoid follicles in the entire mucosa layer" (7). Activity and chronicity scores are shown in the examples of our cases (Figures 3a-c).

Histochemically, the preparations stained with Warthin-Starry were categorized as "none, mild (1+), moderate (2+), or severe (3+)" using the Sydney classification. Accordingly, they were categorized as "None: *H. pylori* was not found; mild: *H. pylori* was found rarely or in less than one-third of the entire length of the sample; moderate: *H. pylori* was found in more than one-third but less than two-thirds of the entire sample length; severe: *H. pylori* was found in clusters throughout the sample." (7) The degree of *H. pylori* presence in the preparations stained histochemically with Warthin-Starry is shown in the samples from our cases (Figures 3d-g).

We obtained data from the patients included in this study from our hospital's electronic database. We recorded the patients' sex, age, and laboratory findings such as hemoglobin, leukocyte, platelet, neutrophil, lymphocyte, and C-reactive protein (CRP) (Advia 120 Siemens Healthcare Diagnostics, Eschborn, Germany). Based on the laboratory results, NLR, platelet-to-lymphocyte ratio (PLR), and SII were calculated (NLR=neutrophil count/lymphocyte count ratio, PLR=platelet count/lymphocyte count ratio, SII=platelet count x neutrophil count/lymphocyte count ratio).

Statistical Analysis

Statistical Package for the Social Sciences (SPSS) version 26.0 for Windows (SPSS, Windows 26.0, Chicago, Illinois, USA) was used. Differences between groups were compared using the chi-square test and Student's t-test. Pearson's correlation test was used to evaluate the correlation among the laboratory data points. The normality of continuous variables was assessed using the Kolmogorov-Smirnov and Shapiro-Wilk tests. The results of statistical analyses are presented as the mean ± standard deviation. Differences were considered statistically significant at $p<0.05$.

Results

Of the 1538 patients included in the study, 1137 were *H. pylori*-positive and 401 were *H. pylori*-negative. There were 410 *H. pylori* 1(+), 327 *H. pylori* 2(+), and 400 *H. pylori* 3(+) patients. There were 437 *H. pylori* (+) male and 700 female patients. There was no difference in the sex distribution between the *H. pylori* (+) and *H. pylori* (-) groups (p=0.327) (Table 1).

Age, lymphocyte count, neutrophil count, platelet count, SII, NLR, and PLR were compared between groups. We found no significant differences between the groups in terms of parameters, including SII, except for age. In terms of age, the *H. pylori* (-) and *H. pylori* 1(+) groups were statistically similar to each other and different from the *Hp*2 (+) and *Hp*3 (+) groups (Table 2 and Figures 4, 5).

The comparison of the "*H. pylori* 1(+) and *H. pylori* 2(+)" groups with the "*H. pylori* 3(+)" group showed a significant difference in age ($p<0.001$). There were no significant differences in terms of SII and other parameters (Table 3).

We found no significant relationship between tissue activity and chronicity score values, inflammatory markers, or the SII (Table 4).

Discussion

The SII, which is calculated using blood cell counts such as platelets, neutrophils, and lymphocytes, is suggested to be a better and more sensitive measure of immune-inflammatory status compared to older ratios like PLR

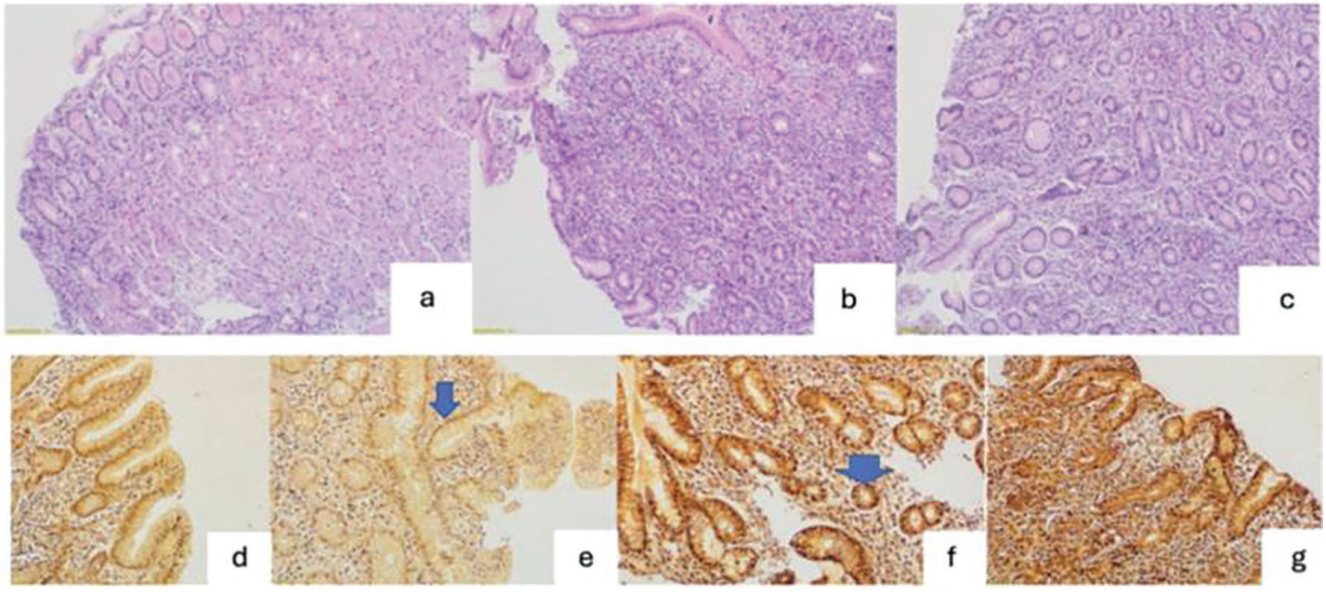

Figure 3. *H.pylori* associated chronic active gastritis; Inflammation severities in activity and chronicity in the antral mucosa. Mild (a), Moderate (b), Severe (c) findings are observed (H&E X10) d, e, f, g: Histochemically, *H.pylori* in mucus with Warthin Starry stain; None (d), Mild (e), Moderate (f), Severe (g)

H. pylori: Helicobacter pylori, H&E: Hematoxylin and eosin

Table 1. *Helicobacter pylori (Hp)* distribution by gender					
	***Hp* (-)**	***Hp*1 (+)**	***Hp*2 (+)**	***Hp*3 (+)**	**p**
Male	150 (25.6)	144 (24.5)	128 (21.8)	165 (28.1)	0.327
Female	251 (26.4)	266 (28.0)	199 (20.9)	235 (24.7)	

Table 2. Comparison of demographic, laboratory data and inflammatory markers of *Hp* (-) and *Hp* (+) groups and *Hp* (+) subgroups					
	***Hp* (-)**	***Hp*1 (+)**	***Hp*2 (+)**	***Hp*3 (+)**	**p**
Age	51.78±15.87[a] 53 (42-63)	50.67±15.36 51 (40-62)	46.67±14.99[b] 45 (35-56)	44.35±14.38 43 (33-53)	<0.00
Lymphocyte	2.26±0.75 2.2 (1.7-2.7)	2.34±0.77 2.3 (1.8-2.8)	2.45±1.8. 2.3 (1.8-2.8)	2.37±0.73 2.3 (1.9-2.8)	0.303
Neutrophil	4.44±1.64 4.1 (3.4-5.2)	4.77±2.59 4.3 (3.4-5.6)	4.57±1.83 4.2 (3.4-5.3)	4.51±1.7 4.3 (3.4-5.1)	0.406
Platelet	269.82±73.95 266 (218-309)	274.99±78 265.5 (225-315)	280.25±71.22 271(232-324)	275.99±70.68 271.5 (224.5-316.5)	0.808
SII	603.15±435.01 474.1 (359.4-684.9)	628.8±548.21 500.5 (374.2-734)	595.59±379.65 502 (374-700.5)	569.21±365.7 495.7 (362.8-670)	0.440
NLR	2.24±1.59 1.8 (1.5-2.6)	2.33±2.14 1.9 (1.4-2.6)	2.14±1.23 1.9 (1.5-2.5)	2.05±1.13 1.8 (1.5-2.3)	0.922
PLR	131.14±54.63 120.2 (96.2-153.3)	128.33±51.84 119.5 (93.7-152.9)	128.19±45.66 120.7 (97.6-151.6)	124.78±42.28 117.2 (95-147.2)	0.644

*p value from ANOVA and all others from Kruskal Wallis test. [a], and [a],[b] denotes statistically significant difference between means or medians
Mean ± standard deviation and Median 25-75%
NLR: Neutrophil-lymphocyte ratio, PLR: Platelet-lymphocyte ratio, SII: Systemic immune-inflammatory index, Hp: Helicobacter pylori,

and NLR. Elevated SII levels have been linked to increased disease activity and poor clinical outcomes in various inflammatory and malignant conditions (8-10). However, our findings revealed no significant differences in SII values between *H. pylori*-positive and *H. pylori*-negative patients, nor among patients with varying histopathological severity. The present study suggests that in the absence of systemic comorbidities, localized gastric inflammation induced by *H. pylori* may not lead to a measurable systemic inflammatory response. Although SII is often linked to

overall inflammation in the body and predicts mortality in patients with solid tumors, it may only be important in more serious cases where the immune system is not working properly. In our carefully chosen group of people who do not have cancer or long-term illnesses, it seems that the inflammation stays in the stomach lining and does not affect systemic health indicators. While an elevated SII may signal a higher risk of cancer in some individuals, its utility in identifying such risk among otherwise healthy subjects remains unclear. Prior research has identified SII as an independent prognostic marker in solid tumors (11). In the area around tumors, neutrophils help create substances that cause inflammation, and lymphocytes inhibit tumor growth, while platelets promote tumor

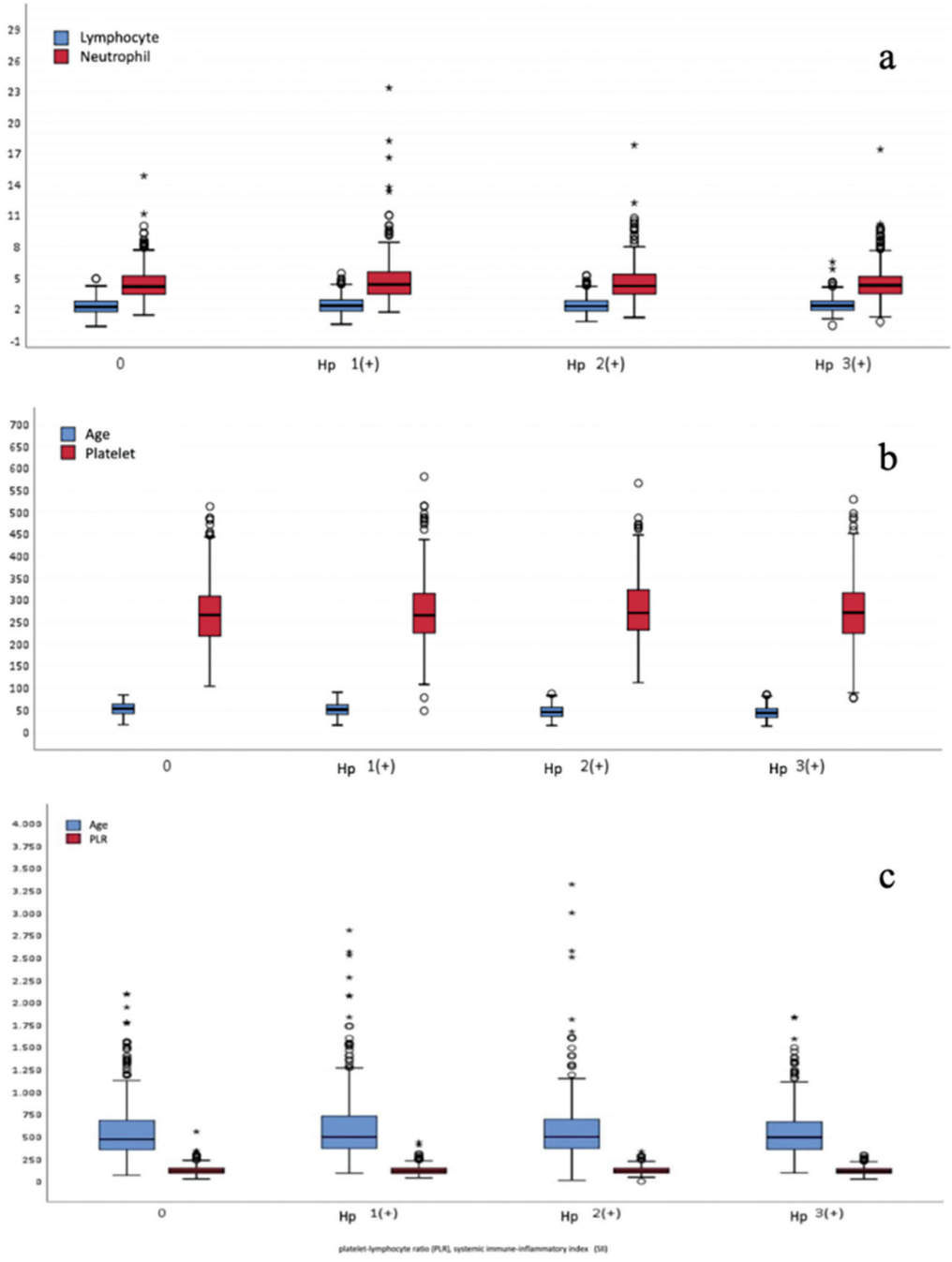

Figure 4. a. Relationship between neutrophil and lymphocyte count and H.pylori density b. Relationship between age, platelet count and H.pylori density c. Relationship between H.pylori density and SII and PLR

PLR: Platelet-lymphocyte ratio, SII: Systemic immune-inflammatory index, Hp: Helicobacter pylori

spread by releasing growth factors (6,12). In gastric cancer patients, increased SII has been associated with disease stage and outcome. Although threshold values ranging from 330 to 1600 have been proposed to predict prognosis, a universally accepted cut-off has yet to be determined (13).

Studies have indicated that SII can be used as a simple, easily accessible, and promising parameter to evaluate disease activity and the extent of disease involvement when colonoscopy cannot be performed for ulcerative colitis (14,15). However, the lack of a strong link between SII and *H. pylori*-related tissue changes, like activity and chronicity scores, raises doubts about the utility of SII in localized gastric inflammation. This finding may point to a threshold effect, below which localized inflammation does not manifest as systemic immune activation. This information indicates that SII cannot differentiate between patients who are positive for *H. pylori* and those who are negative, nor can it assess the link between *H. pylori* density and SII levels in patients who are positive for *H. pylori.*

It is known that CRP values increase in *H. pylori*-positive individuals (16-19). Research has found that substances in the body that respond to inflammation, like erythrocyte sedimentation rate, fibrinogen, and CRP, are elevated in *H. pylori*-positive patients, and there is a strong link between *H. pylori* and the body's overall inflammatory response (20,21). In our study, no significant difference was found between the *H. pylori*-positive and *H. pylori*-negative groups in terms of CRP levels. However, this may be because the preprocedural CRP value was not checked in every patient. Another important consideration is the potential disconnect between histological scoring systems and peripheral markers. While the Sydney system provides a valuable semi-quantitative assessment of gastric mucosal inflammation, it may not align proportionally with systemic indices like SII. The subjective nature of pathology scoring may contribute to this variability.

Table 3. Comparison of demographic and laboratory data and inflammatory markers between *Hp* "mild and moderate" and "severe" groups

	Mean	SD	Median	Mean	SD	Median	p
Age	48.89	15.32	49.00	44.35	14.38	43.00	<0.001
Lymphocyte	4.68	2.29	4.25	4.51	1.70	4.27	0.547
Neutrophil	2.39	1.33	2.28	2.37	0.73	2.29	0.740
Platelet	277.32	75.06	268.00	275.99	70.68	271.50	0.934
SII	614.06	480.75	501.95	569.21	365.70	495.71	0.293
NLR	2.25	1.80	1.86	2.05	1.13	1.83	0.405
PLR	128.27	49.16	120.25	124.78	42.28	117.21	0.561

NLR: Neutrophil-lymphocyte ratio, PLR: Platelet-lymphocyte ratio, SII: Systemic immune-inflammatory index, *Hp: Helicobacter pylor*i, SD: Standard deviation

Table 4. Comparison of tissue activity and chronicity and inflammatory markers

		SII		NLR		PLR	
Tissue activity	None	603.15±435.01	474.08 (72.69-4574.64)	2.24±1.59	1.85 (0.53-21.03)	131.14±54.63	120.19 (31.9-559.38)
	Mild	630.71±557.31	501.51 (97.01-7394.17)	2.33±2.17	1.86 (0.57-27.8)	128.41±50.57	118.39 (42.33-441.12)
	Moderate	609.78±448.7	516,95 (16.26-5186.21)	2.21±1.47	1.91 (0.14-16.89)	128.27±47.45	121.16 (3.61-408)
	Severe	555.29±281.05	490,92 (99.9-1831.16)	1.99±0.87	1.81 (0.32-6.73)	124.52±42.3	117.48 (30-302.63)
p		0.501		0.319		0.816	
Tissue chronicity	None	594.83±455.83	471.19 (72.69-4574.64)	2.2±1.68	1.84 (0.53-21.03)	130.05±56.43	117.35 (31.9-559.38)
	Mild	642.85±561.45	496.91 (136.42-7394.17)	2.42±2.22	1.88 (0.62-27.8)	130.21±50.08	121.16 (44.56-316.67)
	Moderate	580.3±352.14	498.51 (16.26-3324.53)	2.09±1.14	1.84 (0.14-10.22)	127.24±48.12	119,63 (3.61-441.12)
	Severe	586.16±403.39	499.46 (99.9-5186.21)	2.1±1.26	1.83 (0.32-16.89)	125.47±41.91	118.5 (30-298.06)
p		0.477		0.206		0.923	

NLR: Neutrophil-lymphocyte ratio, PLR: Platelet-lymphocyte ratio, SII: Systemic immune-inflammatory index

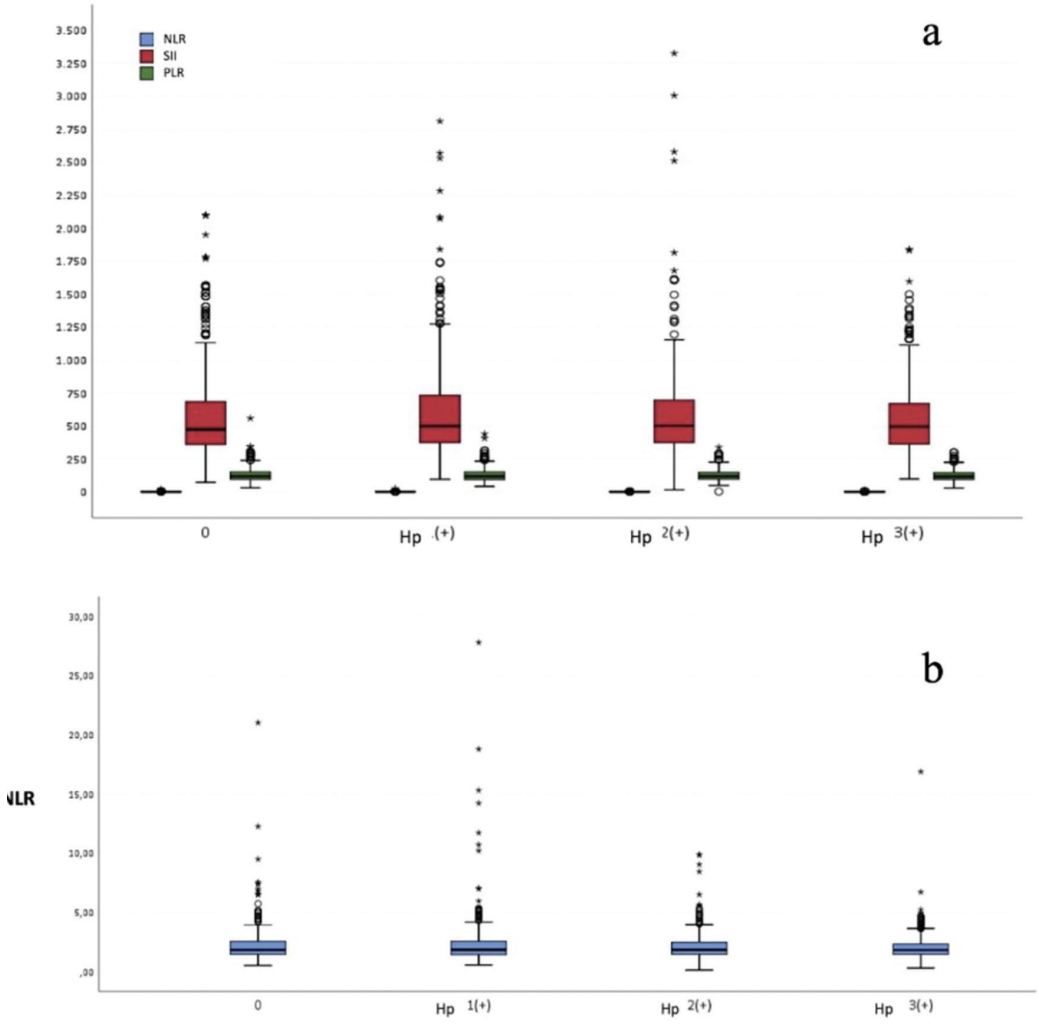

Figure 5. a. Relationship between *H.pylori* density and SII, NLR and PLR b. Relationship between *H.pylori* density and NLR

H. pylori: Helicobacter pylori,, NLR: Neutrophil-lymphocyte ratio, SII: Systemic Immune-inflammatory index, PLR: Platelet-lymphocyte ratio,

There are conflicting findings in the literature regarding CRP levels. This is because CRP levels may vary depending on age, sex, ethnicity, lifestyle, and comorbidities (22). In some studies, an insufficient number of patients may have caused this situation. Studies may differ in their CRP calculation methods, with many lacking a detailed explanation. The large number of patients in our study minimized the effect of at least one of these limitations, although it did not evaluate the CRP level of each patient before endoscopy, preventing the acquisition of clear data in the findings.

In recent years, the NLR has been frequently used as a parameter to indicate the presence of inflammation. In light of our results, tissue biopsy remains essential for the accurate assessment of *H. pylori* infection severity. Even though SII shows potential as a useful marker in various inflammatory and cancer situations, our results indicate that it cannot effectively distinguish mild to moderate mucosal inflammation from *H. pylori*, particularly in healthy people. In one study, a correlation was found between the severity of *H. pylori* gastritis and NLR (21). In another study, leukocyte, neutrophil, and NLR rates were higher in *H. pylori*-positive patients than in *H. pylori*-negative patients (4,22). It has been reported in the literature that platelets are activated by *H. pylori* infection and that *H. pylori* damage to tissues is associated with inflammatory mediators resulting from platelet activation (23,24). Researchers conducted these studies with a limited number of patient groups.

In our study, no significant differences were observed in NLR and PLR values between *H. pylori*-positive and *H. pylori*-negative groups. This finding may be attributed to the relatively large and heterogeneous study population, which likely minimized the influence of minor inflammatory

fluctuations. When we looked at the SII values for both *H. pylori*-positive and *H. pylori*-negative patients, and also within the *H. pylori*-positive groups based on bacterial density [1 (+), 2 (+), and 3 (+)], we did not find any significant differences. These results further suggest that localized gastric inflammation may not sufficiently impact systemic inflammatory markers in otherwise healthy individuals.

It is known that *H. pylori* can cause gastric cancer. Our study found no relationship between the SII and the severity of *H. pylori*. This may be related to the fact that the tumor microenvironment has not been formed, or it may be that SII is associated with cancer only above a certain threshold value. The inflammatory reaction caused by *H. pylori* may have remained below this threshold.

Study Limitations

Although only patients with dyspeptic complaints and no known comorbidities were included, the limitations of this study include the lack of information about the conditions that may affect SII, such as the patients' nutritional characteristics and the recent use of antibiotics and anti-inflammatory drugs. In addition, the retrospective design and reliance on single-timepoint laboratory values without serial follow-up may have reduced the sensitivity of inflammatory indices in reflecting disease severity. Despite these limitations, this study has several strengths. First, the large sample size and histologically verified diagnosis of *H. pylori* infection provide robust and reliable data. Second, the strict exclusion of patients with chronic diseases or malignancies allowed for a more focused assessment of the relationship between local gastric inflammation and systemic inflammatory markers in otherwise healthy individuals. Finally, the use of standardized histopathological grading according to the updated Sydney system enhances the reproducibility and clinical relevance of the findings.

Conclusion

We concluded that SII and other inflammation values are not sufficient to distinguish *H. pylori*-positive patients or to serve as a standalone prognostic marker in *H. pylori* follow-up. Tissue biopsy continues to be the most reliable method for diagnosing *H. pylori* and assessing its density. This study is the first to compare the tissue activity and chronic severity of an *H. pylori* infection with serum inflammatory markers. We have observed that the findings and severity of a gastric biopsy do not correlate with inflammatory values in the blood.

Ethics

Ethics Committee Approval: This study was approved by the Aksaray University Faculty of Medicine Clinical Research Ethics Committee (approval no.: 2022/07-02, date: 07.04.2022)

Informed Consent: The informed consent process was not required for the study because the research is a retrospective analysis.

Footnotes

Authorship Contributions

Surgical and Medical Practices: M.E., Concept: M.O., Design: M.O., G.O., Data Collection or Processing: M.O., G.O., Analysis or Interpretation: M.O., G.O., M.E., Literature Search: M.O., G.O., M.E., Writing: M.O., M.E.

Conflict of Interest: The authors have no conflict of interest to declare.

Financial Disclosure: This study received no financial support.

References

1. Ali A, AlHussaini KI. *Helicobacter pylori*: a contemporary perspective on pathogenesis, diagnosis and treatment strategies. Microorganisms. 2024;12:222.
2. Haider SI, Akhtar N, Saleem M, et al. Diagnosis of "'*Helicobacter pylori* infection of the gastric biopsy'" by rapid urease test, histopathology and Raman spectroscopy. Diagn Microbiol Infect Dis. 2024;108:116129.
3. Stolte M, Meining A. The updated Sydney system: classification and grading of gastritis as the basis of diagnosis and treatment. Can J Gastroenterol. 2001;15:591-8.
4. Akarsu M, Dikker O. Relationship between Helicobacter pylori positivity, neutrophil-to-lymphocyte ratio, and mean platelet volume. Dicle Med J. 2019;46:201-7.
5. Gawaz M, Langer H, May AE. Platelets in inflammation and atherogenesis. J Clin Invest. 2005;115:3378-84.
6. Koyasu S, Moro K. Role of innate lymphocytes in infection and inflammation. Front Immunol. 2012;3:101.
7. Hirahara N, Tajima Y, Matsubara T, et al. Systemic immune-inflammation index predicts overall survival in patients with gastric cancer: a propensity score-matched analysis. J Gastrointest Surg. 2021;25:1124-33.
8. Yang R, Chang Q, Meng X, Gao N, Wang W. Prognostic value of systemic immune-inflammation index in cancer: a meta-analysis. J Cancer. 2018;9:3295-302.
9. Zhong JH, Huang DH, Chen ZY. Prognostic role of systemic immune-inflammation index in solid tumors: a systematic review and meta-analysis. Oncotarget. 2017;8:75381-8.
10. Menyhart O, Fekete JT, Győrffy B. Inflammation and colorectal cancer: a meta-analysis of the prognostic significance of the systemic immune-inflammation index (SII) and the systemic inflammation response index (SIRI). Int J Mol Sci. 2024;25:8441.
11. Fest J, Ruiter R, Mulder M, et al. The systemic immune-inflammation index is associated with an increased risk of incident cancer-a population-based cohort study. Int J Cancer. 2020;146:692-8.

12. Hou R, Wu X, Wang C, et al. Tumor-associated neutrophils: critical regulators in cancer progression and therapeutic resistance (Review). Int J Oncol. 2025;66:28.
13. Ercan Z, Ayli MD. Systemic immune inflammation index may predict mortality in dialysis patients. Turk J Clin Lab. 2023;14:392-8.
14. Adanır H, Akıncıoğlu P. The significance of the systemic immune-inflammation index in ulcerative colitis. Dicle Medical Journal. 2022;49:521–8.
15. Zeng QY, Qin Y, Shi Y, et al. Systemic immune-inflammation index and all-cause and cause-specific mortality in sarcopenia: a study from National Health and Nutrition Examination Survey 1999-2018. Front Immunol. 2024;15:1376544.
16. Alpsoy Ş, Oran M, Mete R, et al. The relationship between endothelial dysfunction and serum levels of CRP and vitamin B12 in *Helicobacter pylori*-positive subjects. Genel Tıp Derg. 2013;23:81-6.
17. Huang J, Yang X, Zhang J, Wang F. The correlation between *Helicobacter pylori* immunoglobulin G seropositivity and plasma homocysteine levels in adults. Mediators Inflamm. 2023;2023:7590549.
18. Jafarzadeh A, Hassanshahi GH, Nemati M. Serum levels of high-sensitivity C-reactive protein (hs-CRP)in *Helicobacter pylori*-infected peptic ulcer patients and its association with bacterial CagA virulence factor. Dig Dis Sci. 2009;54:2612-6.
19. Ishida Y, Suzuki K, Taki K, et al. Significant association between *Helicobacter pylori* infection and serum C-reactive protein. Int J Med Sci. 2008;5:224-9.
20. Jia Y, Qu B, Wang Z, Han X, Ren G. Effects of active and latent *H. pylori* infection coupled with chronic alcohol ingestion on cytokine profiles and markers of oxidative balance in men seropositive for *H. pylori* CagA Ab: an observational study. Medicine (Baltimore). 2018;97:e11991.
21. Guclu M, Faruq Agan A. Association of severity of *Helicobacter pylori* infection with peripheral blood neutrophil to lymphocyte ratio and mean platelet volume. Euroasian J Hepatogastroenterol. 2017;7:11-6.
22. Farah R, Khamisy-Farah R. Association of neutrophil to lymphocyte ratio with presence and severity of gastritis due to *Helicobacter pylori* infection. J Clin Lab Anal. 2014;28:219-23.
23. Watanabe J, Kotani K. The effect of *Helicobacter pylori* eradication on C-reactive protein: results from a meta-analysis. Arch Med Sci. 2021;18:958-64.
24. de Maat MP, Kluft C. Determinants of C-reactive protein concentration in blood. Ital Heart J. 2001;2:189-95.

Original Article

DOI: 10.4274/haseki.galenos.2025.19484
Med Bull Haseki 2025;63(3):159-164

The Effects of a Probiotic-Focused Diet on Cardiovascular Risk Markers in Obese Individuals: A 12-Week Intervention Study

Bercem Aycicek*, Suleyman Akkaya**, Mesude Uzun**, Umit Cavdar**, Halit Diri**

*Istinye University Faculty of Medicine, Department of Endocrinology and Metabolism Diseases, Istanbul, Türkiye

**University of Health Sciences Türkiye, Diyarbakir Gazi Yasargil Training and Research Hospital, Department of Cardiology, Diyarbakir, Türkiye

Abstract

Aim: Obesity is a key risk factor for cardiovascular diseases, contributing to endothelial dysfunction and atherosclerosis. Pericardial fat, a visceral fat depot surrounding the heart, has emerged as an important cardiovascular risk factor. Its proximity to coronary vessels and its role in inflammatory processes suggest its potential impact on vascular health. This study investigates the effects of a probiotic-focused diet on cardiovascular risk markers, including carotid intima-media thickness (CIMT), flow-mediated dilation (FMD), and pericardial fat thickness in obese individuals.

Methods: This was a 12-week prospective dietary intervention study designed to evaluate the effects of a probiotic-rich diet in 100 obese participants. The participants followed a probiotic-focused diet, and cardiovascular risk markers-including CIMT, FMD, and pericardial fat thickness-were measured before and after the intervention. Metabolic markers such as body weight, insulin levels, and high-sensitivity C-reactive protein (hsCRP) were also evaluated.

Results: The probiotic group showed significant improvements in body weight, body mass index, insulin sensitivity, and hsCRP levels. However, no significant differences were observed in CIMT, FMD, or pericardial fat thickness between the two groups. Pericardial fat thickness showed a slight reduction in the probiotic group, but this change was not statistically significant.

Conclusion: A probiotic-rich diet significantly improves metabolic health but does not result in significant changes in cardiovascular markers such as CIMT, FMD, or pericardial fat thickness in obese individuals over a 12-week period. Further research with longer intervention periods is necessary to explore the impact of probiotics on vascular health and pericardial fat accumulation in obesity.

Keywords: Obesity, probiotics, cardiovascular diseases, inflammation, endothelium, vascular

Introduction

Obesity is a major global health problem and a leading risk factor for cardiovascular disease (CVD). It is closely associated with subclinical vascular changes, including increased carotid intima-media thickness (CIMT) and impaired endothelial function, both of which contribute to the development of atherosclerosis. Carotid intima-media thickness reflects arterial wall thickening and has been recognized as a surrogate marker for atherosclerotic burden and cardiovascular events such as myocardial infarction and stroke (1). Flow-mediated dilation (FMD), on the other hand, is a non-invasive assessment of endothelial function and an early indicator of vascular health (2,3).

Pericardial fat, a visceral fat depot surrounding the heart, has emerged as a significant cardiovascular risk factor due to its inflammatory activity and anatomical

Corresponding Author: Bercem Aycicek, Istinye University Faculty of Medicine, Department of Endocrinology and Metabolism Diseases, Istanbul, Türkiye
E-mail: bercemay@gmail.com **ORCID:** orcid.org/0000-0001-7085-5846
Received: 27.01.2025 **Accepted:** 11.07.2025 **Epub:** 12.08.2025 **Publication Date:** 29.08.2025

Cite this article as: Aycicek B, Akkaya S, Uzun M, Cavdar U, Diri H. The effects of a probiotic-focused diet on cardiovascular risk markers in obese individuals: a 12-week intervention study. Med Bull Haseki. 2025;63(3):159-164

proximity to coronary vessels (4). Several studies have reported a strong association between pericardial fat and endothelial dysfunction, particularly in obese individuals (5). Meanwhile, probiotics have gained attention for their ability to modulate the gut microbiota and reduce systemic inflammation. They have shown promise in improving insulin resistance, lipid profiles, and pro-inflammatory markers-key contributors to cardiovascular risk (6-8).

We hypothesized that a probiotic-focused diet would improve metabolic parameters and potentially exert favorable effects on vascular health in obese individuals. The aim of this study was to evaluate the effects of a 12-week probiotic-rich dietary intervention on metabolic markers, endothelial function (assessed by FMD), and subclinical atherosclerosis (assessed by, CIMT) in obese individuals. This approach may contribute to cardiovascular prevention strategies by offering a non-pharmacological, probiotic-targeted dietary tool to improve cardiometabolic health in high-risk populations.

Materials and Methods

Compliance with Ethical Standards

This study was conducted in accordance with the Declaration of Helsinki. Ethical approval was obtained from the Clinical Research Ethics Committee of the University of Health Sciences Türkiye, Diyarbakir Gazi Yasargil Training and Research Hospital (approval no.: 154, date: 05.10.2018). All participants provided written informed consent prior to their inclusion in the study.

Study Design

This 12-week study was conducted between January and June 2019 at University of Health Sciences Türkiye, Diyarbakir Gazi Yasargil Training and Research Hospital. The participants followed a probiotic-focused diet, and cardiovascular risk markers-including CIMT, FMD, and pericardial fat thickness were measured before and after the intervention. The primary aim was to assess the impact of a probiotic-rich diet on these markers of metabolic and vascular health. Previous research has shown that probiotics can influence metabolic parameters such as insulin sensitivity and inflammation (1), and that pericardial fat, in particular, is a strong predictor of cardiovascular risk, particularly in obese individuals (2). By measuring these markers, this study aimed to explore how dietary interventions might improve cardiovascular health in obese populations.

Inclusion Criteria

Participants aged 18-65 years with a body mass index (BMI) of ≥30 kg/m² were included. None of the participants had a history of CVD, diabetes, or other significant medical conditions.

Dietary Intervention

Participants followed a diet enriched with probiotic foods aimed at improving gut microbiota health and overall metabolic function. The diet included regular consumption of foods such as yogurt, kefir, and fermented vegetables (e.g., sauerkraut, kimchi). Participants were instructed to incorporate these foods into their daily meals, with portion sizes and frequency tailored to their caloric and nutritional requirements as determined by a dietitian.

To ensure adherence and provide personalized guidance, participants met individually with a dietitian three times per week. During these sessions, the dietitian monitored dietary compliance, provided tailored advice, and addressed any challenges faced by the participants. Dietary logs were also maintained by participants to document daily food intake, which was reviewed during the consultations.

Assessment of Carotid Intima-Media Thickness

Assessment of CIMT was measured using high-resolution B-mode ultrasonography [e.g., (state the device, such as GE Vivid E9, Philips iE33)]. Measurements were taken on the far wall of the distal 1 cm of the common carotid artery, approximately 1-2 cm proximal to the bifurcation (9). Three measurements were obtained bilaterally, and the mean CIMT value was calculated for analysis. All measurements were performed by an experienced physician blinded to the study groups and intervention status to ensure objectivity and reduce bias. The intra-observer variability was assessed and found to be within acceptable limits (<5%).

Assessment of Flow-Mediated Dilation

Assessment of FMD, an established marker of endothelial function, was assessed using a standardized non-invasive ultrasound technique. Participants were instructed to fast and refrain from consuming caffeine or alcohol or engaging in physical activity for at least 12 hours prior to the procedure. Measurements were conducted in a temperature-controlled room after 10 minutes of rest in the supine position. Reactive hyperemia was induced by inflating a forearm cuff to 50 mmHg above systolic blood pressure for 5 minutes, followed by sudden deflation. The brachial artery diameter was measured at baseline and during hyperemia. FMD was calculated as the percentage increase in arterial diameter from baseline (9). All measurements were performed by the same experienced physician, who was blinded to the study groups and intervention details.

Assessment of Pericardial Fat Thickness

Pericardial fat thickness, a key marker of visceral adiposity surrounding the heart, was measured using transthoracic echocardiography. Parasternal long-axis

and short-axis views were obtained, and pericardial fat thickness was defined as the echo-lucent space between the outer wall of the myocardium and the visceral layer of the pericardium. Measurements were taken perpendicularly to the right ventricular free wall at end-diastole. Three consecutive cardiac cycles were averaged for the final value. An experienced and blinded physician performed all echocardiographic evaluations, ensuring consistency and minimizing measurement bias.

Quality Assurance

To ensure high data quality, all imaging devices were calibrated before each session. The physician conducting the measurements underwent inter-rater reliability validation at the study's outset. Blinded duplicate measurements were performed in a random subset of 10% of participants to assess reproducibility, yielding intraclass correlation coefficients >0.9 for all parameters.

Metabolic Markers

Body weight, BMI, insulin levels, and high-sensitivity C-reactive protein were also measured. Homeostasis model assessment of insulin resistance (HOMA-IR) was calculated as a measure of insulin resistance.

Statistical Analysis

Continuous variables were assessed for normality using the (e.g., Shapiro-Wilk test or Kolmogorov-Smirnov test). Normally distributed variables were presented as mean ± standard deviation, while non-normally distributed variables were summarized as median and interquartile range. For within-group comparisons, paired t-tests were applied to normally distributed data, and the Wilcoxon signed-rank test was used for non-parametric data.

Effect sizes were calculated to assess the magnitude of observed changes, reported as Cohen's d for paired t-tests or matched rank biserial correlation for non-parametric tests. Missing data were managed using (state method, e.g., pairwise deletion or imputation), and sensitivity analyses were conducted where appropriate. Continuous variables were assessed for normality using the Shapiro-Wilk test.

The statistical significance level was set at $p<0.05$. All analyses were performed using (state statistical software, e.g., Statistical Package for the Social Sciences version 25.0, R version 4.3.1). Confidence intervals of 95% were provided for key outcomes.

Results

The baseline characteristics of both groups were similar, with no significant differences observed in terms of age, gender, or baseline cardiovascular risk markers. Table 1 present all the data for both baseline and post-intervention measurements.

Participants in the probiotic group experienced a significant reduction in body weight (-6.57 kg, $p<0.001$) and BMI (-1.63 kg/m², $p<0.01$). Insulin levels decreased by 3.42 μU/mL ($p<0.01$), and HOMA-IR improved by 1.23 ($p<0.01$), while the control group showed no significant changes in these metabolic parameters.

Regarding cardiovascular risk markers, although changes in CIMT and pericardial fat thickness were observed in both groups, no significant differences were found between the probiotic and control groups (CIMT: -0.02 mm, $p=0.35$, pericardial fat: -0.01 cm, $p=0.26$). Flow-mediated dilation showed a slight improvement in the probiotic group (+0.18%), but this change was not statistically significant ($p=0.12$).

Discussion

The findings of this study suggest that a probiotic-focused diet can significantly improve body composition and insulin sensitivity in obese individuals. These results are consistent with previous studies that have demonstrated the benefits of probiotics in reducing metabolic dysfunctions such as insulin resistance and inflammation (10).

Recent clinical trials have also shown that probiotics may modulate gut microbiota composition, promote short-chain fatty acids production, and reduce systemic inflammation, all of which contribute to improved insulin signaling and energy metabolism (11-13).

Probiotics have also been shown to improve lipid profiles and reduce markers of systemic inflammation, which are known to contribute to cardiovascular risk (14). However, our study did not observe significant improvements in CIMT or FMD, which suggests that longer intervention periods or more intensive dietary changes may be necessary to elicit meaningful changes in vascular function. This may be due to the fact that vascular remodeling and endothelial repair are complex processes, typically requiring sustained anti-inflammatory stimuli and structural adaptation, which may not be fully achieved within a 12-week timeframe (14).

One aspect that could have influenced our findings is the potential role of pericardial fat, which is an important but often overlooked marker of cardiovascular risk in obesity. Pericardial fat is known to contribute to systemic inflammation and endothelial dysfunction, both of which are critical factors in the development of atherosclerosis and CVD (4). Previous studies have highlighted that pericardial fat is a strong predictor of coronary artery disease and other cardiovascular events, particularly in individuals with visceral obesity (5). Recent studies have demonstrated that pericardial fat volume independently predicts coronary artery calcification and impaired FMD, even after adjusting for BMI and visceral adiposity (1,15,16).

Table 1. Baseline and post-intervention analysis of cardiovascular and metabolic variables

Variable	Group 1 (Mean ± SD) (Baseline)	Group 2 (Mean ± SD) (Post-intervention)	*p-value
Body weight	83.96±15.41	77.39±15.61	p<0.001
Height	1.65±0.08	-	-
BMI (kg/m²)	30.78±5.07	29.15±4.91	p<0.001
FBG (mg/dL)	94.37±11.78	96.90±8.81	0.535
PPBG (mg/dL)	104.69±24.53	96.60±17.43	0.208
HBA1C (%)	5.36±0.29	5.37±0.33	0.465
ALT (U/L)	16.50±14.84	17.00±16.07	0.972
Fibrinogen (mg/dL)	285.50±69.14	271.50±70.02	0.964
Homocystein (µmol/L)	695±476	654±375	0.757
Total cholesterol (mg/dL)	99.00±24.36	96.60±17.43	0.401
Triglycerides (mg/dL)	130.00±55.37	127.00±45.91	0.542
HDL (mg/dL)	45.00±12.15	47.00±9.30	0.280
LDL (mg/dL)	123.00±38.55	120.00±28.44	0.889
VLDL (mg/dL)	30.00±6.12	30.00±7.09	0.919
Pericardial fat thickness (cm)	0.67±0.17	0.63±0.18	0.26
Vitamin D (ng/mL)	16.48±8.23	15.94±5.68	0.189
C-reactive protein (mg/dL)	2.58±0.78	2.55±0.72	0.543
Cinc (µg/dL)	88.59±10.35	94.70±7.59	0.041
Magnesium (mg/dL)	1.95±0.20	1.93±0.10	0.280
C-peptide (ng/mL)	2.64±1.01	2.39±0.81	0.068
Insulin (µU/mL)	13.32±8.67	10.59±6.67	0.052
CIMT (mm)	0.07±0.09	0.0546±0.01527	0.288
FMD1 (%)	0.37±0.07	0.3829±0.07668	0.525
FMD2 (%)	0.40±0.08	0.4071±0.08956	0.816
FMDs (%)	-72.35±22.35	-69.9427±25.49738	0.646

*Statistical comparisons were performed using paired t-test and Wilcoxon signed-rank test as appropriate
BMI: Body mass index, FBG: Fasting blood glucose, PPBG: Postprandial blood glucose, HBA1C: Hemoglobin A1C, ALT: Alanine transaminas, HDL: High-density lipoprotein, LDL: Low-density lipoprotein, VLDL: Very low density lipoprotein, FMD: Flow-mediated dilation, CIMT: Carotid intima-media thickness, SD: Standard deviation

Given that our study did not measure pericardial fat, it is possible that incorporating this marker could have revealed additional insights into the diet's impact on cardiovascular risk. Moreover, longer intervention periods and more intensive dietary changes may be required to see significant changes in vascular health markers like CIMT and FMD, as these are often influenced by multiple factors such as systemic inflammation, lipid metabolism, and genetic predisposition (17,18).

Overall, while probiotic diets can improve metabolic health, their direct effects on vascular markers, including those influenced by pericardial fat, remain inconclusive. Further research with longer durations and larger sample sizes is needed to determine the long-term benefits of such diets on CVD risk.

Study Limitations

This study suggests that a probiotic-focused diet can improve metabolic health and insulin sensitivity in obese individuals. However, the effects on cardiovascular markers, such as CIMT and FMD, were not significant. Small sample size and short study duration are key limitations. The absence of a control group limits the ability to definitively attribute observed changes to the probiotic intervention, as potential confounding factors or regression to the mean cannot be excluded. Future studies should address these issues by including larger cohorts, longer interventions, and evaluating pericardial fat and gut microbiota to provide a clearer understanding of how probiotics influence cardiovascular risk. Despite these limitations, the study's strengths include its prospective

design, objective vascular measurements, and close dietary supervision.

Conclusion

A probiotic-focused diet led to significant improvements in body composition and insulin sensitivity in obese individuals. Although no significant changes were observed in CIMT, FMD, or pericardial fat, the metabolic improvements observed in this study suggest that probiotic-rich diets may help reduce obesity-related cardiovascular risk in the long term. Future studies should explore the long-term effects of probiotic diets on cardiovascular health, with a focus on endothelial function and arterial stiffness.

Ethics

Ethics Committee Approval: This study was conducted in accordance with the Declaration of Helsinki. Ethical approval was obtained from the Clinical Research Ethics Committee of the University of Health Sciences Türkiye, Diyarbakir Gazi Yasargil Training and Research Hospital (approval no.: 154, date: 05.10.2018).

Informed Consent: Informed consent was obtained from all participants prior to their enrollment in the study.

Acknowledgements: We would like to thank all participants for their contribution to this study.

Footnotes

Authorship Contributions

Concept: B.A., Design: B.A., Data Collection or Processing: S.A., M.U., U.C., H.D., Analysis or Interpretation: S.A., M.U., H.D., Literature Search: B.A., M.U., U.C., Writing: B.A., U.C.

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

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

References

1. Willeit P, Tschiderer L, Allara E, et al. Carotid intima-media thickness progression as surrogate marker for cardiovascular risk: meta-analysis of 119 clinical trials involving 100,667 patients. Circulation. 2020;142:621-42.
2. Mućka S, Miodońska M, Jakubiak GK, Starzak M, Cieślar G, Stanek A. Endothelial function assessment by flow-mediated dilation method: a valuable tool in the evaluation of the cardiovascular system. Int J Environ Res Public Health. 2022;19:11242.
3. Arrebola-Moreno AL, Laclaustra M, Kaski JC. Noninvasive assessment of endothelial function in clinical practice. Rev Esp Cardiol (Engl Ed). 2012;65:80-90.
4. Li C, Liu X, Adhikari BK, et al. The role of epicardial adipose tissue dysfunction in cardiovascular diseases: an overview of pathophysiology, evaluation, and management. Front Endocrinol (Lausanne). 2023;14:1167952.
5. Mookadam F, Goel R, Alharthi MS, Jiamsripong P, Cha S. Epicardial fat and its association with cardiovascular risk: a cross-sectional observational study. Heart Views. 2010;11:103-8.
6. Peña-Durán E, García-Galindo JJ, López-Murillo LD, et al. Microbiota and inflammatory markers: a review of their interplay, clinical implications, and metabolic disorders. Int J Mol Sci. 2025;26:1773
7. Tripolt NJ, Leber B, Blattl D, et al. The effects of probiotics on inflammation, endothelial dysfunction, and atherosclerosis progression: a mechanistic overview. Heart Lung Circ. 2022;31:e26-e33.
8. Pires L, González-Paramás AM, Heleno SA, Calhelha RC. The role of gut microbiota in the etiopathogenesis of multiple chronic diseases. Antibiotics (Basel). 2024;13:392.
9. Sharma K, Blaha MJ, Blumenthal RS, Musunuru K. Clinical and research applications of carotid intima-media thickness. Am J Cardiol. 2009;103:1316-20.
10. Ding LN, Ding WY, Ning J, Wang Y, Yan Y, Wang ZB. Effects of probiotic supplementation on inflammatory markers and glucose homeostasis in adults with type 2 diabetes mellitus: a systematic review and meta-analysis. Front Pharmacol. 2021;12:770861.
11. Tabrizi R, Ostadmohammadi V, Akbari M, et al. The effects of probiotic supplementation on clinical symptom, weight loss, glycemic control, lipid and hormonal profiles, biomarkers of inflammation, and oxidative stress in women with polycystic ovary syndrome: a systematic review and meta-analysis of randomized controlled trials. Probiotics Antimicrob Proteins. 2022;14:1-14.
12. Cao N, Zhao F, Kwok LY, Wang H, Sun Z. Impact of probiotics on weight loss, glucose and lipid metabolism in overweight or obese women: a meta-analysis of randomized controlled trials. Curr Res Food Sci. 2024;9:100810.
13. Salles BIM, Cioffi D, Ferreira SRG. Probiotics supplementation and insulin resistance: a systematic review. Diabetol Metab Syndr. 2020;12:98.
14. Wang Q, Wang Z, Pang B, et al. Probiotics for the improvement of metabolic profiles in patients with metabolic-associated fatty liver disease: a systematic review and meta-analysis of randomized controlled trials. Front Endocrinol (Lausanne). 2022;13:1014670.
15. Veglia F, Malagoni AM, Amato M, et al. Enhancing cardiovascular risk prediction with a simplified carotid IMT protocol: evidence from the IMPROVE study. Biomedicines. 2025;13:584.
16. Nafakhi H, Elwali HQ, Al-Sharea KMK, Al-Mosawi AA. Relationship of cardiovascular risk factors, pericardial fat, and

carotid thickness with coronary plaque type in patients with diabetes mellitus. J Diabetes Metab Disord. 2023;22:713-9.

17. Kim SH, Lee SJ, Kang ES, et al. Effects of lifestyle modification on metabolic parameters and carotid intima-media thickness in patients with type 2 diabetes mellitus. Metabolism. 2006;55:1053-9.

18. de Vries P, Hasbani N, Heath A, et al. A multi-trait genome-wide association study of coronary artery disease and subclinical atherosclerosis traits. Res Sq [Preprint]. 2025:rs.3.rs-6456056.

Original Article

DOI: 10.4274/haseki.galenos.2025.58076
Med Bull Haseki 2025;63(3):165-170

Long-term Follow-up of Valvular Involvement in Children with Acute Rheumatic Fever Carditis: 15-year Results

Muhammet Hamza Halil Toprak*, Abdulhakim Gunes**, Abdusselam Genc***, Fahrettin Uysal****, Ozlem Mehtap Bostan****

*University of Health Sciences Türkiye, Basaksehir Cam ve Sakura City Hospital, Clinic of Pediatric Cardiology, Istanbul, Türkiye

**Siirt Hayat Hospital, Clinic of Pediatrics, Siirt, Türkiye

***Bursa City Hospital, Clinic of Pediatric Cardiology, Bursa, Türkiye

****Bursa Uludag University Faculty of Medicine, Department of Pediatric Cardiology, Bursa, Türkiye

Abstract

Aim: Acute rheumatic fever (ARF) is a multisystemic inflammatory disease and remains an important health problem in low- and middle-income countries. The most important and potentially permanent adverse effect of the disease is cardiac involvement. This study aims to investigate the course of valvular lesions in ARF carditis and the factors that may affect their course during a long-term follow-up.

Methods: In this cross-sectional study, which is a type of observational study, the records of children diagnosed with ARF in our hospital between August 2003 and December 2018 were retrospectively reviewed. Valvular insufficiency was examined on echocardiography and compared with the previous findings. The changes in the valvular lesions and the factors that could affect them were evaluated. The results were analyzed statistically.

Results: The study included 90 patients with ARF who had been followed up for over a year. Of our patients, 37 (41%) were male and 53 (59%) were female. The mean age of our patients was 16.5±3.2 years. The mean age at diagnosis was 11 (4.2-18 years). The mean follow-up period was 67 (14-184) months. Carditis was present in 86 (95.6%) patients. Echocardiographically, mitral regurgitation (MR) was present in 83 (96.5%) patients and aortic regurgitation (AR) was present in 56 (65%) patients. 54 (62.7%) patients had both MR and AR. During the follow-up, MR persisted or progressed in 19 (22.9%) patients, regressed in 39 (47%) patients, and completely recovered in 25 (30.1%) patients. Aortic regurgitation persisted or progressed in 13 (23.2%) patients, regressed in 4 (7.1%) patients, and completely recovered in 39 (69.7%) patients.

Conclusion: The recovery rate of aortic valve lesions was found to be much higher than that of mitral valve lesions. Initially, The high rate of mild involvement in aortic valve lesions maybe effective in this context. Apart from this, the valve lesion types that have a negative course and the factors affecting them still need to be studied further. Benzathine penicillin prophylaxis remains the most effective method of preventing recurrence.

Keywords: Rheumatic fever, rheumatic heart disease, heart valve diseases

Introduction

Acute rheumatic fever (ARF) is a systemic autoimmune syndrome involving a combination of the joints, heart, brain, skin, and and subcutaneous tissues that develops following exposure to *Group A Streptococcus* [*Streptococcus pyogenes (GAS)*] pharyngitis or impetigo (1). It remains an important health problem in low- and middle-income countries. The most important and potentially permanent complication of the disease is cardiac involvement. The condition often occurs in valvular regurgitation of the left heart valves.

At disease onset, valvular insufficiency is the predominant finding, while in the chronic form, valve

Corresponding Author: Muhammet Hamza Halil Toprak, MD, University of Health Sciences Türkiye, Basaksehir Cam ve Sakura City Hospital, Clinic of Pediatric Cardiology, İstanbul, Türkiye

E-mail: muhammedhamzatoprak@hotmail.com **ORCID:** orcid.org/0000-0003-4579-963X

Received: 02.01.2025 **Accepted:** 23.06.2025 **Epub:** 12.08.2025 **Publication Date:** 29.08.2025

Cite this article as: Toprak MHH, Gunes A, Genc A, Uysal F, Bostan OM. Long-term follow-up of valvular involvement in children with acute rheumatic fever carditis: 15-year results. Med Bull Haseki. 2025;63(3):165-170

stenosis can also be seen. Some of the valvular lesions improve, while others progressively worsen over time and may even require surgical intervention (2).

The aim of this study was to investigate the course of valvular lesions and the factors that may affect it during a long-term follow-up.

Materials and Methods

Compliance with Ethical Standards

The study protocol was approved by the Uludag University Faculty of Medicine Clinical Research Ethics Committee (approval no.: 2023-9/16, date: 26.04.2023). Written consent was obtained from all participants, and the study complied with the Declaration of Helsinki.

Study Design

In this retrospective study the records of children diagnosed with ARF in our hospital between August 2003 and December 2018 were retrospectively reviewed. Patients with a follow-up period of less than 1 year were excluded from the study. Acute rheumatic fever diagnoses for all the patients were made in our hospital and we had records corresponding to their initial episode. All diagnoses were made according to the modified Jones criteria, and the revised criteria were used for patients diagnosed after 2015 (3,4). Demographic data and clinical, laboratory, and echocardiographic findings of our patients were all present and examined from the patient files. Clinical procedures performed on the patients all included detailed cardiovascular examinations, and benzathine penicillin prophylaxis regularity was thoroughly checked. Electrocardiography and echocardiography were performed for each patient at each follow-up. Echocardiographic reports were reviewed by the same physician, Muhammet Hamza Halil Toprak.

When evaluating valvular insufficiency, trivial insufficiencies that did not involve changes in the valve morphology (thickening, limitation of chordate movements) weren't considered pathological. Pansystolic mitral regurgitation (MR) and pandiastolic aortic regurgitation (AR) with a peak velocity greater than 3 m/ss were considered significant. Mitral regurgitation was classified into severity groups as mild, moderate, and severe based on the ratio of maximum insufficiency jet area to atrium area. Those with a rate of <20% were considered mild, those between 20% and 40% were considered moderate, and those >40% were considered severely insufficient. Aortic insufficiency was classified as mild, moderate, or severe based on the parasternal long axis insufficiency ratio, defined as the jet width divided by the left ventricular outflow tract width. Disabilities with a rate of <25% were considered mild, those between 25% and 45% were considered moderate, and those >45% were considered severe. At each follow-up, valvular insufficiency was examined on echocardiography and compared with the previous findings.

Statistical Analysis

Statistical evaluation of the data was performed with Statistical Package for the Social Sciences 16 for Windows. Categorical data is divided into frequency and percentage (n, %). Continuous variable data were presented as mean ± standard deviation or median (min-max). The differences between categorical variable frequencies were investigated by the chi-square test. Normality analysis of the variables was performed with the Shapiro-Wilk test. Continuous variables that did not show normal distribution were compared with the Mann-Whitney U test. The significance level was set at α=0.05 ($p<0.05$).

Results

A total of 90 patients who were diagnosed with ARF in our hospital and had been followed up for more than a year were included in the study. Among the patients, 37 (41%) were male and 53 (59%) were female. The mean age of our patients was 16.5±3.2 years. The mean age at diagnosis was 11 (4.2-18) years. The diagnosis season was spring in 26 (28.9%) patients, summer in 23 (25.6%), autumn in 12 (13.3%), and winter in 29 (32.2%) patients. Only 6 (6.7%) patients had a family history of ARF.

The mean follow-up period was 67 (14-184) months. Carditis was present in 86 (95.6%) patients. The frequencies of other major findings are given in (Table 1). Echocardiographically, MR was present in 83 (96.5%) patients, and AR in 5656 (65%) patients. Fifty-four (62.7%) patients had both MR and AR. A total of 136 valve involvements was determined in 86 patients. The distribution and severity of valvular involvement at the time of diagnosis are given in (Figure 1). During the follow-up, MR persisted or progressed in 19 (22.9%) patients, regressed in 39 (47%) patients, and completely recovered in 25 (30.1%) patients. Aortic regurgitation persisted or progressed in 13 (23.2%) patients, regressed in 4 (7.1%), and completely recovered in 39 (69.7%). The progression of valvular lesions is given in (Figure 2).

The rate of complete recovery of aortic valve lesions was higher than that of mitral valve lesions, and the difference was statistically significant (69.7%, 30.1%, $p<0.01$).

Table 1. The frequencies of the major findings

Finding	n	%
Carditis	86	95.6
Arthritis	52	59.1
Corea	18	20.2
Subcutaneous nodules	1	1.1
Erithema marginatum	0	0

When the improvement and deterioration rates of mitral and aortic valve lesions were compared based on gender, no significant difference was detected for either of the valve lesions (p=0.68, p=0.12). When the improvement and deterioration rates of mitral and aortic valve lesions were compared based on age, no significant difference was detected for either of the valve lesions (p=0.89, p=0.98).

No significant difference was detected for either of the mitral and aortic valve lesions when the improvement and deterioration rates were compared according to age at the time of diagnosis (p=0.18, p=0.65).

Similarly, there was no statistically significant difference for the improvement and deterioration rates of mitral and aortic valve lesions based on the presence of chorea foreither of the lesion types (p=0.90, p=0.14).

Similarly, when the improvement and deterioration rates of mitral and aortic valve lesions were compared according to antistreptolysin O (ASO) levels at the time of diagnosis, no statistically significant difference could be detected for any of the valve lesions (p=0.90, p=0.14). The comparison of the factors affecting the course of aortic and mitral valve lesions ispresented in (Tables 2,3)

Recurrence was observed in 7 patients. Of these, 2 were seen in patients compliant with benzathine penicillin prophylaxis, and 5 were seen in patients who were non-compliant with benzathine penicillin prophylaxis. Two patients underwent surgical valve replacement due to cardiac failure. One patient underwent mitral valve replacement (MVR), and the other patient underwent simultaneous aortic and MVR.

Discussion

Acute rheumatic fever is a non-suppurative disease that occurs because of an autoimmune response to *GAS* pharyngitis. The disease mainly affects the heart, joints, and nervous system. It is more commonly encountered in children and young adults. Non-cardiac involvement of the disease is transient, while the most important permanent damage occurs as a result of cardiac involvement (5). For this reason, research on the understanding and treatment of heart lesions is important. Previously, the criteria for detecting carditis were based on physical examination for which the sensitivity was low (3). In recent years, with the widespread use of echocardiography to detecting carditis, it has been understood that cardiac involvement is actually

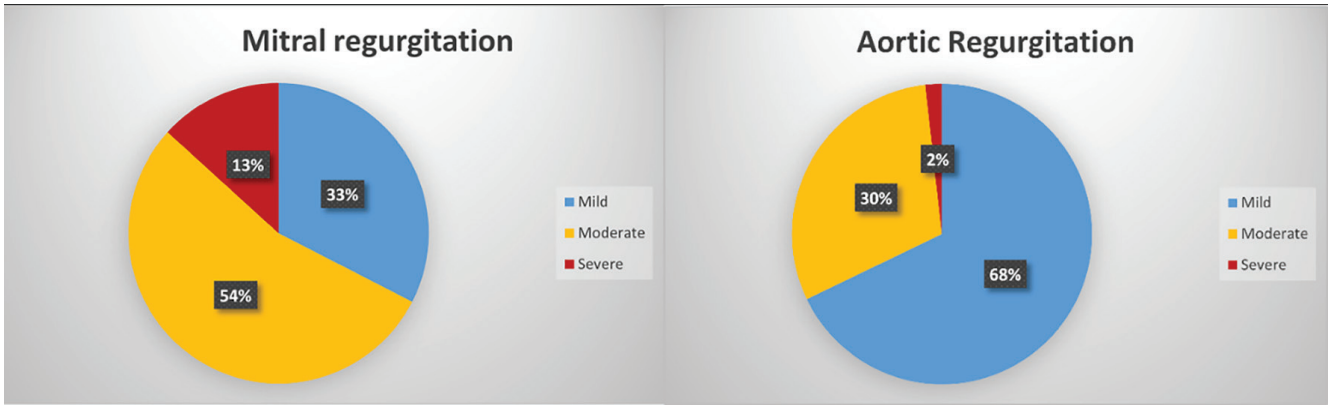

Figure 1. The distribution and severity of valvular involvement at the time of diagnosis

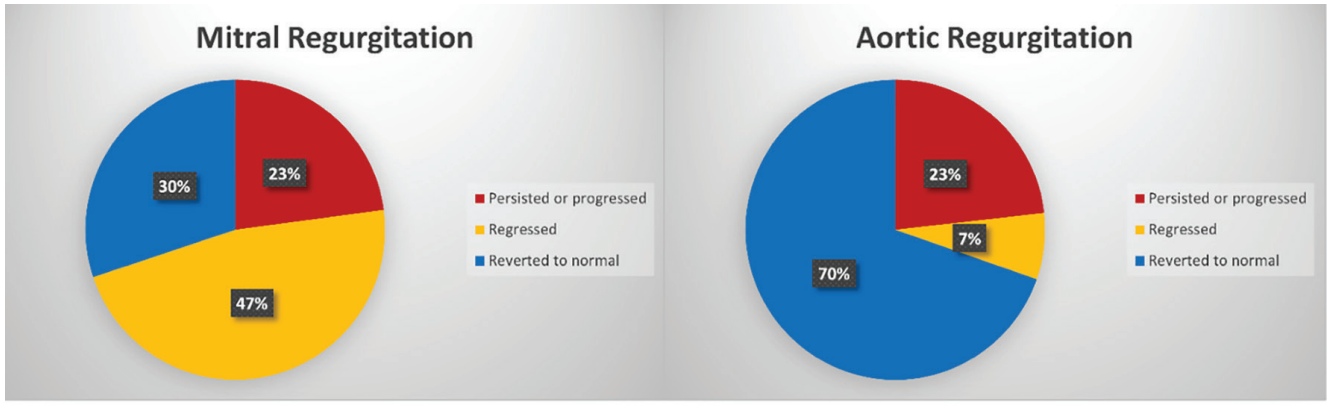

Figure 2. The progression of valvular lesions

Table 2. The comparision of the factors that affecting the course of mitral valve lesions				
Variable	**Cathegory**	**Persisted or progressed**	**Regressed**	**P-value**
Gender	Male	7	26	0.68
	Female	12	38	
Age		199.47±36.53	200.85±37.51	0.89
Age at diagnosis		141.41±23.20	129.32±35.16	0.18
ASO (IU/mL)		936	891	0.67
Corea	Positive	4	14	0.90
	Negative	14	50	
ASO: Antistreptolysin O				

Table 3. The comparision of the factors that affecting the course of aortic valve lesions				
Variable	**Cathegory**	**Persisted or progressed**	**Regressed**	**P-value**
Gender	Male	8	18	0.12
	Female	5	25	
Age		203.08±45.58	203.32±33.5	0.98
Age at diagnosis		130.41±30.5	135.58±35.7	0.65
ASO (IU/mL)		920.5	966	0.74
Corea	Positive	3	4	0.14
	Negative	10	39	
ASO: Antistreptolysin O				

more significant than previously believed. For this reason, echocardiographic findings have taken their place in the revised diagnostic criteria and are explained in detail in the literature (4).

Acute rheumatic carditis is a chronic process that begins in childhood and young adulthood and usually continues into adulthood and is still one of the most important causes of acquired heart valve diseases in low- and middle-income countries (6). The worsening of valvular lesions and the factors affecting it are of scientific interest and represent the most important points to be explained in preventing disease sequelae. Our study aimed to monitor the change in valvular involvement over time and to investigate the factors that may affect it.

In the study conducted by Meira et al., (7) only 9 patients (6.2%) in the follow-up of 146 patients with carditis had normal echocardiography findings. In that study, the recovery rates of the aortic and mitral valves were not given separately. In our study, the rate of complete recovery of aortic valve lesions was higher than that of mitral valve lesions, and the difference was statistically significant. (69.7%, 30.1%, p<0.01). It was initially thought that the high recovery rate in aortic valve lesions may be related to the high incidence of mild involvement.

Whether gender influences the healing of valvular lesions has been investigated in many studies, but no significant difference has been found (8,9). In our study, when the improvement and deterioration rates of mitral and aortic valve lesions were compared according to gender, there was no statistically significant difference for either of the valve lesions (p=0.68, p=0.12).

In the study conducted by Beaton et al. (8) on 60 patients with a follow-up period of 2 years, younger age was defined as a risk factor for the persistence of valvular lesions. In our study, which had a much longer follow-up period, there was no statistically significant difference between the valve lesions when the improvement and deterioration rates of mitral and aortic valve lesions were compared based on age (p=0.89, p=0.98).

In a prospective study conducted by Zühlke et al. (9) with a follow-up period of 5 years, no relationship was found between age at diagnosis and progression of valvular lesions. The results of our study were in line with their findings.

Contrary to the study (8), which showed that the recovery rate of valve insufficiency was lower in patients with higher ASO values at the time of initial diagnosis, our study showedshowed the baseline ASO values for both mitral and aortic valves did not affect the rate of improvement or deterioration (p=0.67, p=0.74). The follow-up period in that study was 2 years, while the mean follow-up period in our study was 67 months (approximately 5 and a half years).

Benzathine penicillin prophylaxis is recognized as the most effective treatment in preventing recurrences of ARF. It is known that the recurrence rate is higher in patients who do not comply with treatment (10-12). In our study, 5 out of 7 patients with recurrence did not comply with regular prophylaxis. There are studies showing that valvular lesions are aggravated and that patients with recurrence experience greater worsening (13). On the contrary, in our study, all 7 patients with recurrence were found to have decreased valve involvement. We could not find any reason to explain this.

Surgical valve replacement emerges as a treatment option for valve insufficiency in the presence of medically uncontrolled cardiac insufficiency. The underlying pathology may be acute chordae rupture or significant valve regurgitation that does not respond to steroid therapy (14). Two of our patients underwent surgical valve replacement. In one patient, MVR and aortic valve replacement were performed together, and in the other patient, only MVR was performed. Neither of these patients had recurrence.

Study Limitations

This study has several limitations. The study was conducted at a single center, and the relatively small number of patients included constitutes the main limitation. Despite these limitations, we believe that this study will make a significant contribution to the literature due to its long follow-up period.

Conclusion

Most of the valve lesions that occur due to ARF carditis either regress or completely resolve in the long term. Our study had a long follow-up period of 15 years, and the recovery rate of aortic valve lesions was found to be much higher than that of mitral valve lesions. The high rate of mild involvement in aortic valve lesions initially may be effective in this. Apart from this, the valve lesion types that have a negative course, and the factors affecting them, still need to be studied further. Benzathine penicillin prophylaxis remains the most effective method of preventing recurrence.

Ethics

Ethics Committee Approval: The study protocol was approved by the Uludag University Faculty of Medicine Clinical Research Ethics Committee (approval no.: 2023-9/16, date: 26.04.2023).

Informed Consent: Written consent was obtained from all participants, and the study complied with the Declaration of Helsinki.

Footnotes

Authorship Contributions

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

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

Financial Disclosure: This study received no financial support

References

1. Hirani K, Rwebembera J, Webb R, et al. Acute rheumatic fever. Lancet. 2025;405:2164-78.
2. Reményi B, Wilson N, Steer A, et al. World Heart Federation criteria for echocardiographic diagnosis of rheumatic heart disease an evidence-based guideline. Nat Rev Cardiol. 2012;9:297-309.
3. Guidelines for the diagnosis of rheumatic fever. Jones Criteria,1992 update. Special Writing Group of the Committee on Rheumatic Fever, Endocarditis, and Kawasaki Disease of the Council on Cardiovascular Disease in the Young of the American Heart Association. JAMA. 1992;268:2069-73.
4. Gewitz MH, Baltimore RS, Tani LY, et al. American Heart Association Committee on Rheumatic Fever, Endocarditis and Kawasaki Disease of the Council on Cardiovascular Disease in the Young. Revision of the Jones criteria for the diagnosis of the rheumatic fever in the era of Doppler echocardiography: a scientific statement of the American Heart Association. Circulation. 2015;131:1806-18.
5. Carapetis JR, Beaton A, Cunningham MW, et al. Acute rheumatic fever and rheumatic heart disease. Nat Rev Dis Primers. 2016;2:15084.
6. Paar JA, Berrios NM, Rose JD, et al. Prevalence of rheumatic heart disease in children and young adults in Nicaragua. Am J Cardiol. 2010;105:1809-14.
7. Meira ZM, Goulart EM, Colosimo EA, Mota CC. Long term follow-up of rheumatic fever and predictors of severe rheumatic valvar disease in Brazilian children and adolescents. Heart. 2005;91:1019-22.
8. Beaton A, Okello E, Aliku T, et al. Latent rheumatic heart disease: outcomes 2 years after echocardiographic detection. Pediatr Cardiol. 2014;35:1259-67.
9. Zühlke L , Engel ME, Lemmer CE, et al. The natural history of latent rheumatic heart disease in a 5 year follow-up study: a prospective observational study. BMC Cardiovascular Disord. 2016;16:46.

10. Engelman D, Wheaton GR, Mataika RL, et al. Screening-detected rheumatic heart disease can progress to severe disease. Heart Asia. 2016;8:67-73.

11. Fabi M, Calicchia M, Miniaci A, et al. Carditis in acute rheumatic fever in a high-income and moderate-risk country. J Pediatr. 2019;215:187-91.

12. Haran S, Crane N , Kazi S , Axford-Haines L , White A. Effect of secondary penicillin prophylaxis on valvular changes in patients with rheumatic heart disease in Far North Queensland. Aust J Rural Health. 2018;26:119-25.

13. Majeed HA, Batnage S, Yousof AM, Khuffash F, Yusuf AR. Acute rheumatic fever and the evolution of rheumatic heart disease: a prospective 12 year follow-up report. J Clin Epidemiol. 1992;45:871-5.

14. Bonow RO, Carabello BA, Chatterjee K et al. 2008 Focused update incorporated into the ACC/AHA 2006 Guidelines for the management of patients with valvular heart disease: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Revise the 1998 Guidelines for the Management of Patients With Valvular Heart Disease): endorsed by the Society of Cardiovascular Anesthesiologists, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. Circulation. 2008;118:e523-661.