ISSN 2146-6505 • EISSN 2147-1894

Indexed in Web of Science

jarem

JOURNAL OF ACADEMIC RESEARCH IN MEDICINE

Original Investigations

VOLUME: 16 | ISSUE: 1 | April 2026

UNIVERSITY OF HEALTH SCIENCES TÜRKİYE GAZİOSMANPAŞA TRAINING AND RESEARCH HOSPITAL

galenos yayınevi

jarem.org

JAREM

JOURNAL OF ACADEMIC RESEARCH IN MEDICINE

Editor in Chief

Editors

English Language Editors

• **Owned by on behalf of the University of Health Sciences Türkiye, Gaziosmanpaşa Training and Research Hospital:** Prof. Dr. Ömer N. Develioğlu
• **Editor in Chief:** Prof. Dr. Ömer N. Develioğlu
• **Publication Type:** Periodical

JAREM
JOURNAL OF ACADEMIC RESEARCH IN MEDICINE

International Editorial Board

JOURNAL OF ACADEMIC RESEARCH IN MEDICINE

International Editorial Board

The previous Editor in Chief was Prof. Dr. Barış Nuhoğlu

İstanbul Yeniyüzyıl University, Gaziosmanpaşa Hospital Head of Department of Urology
(The editor-in-chief duty, which started in September 2011, was transferred to Ömer N Develioğlu in March 2016)

"Please refer to the journal's webpage (https://www.jarem.org/home) for "Journal Policies", "Instructions to Authors" and "About Us".

The Journal of Academic Research in Medicine and/or its editors are members of ICMJE, COPE, WAME, CSE and EASE, and follow their recommendations. Journal of Academic Research in Medicine is indexed in Web of Science-Emerging Sources Citation Index, TÜBİTAK ULAKBİM TR Dizin, EBSCO, Gale, CINAHL, J Gate, Türk Medline, Embase and CAB International (CABI).

The journal is published online.

Owner: *University of Health Sciences Türkiye, Gaziosmanpaşa Training and Research Hospital*

Responsible Manager: *Ömer N. Develioğlu*

galenos yayınevi

Publisher Contact
Address: Molla Gürani Mah. Kaçamak Sk. No: 21/1
34093 İstanbul, 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: April 2026

ISSN: 2146-6505 E-ISSN: 2147-1894

International periodical journal published three times in a year.

Jarem
JOURNAL OF ACADEMIC RESEARCH IN MEDICINE

Contents

DOI: 10.4274/jarem.galenos.2026.68926
J Acad Res Med 2026;16(1):1-3

Artificial Intelligence in Healthcare: A Critical Moment for Responsible Transformation

Seray Öney Doğanyiğit[1], Elgiz Yılmaz Altuntaş[2]

[1]İstanbul Bilgi University, Faculty of Communication, Department of Public Relations and Publicity, İstanbul, Türkiye
[2]Galatasaray University Faculty of Communication, Department of Interpersonal Communication, İstanbul, Türkiye

Cite this article as: Öney Doğanyiğit S, Yılmaz Altuntaş E. Artificial intelligence in healthcare: a critical moment for responsible transformation. J Acad Res Med. 2026;16(1):1-3

Keywords: Artificial intelligence, healthcare management, healthcare innovation, digital transformation, AI in healthcare

INTRODUCTION

Artificial intelligence (AI) is rapidly reshaping the global healthcare landscape. As rising costs, aging populations, chronic disease burden, and workforce shortages intensify pressure on health systems, AI has emerged as a powerful catalyst for innovation. From clinical applications—such as diagnostic imaging, pathology, and robotic surgery—to administrative automation and predictive analytics, AI is now embedded across multiple layers of healthcare.

However, its accelerated adoption also raises urgent questions regarding ethics, regulation, liability, data security, and workforce implications. This editorial synthesizes current evidence on AI's transformative potential while highlighting critical risks that must be addressed to ensure safe, equitable, and sustainable integration—an imperative especially relevant for healthcare systems across Europe and the Balkans.

The recent emergence of specialized generative AI frameworks, such as OpenAI's healthcare-focused initiatives, represents a pivotal shift from analytical tools to interactive clinical intelligence. While these large language models (LLMs) offer unprecedented support in reducing administrative burnout and enhancing clinical decision-making, they also amplify the necessity for rigorous "human-in-the-loop" governance—particularly in linguistically and structurally diverse regions like the Balkans.

AI's Expanding Role in Healthcare Administration

Administrative complexity contributes significantly to inefficiency, delays, and clinician burnout. AI-supported tools offer powerful opportunities to reduce this burden by automating routine tasks, flagging errors, and improving patient flow.

Streamlining Documentation and EHR Processes

Natural language processing systems now transcribe clinical encounters, generate documentation, and organize electronic health record (EHR) data with increasing accuracy, reducing manual workload (1). Predictive analytics integrated into EHRs can identify high-risk patients, enabling proactive intervention and improved resource allocation.

Improving Patient Flow and Emergency Response

AI-driven patient flow models have demonstrated improved bed utilization and reduced hospital congestion (2). In emergency settings, Internet of Things-enabled monitoring and AI-based triage algorithms can shorten ambulance response times and accelerate critical interventions (3).

Such applications are particularly relevant in low-resource or geographically dispersed health systems where staff shortages remain acute.

AI-supported Remote and Preventive Care

Virtual health assistants and remote monitoring systems represent a major shift toward decentralized care. AI-enabled platforms assist patients with medication adherence, symptom tracking, appointment reminders, and chronic disease monitoring (4).

Examples such as "Florence", a virtual nursing assistant, illustrate how AI can extend clinical support beyond traditional settings (5). These tools promote treatment adherence, reduce unnecessary hospital visits, and enhance preventive care—outcomes especially valuable in chronic disease management and elderly populations.

ORCID IDs of the authors: S.Ö.D. 0000-0002-8175-0764; E.Y.A. 0000-0002-3486-1701

Corresponding Author: Elgiz Yılmaz Altuntaş, Prof. Dr.;
E-mail: elyilmaz@gsu.edu.tr

Received Date: 21.11.2025 **Accepted Date:** 13.04.2026
Publication Date: 28.04.2026

J Acad Res Med 2026;16(1):1-3

Clinical Applications with Significant Impact

Early Diagnosis and Cancer Detection

Early detection remains one of AI's most successful domains. Deep learning models have matched dermatologist-level accuracy in classifying skin lesions (6) and achieved high performance in interpreting mammography images (7). Machine learning algorithms have also shown promise in predicting tumor recurrence using large-scale pathological datasets (8).

These tools offer substantial benefits in countries facing limited specialist availability or uneven distribution of diagnostic services.

Radiology and Imaging

Radiology has become a central testing ground for AI innovation. AI-enhanced ultrasound systems now identify anatomical structures in real time and provide differential diagnosis support, improving accuracy and efficiency in lung and obstetric imaging.

Machine learning algorithms have also been successfully applied to maternal health, predicting postpartum depression and fetal anomalies using EHR data (9). These innovations highlight AI's potential to support maternal and perinatal care, which remains an ongoing challenge in many regions.

OpenAI's next-generation models appear capable of analyzing text, audio, and images simultaneously (for example, audio clinical notes alongside radiology images). These features, which support the concept of interactive clinical intelligence, exemplify how AI is no longer a passive tool but an active "clinical partner".

Robotic Surgery and Clinical Robotics

Surgical robotics continues to expand rapidly. Systems such as the da Vinci Surgical System offer reduced complications, shorter hospital stays, and improved surgeon ergonomics (10). AI-powered robotic platforms also assist with routine tasks, medication delivery, and early sepsis detection, increasing safety in intensive care settings (11).

Drug Discovery and Precision Medicine

AI accelerates drug discovery through rapid target identification, molecular modeling, and drug repurposing. Partnerships between industry and AI firms—such as GlaxoSmithKline and Exscientia—have reduced development timelines by up to 75% (12).

AI-supported pharmacogenomics enables personalized medicine by predicting individual responses to medications, including psychopharmacological treatments (13). These advances promise to reduce trial-and-error prescribing and improve treatment outcomes.

Challenges and Risks of Accelerated AI Adoption

Despite its potential, AI introduces substantial risks that must be managed through careful oversight.

Workforce Displacement and Skill Shifts

While AI reduces clinician workload, it also automates tasks previously performed by healthcare professionals. This shift may result in job displacement or require reskilling, potentially creating anxiety and resistance within the workforce (7).

Loss of Humanistic Care

AI cannot replicate empathy, emotional intelligence, or relational communication. Overreliance on automation risks weakening the doctor-patient relationship, especially in communities where trust in healthcare providers is culturally significant (14).

Bias, Inequity, and Fairness

AI models trained on biased datasets may perpetuate or exacerbate inequalities in diagnosis, treatment recommendations, and recruitment (4). Ensuring diverse and representative data is essential for equitable AI performance.

High Implementation Costs and Security Risks

Deployment of AI systems requires significant investment in infrastructure, workforce training, and cybersecurity. Data breaches remain a major concern, particularly as health systems across the Balkans modernize digital platforms (7).

Beyond traditional data security, the rise of specialized generative AI—exemplified by OpenAI's healthcare frameworks—introduces the challenge of epistemic security. As LLMs transition from analytical tools to interactive clinical partners, the risk of "hallucinations"—the generation of plausible but medically inaccurate information—poses a direct threat to the integrity of medical knowledge. In regions with diverse linguistic landscapes, such as the Balkans, ensuring that AI-driven health communication maintains cognitive cybersecurity and prevents the spread of algorithmic disinformation is paramount. Therefore, responsible transformation must prioritize not only the technical accuracy of AI but also the protection of the information ecosystem in which clinicians and patients interact. It may be argued that the risk of AI-generated medical disinformation (hallucinations) infiltrating clinical decision-making processes is not only a data security issue but also an information security issue.

Legal and Ethical Uncertainty

Perhaps the most pressing challenge is liability. When AI-generated recommendations or robotic procedures cause harm, it is unclear whether responsibility lies with developers, healthcare providers, or institutions. This ambiguity underscores the need for standardized regulatory frameworks. For example, the European Union AI Law, which has been frequently cited, addresses not only the "mistakes" that AI makes but also the possibility of presenting misinformation as "correct", thus highlighting the risk of AI tools disrupting the medical information ecosystem (15).

Toward Responsible and Equitable AI Integration

AI's future in healthcare depends not on technological capability alone, but on governance, transparency, and ethical stewardship.

J Acad Res Med 2026;16(1):1-3

Regulation and Governance

Countries must implement regulatory frameworks that address:

- Algorithmic transparency
- Data protection and privacy
- Clinical validation requirements
- Accountability and liability
- Safety monitoring throughout the system lifecycle

Without these safeguards, AI risks undermining public trust and exacerbating inequities.

Human Oversight and Clinical Expertise

AI should augment—not replace—clinicians. Human-in-the-loop models ensure contextual judgment, empathy, and moral reasoning remain central to healthcare delivery. Although AI is known to be incapable of empathy, simulations of "scalable empathy" suggest it could play a supportive role in patient communication (for example, in triage or chronic disease monitoring) (16). According to Schneiders and van der Graaf (17), traditional empathy is a process where one person senses another's feelings, but it is resource-limited (requiring time and energy) and can be tiring (burnout). Scalable empathy, on the other hand, is the simulation of this process at a behavioral level through AI, enabling it to be delivered to millions of patients simultaneously and seamlessly.

Inclusive and Representative Data

Developers must prioritize datasets reflecting diverse populations, ensuring AI systems perform equitably across gender, ethnicity, age, and socioeconomic groups.

Workforce Training and System Preparedness

Healthcare professionals need training in digital literacy, AI interpretation, and ethical use. Health systems must also assess their readiness for technological integration, considering financial, infrastructural, and cultural factors.

CONCLUSION

AI offers extraordinary promise. It can reduce errors, accelerate diagnosis, automate administrative tasks, support overextended workforces, and promote personalized care. Yet without ethical safeguards, regulatory clarity, and equitable implementation, it risks amplifying disparities and creating new vulnerabilities.

Healthcare stands at a pivotal moment. The goal is not merely to adopt AI technologies but to ensure they strengthen the foundations of care—safety, equity, empathy, and trust. A future where AI enhances healthcare for all will require deliberate, responsible, and collaborative action across the global health community.

Footnotes

Author Contributions: Concept - S.Ö.D., E.Y.A.; Design - E.Y.A.; Analysis and/or Interpretation - S.Ö.D., E.Y.A.; Literature Search - S.Ö.D.; Writing - S.Ö.D., E.Y.A.

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

Financial Disclosure: The authors report that no financial support was received for this study.

REFERENCES

1. Akalın B, Veranyurt Ü. Sağlıkta dijitalleşme ve yapay zeka [Digitalization in health and artificial intelligence]. SDÜ Sağlık Yönetimi Dergisi. 2020; 2: 131-41. Turkish.
2. Hu Y, Dong J, Perry O, Cyrus RM, Gravenor S, Schmidt MJ. Use of a novel patient-flow model to optimize hospital bed capacity for medical patients. Jt Comm J Qual Patient Saf. 2021; 47: 354-63.
3. Saha HN, Raun NF, Saha M. Monitoring patient's health with smart ambulance systems using Internet of Things (IoTs). IEMECON. 2017; 8: 91-5.
4. Duran GN. Economic impacts of artificial intelligence in healthcare (dissertation). Ankara: Hacettepe University. 2024.
5. HealthSnap. AI in remote patient monitoring: The top 4 use cases in 2024. (serial online). 2024 (cited 2025 Oct 22). Available from: https://healthsnap.io/ai-in-remote-patient-monitoring-the-top-4-use-cases-in-2024/
6. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017; 542: 115-8.
7. Yorgancıoğlu Tarcan G, Yalçın Balçık P, Sebik NB. Türkiye ve dünyada sağlık hizmetlerinde yapay zeka [Artificial intelligence in healthcare in Türkiye and the world]. Lokman Hekim J Med Hist Folkl Med. 2024; 14: 50-60.
8. Sone K, Toyohara Y, Taguchi A, Miyamoto Y, Tanikawa M, Uchino-Mori M, et al. Application of artificial intelligence in gynecologic malignancies: A review. J Obstet Gynaecol Res. 2021; 47: 2577-85.
9. Larson DB, Chen MC, Lungren MP, Halabi SS, Stence NV, Langlotz CP. Performance of a deep-learning neural network model in assessing skeletal maturity on pediatric hand radiographs. Radiology. 2018; 287: 313-22.
10. Peters BS, Armijo PR, Krause C, Choudhury SA, Oleynikov D. Review of emerging surgical robotic technology. Surg Endosc. 2018; 32: 1636-55.
11. Kalil AJ, Dias VMCH, Rocha CC, Morales HMP, Fressatto JL, Faria RA. Sepsis risk assessment: a retrospective analysis after a cognitive risk management robot (Robot Laura®) implementation in a clinical-surgical unit. Research on Biomedical Engineering. 2018; 34: 310-6.
12. Agrawal P. Artificial intelligence in drug discovery and development. Journal of Pharmacovigilance. 2018; 6: 1000e173.
13. Erden BF, Mutlu O, editörler. Yapay zekâ ve psikofarmakolojide deneysel modeller: teoriden pratiğe. Kocaeli: Kocaeli Üniversitesi Vakfı (KÜV) Yayınevi; 2024.
14. Jarrahi MH. Artificial intelligence and the future of work: human-AI symbiosis in organizational decision making. Business Horizons. 2018: 61; 577-86.
15. European Commission. Laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts. Proposal for a regulation of the European parliament and of the council. 2021.
16. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Internal Medicine. 2023; 83; 589-96.
17. Naing SY, Severin J, Zomer A, Kuntaman K, Suandy I, Sunandar S, et al. Impact of reducing colistin use on colistin resistance in *Escherichia coli* isolated from humans and poultry in Indonesia (COINCIDE): a protocol for a multisectoral, transdisciplinary One Health study. One Health. 2026; 22: 101347.

DOI: 10.4274/jarem.galenos.2025.82474
J Acad Res Med 2026;16(1):4-9

Assessment of the Need for Voice Therapy after Type 3 Thyroplasty: Experience of 31 Patients

Buse Demir[1], Kürşat Yelken[2]

[1]University of Health Sciences Türkiye, Ümraniye Training and Research Hospital, Department of Speech-Language Pathology, İstanbul, Türkiye
[2]Voicest Clinic, Clinic of Otolaryngology, İstanbul, Türkiye

Cite this article as: Demir B, Yelken K. Assessment of the need for voice therapy after type 3 thyroplasty: experience of 31 patients. J Acad Res Med. 2026;16(1):4-9

ABSTRACT

Objective: The aim of the study is to determine the need for voice therapy after type 3 thyroplasty surgery by using the voice self-assessment method.

Methods: A total of 31 male patients who underwent type 3 thyroplasty and 31 male control subjects without voice or pitch complaints were included. Each group's Turkish Voice Handicap Index-10 and Turkish Self-evaluation of Voice Quality Survey were administered online.

Results: There are statistically significant differences between surgery and control groups in both total and subgroup scores of the Turkish Voice Handicap Index-10 Scale and the Turkish Self-evaluation of Voice Quality Survey.

Conclusion: Individuals who have undergone a voice-deepening procedure experience emotional, physical, and functional problems related to their voice. In these patients, voice therapy in support of surgery is necessary to prevent potential hypofunction, improve voice quality, and preserve vocal health.

Keywords: Laryngeal surgery, pitch perception, throplasty, voice disorders, voice training, voice quality

INTRODUCTION

Voice plays a role not only in conveying verbal language to the listener but also in conveying emotions and thoughts. Voice should correspond to a person's age and gender, as it significantly reflects personality, behavior, and mood, and is an important secondary sexual characteristic (1). The relationship between voice and gender involves pitch, intonation, resonance, timbre, articulation, breathiness, intensity, and non-verbal communication. Having a tone of voice perceived as unsuitable for one's gender or appearance can adversely affect social and professional life as well as one's quality of life (2). Lowering the pitch of the voice is the primary treatment goal for individuals who have resisted at least three months of voice therapy, those with puberphonia, sulcus vocalis, vocal fold atrophy, vocal fold scar, various constitutional voice disorders, trans males who desire a masculine voice but do not want hormone therapy or who have insufficient pitch drop despite androgen therapy, and men who, for personal reasons, want to lower their pitch without a pathological condition (3,4).

Type 3 thyroplasty is performed to lower voice pitch (5). It reduces the tension of the vocal folds, leading to a lower fundamental frequency and a deeper voice quality. The effectiveness of type 3 thyroplasty has been well demonstrated in the literature (4,6-10). To our knowledge, the need for postoperative voice therapy has not been studied. To determine treatment approaches for lowering voice pitch in individuals with high-pitched voices, Kizilay and Firat (11) used a voice assessment form and found that voice-related emotional problems were high before treatment, whereas physical scores were high after treatment, indicating that patients had difficulty using their new voices. The aim of this study is to investigate whether individuals require voice therapy after type 3 thyroplasty. For this purpose, the Turkish Voice Handicap Index-10 (TVHI-10) and the Turkish Self-evaluation of Voice Quality Survey (T-SVQS) (Appendix 1) will be used to investigate individuals' emotional, physical, and functional problems related to their voice in their daily lives after type 3 thyroplasty.

ORCID IDs of the authors: B.D. 0000-0002-4376-6909; K.Y. 0000-0001-8133-2717

Corresponding Author: Buse Demir;
E-mail: demiirbuse@gmail.com
Received Date: 03.06.2025 **Accepted Date:** 05.12.2025
Epub: 26.01.2026
Publication Date: 28.04.2026

J Acad Res Med 2026;16(1):4-9

METHODS

Study Desing and Participants

The study was conducted between 2018 and 2023, and included 31 male native Turkish-speaking patients who sought medical attention for a complaint of a high-pitched voice and underwent surgical intervention. Their ages ranged from 20-53 years (mean age 33.39±8.1 years). Postoperative duration was classified into three categories: 19 patients underwent voice-deepening surgery more than 12 months previously, 6 patients underwent voice-deepening surgery between 6 and 12 months previously, and the remaining 6 patients underwent voice-deepening surgery less than 6 months previously. The second author carried out all the surgeries. Among these patients, one had puberphonia, 14 had sulcus vocalis, and 16 received a diagnosis of constitutional high-pitched voice.

Surgical Procedure and Voice Assessment Instruments

All the subjects were examined by means of laryngostroboscopy, and a diagnosis of sulcus vocalis was made when there was a channel or stria that ran parallel to the free edge of the vocal folds, causing unilateral or bilateral glottal incompetence, with decreased or even absent glottal wave amplitude over the lesion. All patients in the surgical group underwent type 3 thyroplasty performed by the same experienced laryngologist to ensure standardization. The procedure was conducted under local anesthesia with sedation. After a horizontal cervical incision at the level of the thyroid cartilage, a small rectangular window was created in the thyroid lamina. The anterior portion of the thyroid cartilage was then approximated to shorten and relax the vocal folds, thereby lowering the pitch. No intraoperative or postoperative complications were observed. The control group consists of 31 individuals matched for age and gender. Control participants did not undergo laryngoscopy; however, all denied vocal symptoms or pitch-related complaints. The exclusion criteria for both the surgery and control groups include being a professional voice user, having extralaryngeal pathology, being female, having previously received voice therapy, and having systemic and/or pulmonary diseases. In this experimental study, voluntary participants were directed, via an online survey, to complete the TVHI-10 scale and the T-SVQS, which were prepared in accordance with the subject and purpose of the study.

Statistical Analysis

All data analyses were conducted using SPSS version 25.0 (IBM Corp., Armonk, NY, USA). The normality of quantitative variables was assessed using the Shapiro-Wilk test. For non-normally distributed data ($p<0.05$), non-parametric statistical tests were used. Descriptive statistics were reported as median ± interquartile range [(IQR), Q1-Q3] for non-normally distributed data and as mean ± standard deviation (SD) for normally distributed data. The IQR value is equal to the diffirence between Q3 and Q1. The Mann-Whitney U test was used for comparisons between two independent groups, and the Kruskal-Wallis test was used for comparisons among more than two groups. For categorical variables, chi-square tests were performed. The internal consistency of the scales was evaluated using Cronbach's alpha coefficient. Construct validity was assessed with exploratory factor analysis (EFA) using Varimax rotation. The Kaiser-Meyer-Olkin (KMO) measure and Bartlett's test of sphericity were used to assess sampling adequacy and the suitability of the correlation matrix for factor analysis, respectively. A p-value of <0.05 was considered statistically significant. The number of participants in each subgroup (n) was clearly stated in all comparison tables.

Participant Information Sheet and Informed Consent Form

This study, titled "Investigation of the Need for Voice Therapy Following Voice Deepening Surgery Using the Self-Assessment Method of Voice," aims to evaluate the self-perceptions of patients who underwent type 3 thyroplasty for voice-deepening and did not receive postoperative voice therapy. The study seeks to assess whether there is a need for voice therapy after surgery based on these self-assessments.

Voluntary participants will be asked to evaluate the quality of their voice and the impact of their voice on quality of life. Voice assessments are typically categorized as objective or subjective. Self-assessment, a form of subjective evaluation, is important because it helps identify the individual's perception of their condition and of the significance of the problem in their life. The purpose of self-assessment is to understand the extent of deviation in voice quality, the impact on professional and social life, and the possible emotional reactions related to voice disorders.

Self-assessment tools are also useful in determining the effectiveness of treatment by measuring how much the difficulties caused by a voice disorder have been alleviated. Among the available self-assessment tools, the most widely used is the VHI. The VHI-10, a shortened version, is a 10-item instrument with a 5-point Likert scale (0=never, 4=always). A higher total score or item score indicates a greater impact of the voice problem on the individual's quality of life.

Another important dimension in voice evaluation is how voice quality is perceived by both the individual and others in their environment. To address this, a 10-question SVQS, developed specifically for this study, will be used. Participants will assess their current voice pitch, quality, loudness, and strength by choosing among the options: "always," "never," or "sometimes."

While the VHI-10 will evaluate the impact of voice on quality of life, the SVQS will assess participants' perception of their voice.

The purpose of this study is to emphasize the need for voice therapy following voice-deepening surgery to stabilize the achieved fundamental frequency, improve voice quality and quality of life, prevent potential hypofunction, increase vocal loudness, and reduce air leakage and vocal fatigue. This study also aims to contribute to the limited literature on this topic.

J Acad Res Med 2026;16(1):4-9

Participants in the experimental group were adult Turkish-speaking male patients (n=31) who underwent type 3 thyroplasty for pitch lowering between 2018-2023. An age- and gender-matched control group (n=31) was also recruited.

Participants will receive an informed consent form, a personal information form, the VHI-10, and the SVQS via email, using the online platform Google Forms. It is estimated that completing the forms will take approximately 5-10 minutes.

There are no known risks or undesirable effects associated with participation in this retrospective, descriptive study. This study aims to investigate whether voice-deepening surgery through self-assessment of voice.

Your participation in this study is completely voluntary, and you have the right to refuse to participate. You may withdraw from the study at any time by notifying the researcher. The researcher also reserves the right to exclude you from the study if necessary. Refusing to participate in or withdrawing from the study will not affect your medical care or your relationship with your physician.

You will not incur any financial costs as part of this study, nor will you receive payment for your participation.

Any data collected from you will be used solely for the purposes of this study. Your identity and confidentiality will be protected throughout all stages of the research.

Participant Declaration

I have been informed about a scientific research study to be conducted by the responsible research team. I have received detailed information regarding the purpose, methods, and procedures of the study. I have been invited to participate as a volunteer.

I understand that any personal or medical information shared during this study will be treated with strict confidentiality and used solely for educational and scientific purposes, ensuring full respect for my privacy.

I am aware that I may withdraw from the study at any time without providing a reason. I also acknowledge that the research team may withdraw me from the study if medically necessary, and that this will not cause any harm to my health.

I will not be held financially responsible for any part of the study, and I understand that I will not receive payment for my participation. In the event of any health issues arising directly or indirectly from the study, I have been assured that all necessary medical interventions will be provided at no cost to me.

I understand that I may contact the research team at any time if I have questions or experience any health concerns during the study.

I confirm that I have not been pressured or coerced into participating, and that my refusal to participate will not affect my medical treatment or relationship with healthcare providers.

Having understood all the information provided and after allowing myself sufficient time to consider it, I voluntarily agree to participate in this study.

Ethics Approval

This study was reviewed and approved by the Non-Interventional Scientific Research Ethics Committee of İstanbul Atlas University (approval no: E-22686390-050.99-28121, date: 15.06.2023). The research titled "Investigating the Need for Voice Therapy After Voice Deepening Surgery Through the Method of Self-Assessment of Voice" was found to be ethically appropriate. All procedures were carried out in accordance with relevant guidelines and regulations. Informed consent was obtained form all individual participants included in the study.

RESULTS

The Shapiro-Wilk test indicated that the quantitative data were not normally distributed ($p<0.05$); therefore, non-parametric statistical tests were applied for all analyses. The surgery group (n=31) and the control group (n=31) differed significantly in both total and subgroup scores on the TVHI-10 and the T-SVQS (Mann-Whitney U test, $p<0.001$). A cut-off value of ≥11 points was considered indicative of a voice handicap, based on the Turkish validation study by Kıyak et al. (12). The mean ± SD of the total TVHI-10 score was 14.26±11.00 (median=14, range 0-36) in the surgery group and 1.94±2.98 (median=1, range 0-12) in the control group. The difference between groups was statistically significant (Mann-Whitney U test, $p<0.001$). According to the Turkish validation study, a cut-off value of ≥11 on the TVHI-10 indicates the presence of a clinically significant voice handicap. Based on this threshold, 17 of 31 patients (54.8%) in the surgery group exceeded the cut-off, whereas only 2 of 31 (6.5%) in the control group did so. This finding demonstrates that, despite having undergone type 3 thyroplasty, more than half of patients still reported some degree of voice handicap, highlighting the importance of subjective voice evaluation during postoperative follow-up. The median (IQR) total TVHI-10 score was higher in the surgery group than in the control group. Similarly, T-SVQS scores were significantly higher in the surgery group (Mann-Whitney U test, $p<0.001$; Table 1). The distribution of responses to each item of the T-SVQS was analyzed using the chi-square test. Statistically significant differences were observed between the surgery and control groups for items 1,2,4,5,6,7,8,9, and 10 (all $p<0.05$). Only item 3 did not show a statistically significant difference ($p>0.05$), which suggests consistent item-level sensitivity across most survey questions (Table 2). To explore the relationship between perceived voice quality and voice handicap in the surgical group, Spearman's rank correlation coefficient was used. Strong positive correlations were observed between the T-SVQS scores and the emotional ($r=0.841$, $p<0.001$), physical ($r=0.836$, $p<0.001$), and functional ($r=0.814$, $p<0.001$) subscales of the TVHI-10, as well as the total score ($r=0.861$, $p<0.001$). All correlations were statistically significant ($p<0.001$; Table 3). The construct validity of the T-SVQS was assessed by EFA. The data were deemed suitable for factor analysis as indicated by a KMO value of 0.725 and a significant result in Bartlett's test of sphericity ($\chi^2=152.594$, $df=45$, $p<0.001$). A single-factor solution was extracted using Varimax rotation, explaining 63.5% of the total

J Acad Res Med 2026;16(1):4-9

Table 1. Comparison of scale and survey scores by groups

Variant	M (Q3-Q1)		p-value
	Surgery (+) (n=31)	Surgery (-) (n=31)	
Emotional	3 (5-0)	0 (0-0)	<0.001*
Physical	3 (8-0)	0 (0-0)	<0.001*
Functional	7 (12-3)	1 (2-0)	<0.001*
TVHI-10	14 (23-4)	1 (2-0)	<0.001*
T-SVQS	9 (14-5)	2 (3-1)	<0.001*

*: p<0.05 there is a statistically significant difference between groups, n: Number, M: Median, Q1; First quartile (25th percentile), Q3: Third quartile (75th percentile), T-SVQS: Turkish Self-evaluation of Voice Quality Survey, TVHI-10: Turkish Voice Handicap Index-10

Table 2. Comparison of Turkish Self-evaluation of Voice Quality Survey questionnaries by groups

Variant	Group	n/%	Group		Total	χ^2	p-value
			Surgery (+) (n=31)	Surgery (-) (n=31)			
Q1: Do you think that your voice pitch is deep enough for your age. gender. work and social life?	Never	n/%	6/19.4%	0/0.0%	6/9.7%	25.268	<0.001*
	Sometimes	n/%	17/54.8%	5/16.1%	22/35.5%		
	Always	n/%	8/25.8%	26/83.9%	34/54.8%		
Q2: Does the pitch of your voice prevent you from practicing your profession?	Never	n/%	15/48.4%	30/96.8%	45/72.6%	21.316	<0.001*
	Sometimes	n/%	14/45.2%	1/3.2%	15/24.2%		
	Always	n/%	2/6.5%	0/0.0%	2/3.2%		
Q3: Is your voice perceived differently on the phone than your gender identifies?	Never	n/%	23/74.2%	27/87.1%	50/80.6%	3.496	>0.05
	Sometimes	n/%	6/19.4%	4/12.9%	10/16.1%		
	Always	n/%	2/6.5%	0/0.0%	2/3.2%		
Q4: Do you ever have trouble controlling the pitch of your voice?	Never	n/%	6/19.4%	24/77.4%	30/48.4%	24.435	<0.001*
	Sometimes	n/%	21/67.7%	7/22.6%	28/45.2%		
	Always	n/%	4/12.9%	0/0.0%	4 /6.5%		
Q5: Do you ever experience a high pitched voice?	Never	n/%	10/32.3%	20/64.5%	30/48.4%	12.334	<0.01*
	Sometimes	n/%	15/48.4%	11/35.5%	26/41.9%		
	Always	n/%	6/19.4%	0/0.0%	6/9.7%		
Q6: Does your voice sound hoarse to you or to those around you?	Never	n/%	9/29.0%	18/58.1%	27/43.5%	15.535	<0.001*
	Sometimes	n/%	13/41.9%	13/41.9%	26/41.9%		
	Always	n/%	9/29.0%	0/0.0%	9/14.5%		
Q7: Do you experience vocal fatigue when speaking?	Never	n/%	8/25.8%	14/45.2%	22/35.5%	14.425	<0.001*
	Sometimes	n/%	14/45.2%	17/54.8%	31/50.0%		
	Always	n/%	9/29.0%	0/0.0%	9/14.5%		
Q8: Do you ever feel breathiness in your voice?	Never	n/%	15/48.4%	26/83.9%	41/66.1%	12.225	<0.01*
	Sometimes	n/%	11/35.5%	5/16.1%	16/25.8%		
	Always	n/%	5/16.1%	0/0.0%	5/8.1%		
Q9: Do you have problems changing the loudness and volume of your voice (shouting. whispering. etc.)?	Never	n/%	9/29.0%	26/83.9%	35/56.5%	24.146	<0.001*
	Sometimes	n/%	14/45.2%	5/16.1%	19/30.6%		
	Always	n/%	8/25.8%	0/0.0%	8/12.9%		
Q10: Do you ever feel weakness in your voice?	Never	n/%	6/19.4%	24/77.4%	30 /48.4%	31.869	<0.001*
	Sometimes	n/%	11/35.5%	7/22.6%	18/29.0%		
	Always	n/%	14/45.2%	0/0.0%	14/22.6%		
Total		n/%	31/100.0%	31/100.0%	62/100.0%		

*: p<0.05 there is a statistically significant difference between groups, n: Number, %: Percentage

J Acad Res Med 2026;16(1):4-9

Table 3. Correlations between scale and survey scores

Points	Value	Emotional	Physical	Functional	TVHI-10
T-SVQS	r	0.841	0.836	0.814	0.861
	p	<0.001*	<0.001*	<0.001*	<0.001*
Emotional	r		0.854	0.925	0.967
	p		<0.001*	<0.001*	<0.001*
Physical	r			0.887	0.939
	p			<0.001*	<0.001*
Functional	r				0.968
	p				<0.001*

*: p<0.05 there is a statistically significant difference between groups, r: correlation coefficient, T-SVQS: Turkish Self-evaluation of Voice Quality Survey, TVHI-10: Turkish Voice Handicap Index-10
Correlations were calculated based on data from participants who had undergone surgery (n=31)

variance. Factor loadings for the items ranged from 0.354 to 0.844. The internal consistency of the T-SVQS was high, with a Cronbach's alpha coefficient of 0.845. Item-total correlation coefficients ranged between 0.307 and 0.763, indicating acceptable reliability for all items. These results support the validity and reliability of the Turkish version of the T-SVQS for assessing self-perceived voice quality.

DISCUSSION

This study investigated the need for voice therapy after type 3 thyroplasty using self-assessment of voice. Type 3 thyroplasty is considered a definitive treatment, this surgery effectively lowers vocal pitch without compromising vocal quality (5). Chandra et al. (13) reported a significant subjective improvement in patients with puberphonia, with VHI scores decreasing from a preoperative 53 to a postoperative 29.

However, the comparison of scale and survey scores among groups in our study revealed a statistically significant difference. Our results indicate that individuals continue to experience emotional, physical, and functional voice problems after type 3 thyroplasty; therefore, voice therapy is necessary.

Kizilay and Firat (11) reported, based on the voice assessment form for patients with puberphonia, that the highest pre-treatment handicap scores were observed in the emotional subgroup, whereas post-treatment the highest scores shifted to the physical subgroup. This increase in post-treatment physical scores reflects the challenges patients encounter when adapting to their new voices and underscores the ongoing need for voice therapy.

Comparison of responses to the T-SVQS across groups revealed that individuals did not experience problems with gender perception despite the decrease in fundamental frequency following surgery. However, they exhibit hypofunctional voice characteristics such as vocal fatigue, breathiness, asthenia, reduced loudness, and difficulty controlling intensity. While the reduction in fundamental frequency and the resolution of gender-perception issues are expected outcomes of type 3 thyroplasty, the decrease in vocal fold tension postoperatively seems to lead to glottic gap formation and discordant vocal fold vibration. This hypofunction results in decreased voice quality. Therefore, voice therapy by a speech and language therapist is essential to ensure proper voice use, to provide respiratory support, and to prevent potential hypofunction after surgery (10).

Findings from correlation analysis between the scales indicated that as T-SVQS scores increased, the TVHI-10 Scale total and subgroup scores also increased. Emotional subgroup scores were associated with increases in the total score and in other subgroup scores. Physical subgroup scores were correlated with increases in total-scale and functional-subgroup scores. Functional subgroup scores were also positively correlated with total scale scores. These analyses suggest a strong relationship between T-SVQS scores and the TVHI-10, showing that the emotional, physical, and functional subgroups are interconnected.

In conclusion, patients who underwent type 3 Thyroplasty exhibited emotional, physical, and functional problems related to their voice. In these patients, adjunctive voice therapy alongside surgery is necessary to prevent potential vocal hypofunction, improve voice quality, preserve vocal health, and reduce emotional, physical, and functional effects associated with voice disorders.

Study Limitations

This study has several limitations. First, using an online platform to administer study questions may have introduced variability in participants' understanding because of sociocultural differences, potentially affecting response consistency. Second, the number of patients included in the study was limited because type 3 thyroplasty is a highly specialized, infrequently performed surgical procedure. This reduced the sample size and may limit the generalizability of the findings. Finally, the limited availability of literature on type 3 thyroplasty, owing to its non-routine application, restricted comparison and corroboration of the study's results with existing evidence.

J Acad Res Med 2026;16(1):4-9

CONCLUSION

This study demonstrated that individuals who underwent type 3 thyroplasty continued to experience emotional, physical, and functional voice difficulties despite the surgical lowering of vocal pitch. The Turkish version of the T-SVQS was found to be a valid and reliable tool in assessing self-perceived voice quality, showing strong correlations with the TVHI-10. The results emphasize that surgery alone may not fully address the complex vocal challenges faced by these patients. Therefore, integrating postoperative voice therapy into the treatment protocol is essential to enhance voice quality, prevent hypofunctional voice patterns, and support long-term vocal health and well-being.

Appendix 1: https://d2v96fxpocvxx.cloudfront.net/beb8919b-f013-4ea1-b1c8-40332e840fe1/content-images/dfd0a53c-fa6d-4c82-a242-72e513648dfe.pdf

Ethics

Ethics Committee Approval: This study was reviewed and approved by the Non-Interventional Scientific Research Ethics Committee of İstanbul Atlas University (approval no: E-22686390-050.99-28121, date: 15.06.2023).

Informed Consent: Informed consent was obtained form all individual participants included in the study.

Acknowledgments

We thank the biostatistics consultant for support with the statistical analyses.

Translation was carried out by one of the authors, Kürşat Yelken.

Footnotes

Author Contributions: Surgical and Medical Practices - K.Y.; Concept - B.D., K.Y.; Design - B.D., K.Y.; Data Collection and/or Processing - B.D.; Analysis and/or Interpretation - B.D., K.Y.; Literature Search - B.D.; Writing - B.D.

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

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

REFERENCES

1. Boone DR, McFarlane SC, Von Berg SL, Zraick RI. Normal voice: anatomy and physiology throughout the lifespan. In: The voice and voice therapy. 9th ed. New Jersey: Pearson; 2014. p. 19-64.
2. Hancock AB. An ICF perspective on voice-related quality of life of American transgender women. J Voice. 2017; 31: 115.e1-115.e8.
3. Saito Y, Nakamura K, Itani S, Tsukahara K. Type 3 thyroplasty for a patient with female-to-male gender identity disorder. Case Rep Otolaryngol. 2018; 2018: 4280381.
4. García-López I, Peñarrocha J, Gavilan J. Tiroplastia tipo III para el tratamiento de la disfonía por elevación del tono [Type III thyroplasty for the treatment of high-pitched voice disorder]. Acta Otorrinolaringol Esp. 2010; 61: 318-20. Spanish.
5. Isshiki N. Mechanical and dynamic aspects of voice production as related to voice therapy and phonosurgery. J Voice. 1998; 12: 125-37.
6. Remacle M, Matar N, Verduyckt I, Lawson G. Relaxation thyroplasty for mutational falsetto treatment. Ann Otol Rhinol Laryngol. 2010; 119: 105-9.
7. Nakamura K, Tsukahara K, Watanabe Y, Komazawa D, Suzuki M. Type 3 thyroplasty for patients with mutational dysphonia. J Voice. 2013; 27: 650-4.
8. Kocak I, Dogan M, Tadihan E, Alkan Cakir Z, Bengisu S, Akpinar M. Window anterior commissure relaxation laryngoplasty in the management of high-pitched voice disorders. Arch Otolaryngol Head Neck Surg. 2008; 134: 1263-9.
9. Chowdhury K, Saha S, Pal S, Chatterjee I. Effects of type 3 thyroplasty on voice quality outcomes in puberphonia. Phil J Otolaryngol Head Neck Surg. 2014; 29: 6-10.
10. Yelken K, Coşkun H. Laryngeal framework surgeries. In: Öz F, editor. Laryngology in all aspects. Ankara: Nobel Medical Bookstores; 2021. p. 389-93.
11. Kizilay A, Firat Y. Puberfoni hastalarinda tedavi şemasi [Treatment algorithm for patients with puberphonia]. Kulak Burun Bogaz Ihtis Derg. 2008; 18: 335-42. Turkish.
12. Kıyak E, Haznedaroğlu Ş, Özcebe E. Reliability and validity of the Turkish version of the Voice Handicap Index-10 (TVHI-10). J Voice. 2016; 30: 644.e9-644.e14.
13. Chandra ST, Rao SM, Kumar AY, Murthy PSN. Puberphonia. Int J Phonosurg Laryngol. 2011; 1: 19-20.

DOI: 10.4274/jarem.galenos.2025.42243
J Acad Res Med 2026;16(1):10-16

Analgesic Efficacy of Systemic versus Interfascial Dexamethasone in Anterior Quadratus Lumborum Block for Laparoscopic Cholecystectomy: A Retrospective Cohort Study

Serpil Şehirlioğlu, Döndü Genç Moralar, Veysel Dinç

University of Health Sciences Türkiye, Gaziosmanpaşa Training and Research Hospital, Department of Anaesthesia and Reanimation, İstanbul, Türkiye

Cite this article as: Şehirlioğlu S, Genç Moralar D, Dinç V. Analgesic efficacy of systemic versus interfascial dexamethasone in anterior quadratus lumborum block for laparoscopic cholecystectomy: a retrospective cohort study. J Acad Res Med. 2026;16(1):10-16

ABSTRACT

Objective: The anterior quadratus lumborum block (aQLB) is commonly used for postoperative pain relief following laparoscopic cholecystectomy (LC). Dexamethasone is known to prolong and enhance analgesia when used as an adjuvant in peripheral nerve blocks. Some studies suggest that systemic administration may be as effective as perineural administration. However, studies directly comparing interfascial administration of dexamethasone in fascial plane blocks with systemic administration remain limited. This study aims to compare the analgesic efficacy and side-effect profiles of interfascial versus systemic administration of dexamethasone with the aQLB.

Methods: This retrospective study included 72 patients who underwent elective LC under general anesthesia. All patients received preoperative bilateral aQLBs. Patients were assigned to Group IV (n=33), which received intravenous dexamethasone, or to Group IF (n=39), which received interfascial dexamethasone. The primary endpoint was the time to first rescue analgesic. Secondary endpoints were 24-hour tramadol use, intraoperative remifentanil consumption, numeric rating scale scores at 1,4,8,12, and 24 hours, and side effects.

Results: No statistically significant difference was observed between groups in the time to first rescue analgesia; however, median times favored Group IF (5.3 hours vs. 4 hours). The median total tramadol consumption in the first 24 hours was 100 mg [interquartile range (IQR): 150 mg] in Group IV and 75 mg (IQR: 100 mg) in Group IF, with no significant difference between the groups (p=0.256). Numeric rating scale scores and remifentanil use were similar. Nausea was significantly more frequent in Group IF (41%) than in Group IV (18%) (p=0.036).

Conclusion: Both systemic and interfascial administration of dexamethasone, when combined with an aQLB in patients undergoing LC, provided comparable postoperative pain relief. However, systemic administration was associated with a lower incidence of postoperative nausea, suggesting it may be preferable for patients at higher risk of this side-effect.

Keywords: Adjuvant, dexamethasone, fascial plane blocks, postoperative analgesia, anterior quadratus lumborum block, laparoscopic cholecystectomy

INTRODUCTION

Laparoscopic cholecystectomy (LC) is widely recognized as one of the most commonly performed routine surgical procedures. Although LC typically causes less intense pain than open cholecystectomy, patients may still experience moderate to severe postoperative pain. This pain may arise from somatic sources at the trocar entry sites, from visceral sources related to gallbladder manipulation, or from carbon dioxide insufflation in the abdomen (1). Postoperative pain is often managed with opioid analgesics. However, opioid use is associated with undesirable side effects such as nausea, vomiting, sedation, itching, dependence, prolonged hospital stays, and increased healthcare costs.

Therefore, the use of truncal blocks as part of multimodal analgesia has increased in recent years (2,3). Anterior quadratus lumborum blocks (aQLB) are among the techniques used for postoperative analgesia after abdominal surgery. Studies have demonstrated effective analgesia in the T7-L2 dermatomal regions (4). Various adjuvants can be used in block applications to enhance the duration and efficacy of analgesia provided by local anesthetics.

ORCID IDs of the authors: S.Ş. 0000-0003-1471-1340; D.G.M. 0000-0002-4229-4903; V.D. 0000-0003-2718-5212

Corresponding Author: Serpil Şehirlioğlu, MD;
E-mail: drserpilsahin@gmail.com

Received Date: 14.07.2025 **Accepted Date:** 15.12.2025
Epub: 06.02.2026
Publication Date: 28.04.2026

J Acad Res Med 2026;16(1):10-16

Dexamethasone is a glucocorticoid known for its ability to prolong the effects of local anesthetics and suppress inflammation when used as an adjuvant (5,6).

Extensive research has been conducted on the perineural use of dexamethasone in extremity surgeries (6-8). Although dexamethasone is widely used, its perineural administration is considered off-label by both the European Medicines Agency and the U.S. Food and Drug Administration, which raises concerns among some clinicians. As a result, intravenous administration has become more common in recent years (9).

Conflicting results have been reported in the literature regarding the perineural versus systemic administration of dexamethasone for peripheral nerve blocks. In a meta-analysis of adjuvants for supraclavicular brachial plexus blocks, systemic administration of dexamethasone provided a longer duration of sensory blockade than perineural administration in upper extremity blocks (10). In the same study, a longer duration of analgesia was also observed in the perineural group; however, this difference was not statistically significant. Furthermore, in a study comparing plasma concentrations of dexamethasone after systemic and perineural administration during upper extremity surgeries, similar plasma levels were detected via both routes, which were attributed to the rich vascularity and close anatomical relationship between nerves and blood vessels, facilitating systemic absorption (11). Similarly, Desai et al. (12), in a meta-analysis of lower-extremity peripheral nerve blocks, reported that the duration of analgesia was longer with perineural than with systemic administration of dexamethasone. This was explained by lower vascularity in the lower-extremity relative to the upper extremity; however, the difference was not clinically significant.

No studies have directly compared the analgesic efficacy of systemic versus interfascial dexamethasone administration in fascial plane blocks. In studies investigating local anesthetic pharmacokinetics, it has been reported that ropivacaine reached high plasma concentrations shortly after administration in the erector spinae plane and serratus intercostal fascial plane blocks, a finding associated with rapid absorption (13). However, studies specifically addressing QLB remain limited. Given the absence of major vascular structures between fascial layers, systemic absorption of dexamethasone in these blocks may be slower than that in extremity blocks, potentially allowing interfascial administration to provide a longer duration of analgesia than the intravenous route.

This retrospective study compared the analgesic efficacy of interfascial versus systemic dexamethasone administration in patients who underwent preoperative aQLB for elective LC.

METHODS

Study Design and Ethical Approval

This retrospective, single-center study was conducted at University of Health Sciences Türkiye, Gaziosmanpaşa Training and Research Hospital. Ethical approval was granted by the Non-Interventional Research Ethics Committee of University of Health Sciences Türkiye, Gaziosmanpaşa Training and Research Hospital (approval no: 86, date: 18.06.2025). The ethics application specified that part of the data for the interfascial adjuvant group would be drawn from a previously approved clinical study at our institution, titled "Comparison of the Postoperative Analgesic Efficacy of Adjuvant Quadratus Lumborum Block in Laparoscopic Cholecystectomies" (approval no: 40, date: 10.05.2023). Permission was granted to use this data for secondary analysis.

After obtaining ethical clearance, a retrospective review was conducted of 84 patients who underwent elective LC between January 1, 2024, and June 1, 2025. Twelve patients were excluded due to incomplete follow-up documentation regarding postoperative analgesia or use of non-tramadol rescue analgesics.

Patient Selection

In this retrospective study, we reviewed the records of patients who underwent elective LC under general anesthesia and who received a preoperative bilateral aQLB. Among these patients, those who had received dexamethasone via either the interfascial or the systemic route were included in the study. Patients over 18 years of age and classified as American Society of Anesthesiologists (ASA) physical status I-II were included. Patients classified as ASA III-IV and those who had received rescue analgesics other than tramadol were excluded from the analysis.

Intervention and Group Descriptions

In our clinic, fascial plane blocks are generally performed to provide postoperative analgesia and reduce intraoperative opioid requirements following LC. In our clinic, aQLBs are routinely performed preoperatively and bilaterally under ultrasound guidance, with patients in the lateral decubitus position. A convex ultrasound probe (2-6 MHz; MyLabseven, Esaote Europe, Netherlands) and a 100-mm nerve block needle (Stimuplex® Ultra; B. Braun, Melsungen, Germany) were used. The standard local anesthetic solution consists of 20 mL of 0.25% bupivacaine, administered bilaterally. For interfascial dexamethasone administration, each side receives 20 mL comprising 10 mL of 0.5% bupivacaine, 9 mL of normal saline, and 1 mL (4 mg) of dexamethasone.

Participants were divided into two groups based on the method of adjuvant delivery.

Patients who received interfascial dexamethasone (8 mg) coadministered with the local anesthetic were included in Group IF, whereas those who received systemic dexamethasone (8 mg) intravenously during the intraoperative period were included in Group IV.

General Anesthesia and Postoperative Analgesia Protocol

In our clinic, a standard general anesthesia protocol is used for ASA I-II patients. Induction is performed by intravenous administration of midazolam (0.04 mg/kg), fentanyl (1 μg/kg),

J Acad Res Med 2026;16(1):10-16

lidocaine (0.5 mg/kg), propofol (2 mg/kg), and rocuronium (0.6 mg/kg). Anesthesia maintenance involves sevoflurane and a continuous infusion of remifentanil at a rate of 0.05-0.2 µg/kg/min, adjusted based on the patient's hemodynamic status. All surgical procedures are carried out using the standard four-port technique with a maximum intra-abdominal pressure of 12 mmHg.

In our clinic, it is routine practice to administer intravenous paracetamol (1 g), tramadol (1 mg/kg), and ondansetron (0.1 mg/kg) to patients at the end of surgery. Neuromuscular blockade is typically reversed with either sugammadex (2 mg/kg) or a combination of neostigmine (0.04 mg/kg) and atropine (0.02 mg/kg), at the discretion of the anesthesiologist. Postoperative pain management in the general surgery ward routinely consists of 1 g paracetamol every 6 hours, with 100 mg intravenous tramadol administered when the resting numeric rating scale (NRS) score is ≥4.

Data Collection

Data were retrospectively collected from patients' anesthesia and postoperative analgesia follow-up forms. Block procedures were recorded in the block registry located in the operating room block area, while postoperative analgesia data were gathered from dedicated follow-up forms maintained for each patient. In our clinic, all patients who receive truncal blocks are visited by an anesthesiologist at 1,4,8,12, and 24 hours postoperatively. The time to first rescue analgesia, total tramadol consumption, resting and dynamic (during coughing) NRS scores, and side effects such as nausea, vomiting, and shoulder pain are documented in the postoperative analgesia follow-up forms. Patients' demographic data and ASA scores are recorded in both the anesthesia and analgesia follow-up forms. Intraoperative remifentanil consumption was extracted from the anesthesia records.

Outcome Measures

The primary objective of this study was to evaluate the effect of interfascial versus systemic administration of dexamethasone on the time to first rescue analgesia in patients undergoing anterior QLB during elective LC. Secondary outcomes included total postoperative tramadol consumption, NRS scores at rest and during movement, intraoperative remifentanil use, and the incidence of adverse events such as nausea, vomiting, and shoulder pain.

Statistical Analysis

Descriptive statistics were presented as mean ± standard deviation, median [interquartile range (IQR)], frequency, and percentage, depending on variable type. Data distribution was assessed using the Kolmogorov-Smirnov test. For variables with a normal distribution, the independent samples t-test was used, while the Mann-Whitney U test was applied to non-normally distributed variables. The chi-square test was used to compare categorical data. Time to first rescue analgesia was analyzed with Kaplan-Meier survival curves. All statistical analyses were performed using SPSS version 28.0 (IBM Corp., Armonk, NY), and a p-value less than 0.05 was considered statistically significant.

RESULTS

A total of 72 patients were included in the analysis. The demographic characteristics of both groups were similar (Table 1).

No statistically significant difference was found between the groups regarding the time to first rescue analgesia (p=0.497; Table 2). The need for rescue analgesia was observed in 54.5% of patients in Group IV and 51.3% of patients in Group IF. Kaplan-Meier analysis was performed for patients who required rescue analgesia. Although the curves suggest that patients in Group IV tended to need rescue analgesia earlier, the log-rank test showed no statistically significant difference between the groups (p=0.463; Figure 1). The Kaplan-Meier curves also indicated that in Group IV, all patients had received rescue analgesia by the 10th postoperative hour, whereas in Group IF all patients had received it by the 13th postoperative hour.

During the first 24 hours postoperatively, the median total tramadol consumption was 100 mg (IQR: 150 mg) in Group IV and 75 mg (IQR: 100 mg) in Group IF, with no statistically significant difference between the groups (p=0.256; Table 2). Additionally, intraoperative remifentanil consumption did not differ significantly between the two groups (Table 2).

Resting NRS and dynamic NRS scores were similar across all evaluated time points (Table 3).

Analysis of side effects showed that nausea occurred in 18% (6 patients) of patients in Group IV and 41% (16 patients) of patients in Group IF. This difference was statistically significant: Group IF exhibited a higher rate of nausea (p=0.036; Table 4). However, there were no significant differences between groups regarding vomiting or shoulder pain (Table 4).

DISCUSSION

This study compared the analgesic effectiveness of interfascial and systemic dexamethasone in patients undergoing elective LC who received a preoperative aQLB for pain relief. The time to first rescue analgesia, total postoperative opioid use, intraoperative remifentanil consumption, and NRS scores were similar between the groups, but the incidence of nausea was significantly lower in the group that received systemic dexamethasone.

QLBs are used for postoperative pain control after surgeries involving the lower thoracic area, abdomen, retroperitoneal space, and inguinal region (14,15). In our clinic, the anterior approach to the QLB technique is commonly preferred for LC procedures. First described by Børglum et al. (16), the aQLB is believed to offer more effective pain relief, possibly because of the high density of mechanoreceptors in the anterior thoracolumbar fascia (4,17).

Previous studies have shown that dexamethasone, when used as an adjuvant in peripheral nerve blocks, may extend analgesic

J Acad Res Med 2026;16(1):10-16

duration and enhance postoperative pain control (6,9). However, research on the effectiveness of adjuvants in fascial plane blocks remains limited.

A randomized controlled trial evaluating the addition of dexamethasone to local anesthetics in aQLB for LC demonstrated a significantly prolonged time to first rescue analgesia compared with the saline control group (18). Arafa et al. (19) randomized pediatric patients undergoing renal surgery into three groups receiving interfascial dexamethasone, intravenous dexamethasone, or aQLB alone. Both dexamethasone groups had significantly lower total morphine consumption, longer time to first rescue analgesia, and lower pain scores than the aQLB-alone group. The time to first rescue analgesia did not differ between the two dexamethasone groups; however, total tramadol consumption was significantly lower in the interfascial dexamethasone group. Consistent with previous findings, our study found that the median total tramadol consumption within the first 24 hours was 100 mg (IQR: 150 mg) in Group IV and 75 mg (IQR: 100 mg) in Group IF, indicating lower consumption in the interfascial group. However, this difference was not statistically significant. Additionally, there was no significant difference in the time to first rescue analgesia between patients who received systemic dexamethasone and those who received interfascial dexamethasone as an adjuvant. The median times to first rescue analgesia were 4 hours in Group IV and 5.3 hours in Group IF, though this was not statistically significant. Kaplan-Meier analysis of the time to first rescue analgesia showed that patients in Group IV requested analgesia earlier than those in Group IF. Although not statistically significant, this finding suggests that interfascial administration of the adjuvant may prolong the time to request analgesia compared with systemic administration.

In this study, the need for rescue analgesia was similar between the groups, with approximately half of the patients requiring it in each arm. These relatively low rates were attributed to an effective multimodal analgesia protocol combined with the aQLB block. NRS scores remained consistently low, staying below 4 at all time points in both groups. Intraoperative remifentanil consumption was likewise comparable, in line with the observed postoperative analgesic outcomes.

In two separate meta-analyses comparing perineural and systemic dexamethasone administration in peripheral nerve blocks—one including 11 studies with a total of 914 patients and the other including 9 studies with a total of 801 patients—perineural administration was shown to prolong the duration of analgesia more effectively than systemic administration (7,8). These effects of dexamethasone on analgesic efficacy are believed to result from several mechanisms, including suppression of inflammatory

Table 1. Comparison of demographic data and operational durations among groups

	Group IV (n=33)		Group IF (n=39)		p*
	n	%	n	%	
Gender					
Female	26	78.8	33	84.6	0.522
Male	7	21.2	6	15.4	
	Mean ± SD		Mean ± SD		
Age (yr)	46.2±11.09		45.3±9.8		0.734
BMI (kg/m^2)	26.9±3.4		27.9±5.1		0.326
Operation time (min)	52.9±12.0		56.2±14.3		0.477

Values are presented as mean ± SD or number and percentage n (%)
Group IV: Intravenous dexamethasone group, Group IF: Interfascial dexamethasone group, BMI: Body mass index, SD: Standard deviation, *: p<0.05

Table 2. Comparison of intraoperative remifentanil consumption, total analgesic consumption, and time to first analgesic use between grup

	Group IV (n=33)	Group IF (n=39)	p*
	Median (IQR)	Median (IQR)	
Intraoperative remifentanil consumption (mcg/kg)	180 (125)	210 (160)	0.261
Total tramadol consumption (mg/day)	100 (150)	75 (100)	0.256
Time to first rescue analgesia (hour)	4.0 (5.5)	5.3 (6.9)	0.497
	n (%)	n (%)	
Patients requiring rescue analgesia	18 (54.5)	20 (51.3)	0.782
Patients not requiring rescue analgesia	15 (45.5)	19 (48.7)	

Values are presented as median (IQR)
Group IV: Intravenous dexamethasone group, Group IF: Interfascial dexamethasone group, IQR: Interquartile range, *: p<0.05

J Acad Res Med 2026;16(1):10-16

mediator release, inhibition of ectopic neuronal discharges, and blockade of potassium channel-mediated discharges in nociceptive C fibers (20-24).

In the experimental animal model study conducted by Matsuda et al. (25), co-administration of dexamethasone with local anesthetics significantly prolonged the duration of analgesia. This effect was observed exclusively with perineural administration, whereas the same dose administered intramuscularly (systemically) did not produce a comparable effect. Specifically, perineural dexamethasone extended the duration of sciatic nerve block with ropivacaine to more than 360 minutes, while systemic administration of dexamethasone failed to achieve this outcome. Histological analyses indicated that the effect was mediated through glucocorticoid receptor activation and was closely associated with the suppression of neuronal nitric oxide synthase (nNOS) expression in the dorsal root ganglion.

Furthermore, perineural administration of the NOS inhibitor L-NAME significantly prolonged block duration, supporting the involvement of peripheral mechanisms. These findings suggest that the analgesic-prolonging effect of dexamethasone is not solely attributable to systemic absorption but also to direct perineural mechanisms. In addition to suppressing NOS production, the vasoconstrictive properties of dexamethasone may further prolong the duration of action of the local anesthetic by delaying its systemic absorption.

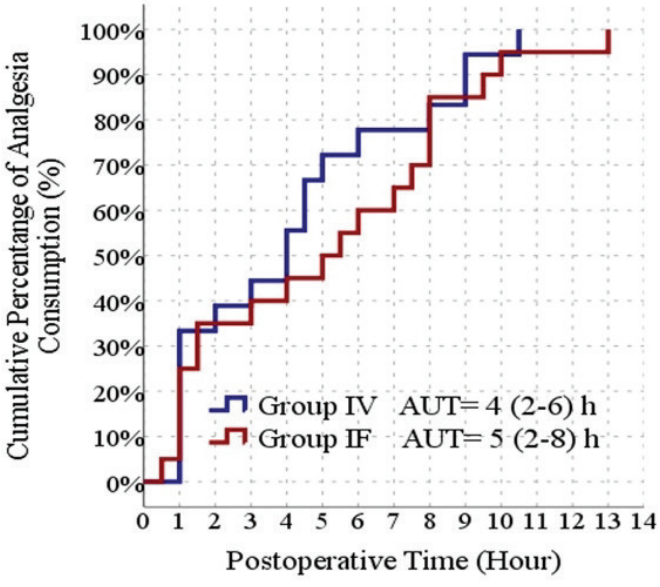

Figure 1. Comparison of time to first analgesic request between Group IV and Group IF: Kaplan-Meier analysis. Median time to first rescue analgesia was 4.0 hours (95% CI: 1.9-6.1) in Group IV and 5.0 hours (95% CI: 1.7-8.3) in Group IF (log-rank test, p=0.463)

Group IV: Intravenous dexamethasone group; Group IF: Interfascial dexamethasone group, CI: Confidence interval, AUT: Analgesic using time

Note: CI values in the figure are rounded to the nearest whole number for visual clarity

Moreover, in the management of rebound pain—a phenomenon that can occur after the resolution of a peripheral nerve block—single-dose dexamethasone administered with the block has been shown to reduce its incidence (26). In addition, intravenous dexamethasone has been reported to decrease rebound pain, demonstrating an effect comparable to that of perineural administration (27). Overall, these findings suggest that the analgesic effect of dexamethasone is multifactorial, with systemic absorption playing a role; however, experimental data highlight that perineural mechanisms, particularly nNOS suppression via glucocorticoid receptor activation, substantially contribute to prolonging block duration.

Additionally, it has been suggested that interfascial dexamethasone administered during anterior QLB can reach the paravertebral space via spread of the block, thereby exerting a central analgesic effect. This effect is thought to aid in reducing central sensitization by inhibiting nuclear factor kappaB (NF-κB) activity. As a transcription factor, NF-κB plays a key role in regulating inflammatory activity and has been associated with the development of chronic and pathological pain (28). However, we were unable to evaluate the long-term effects on chronic pain

Table 3. Comparison of rNRS and dNRS values between groups

	Group IV (n=33)	Group IF (n=39)	p*
	Median (IQR)	Median (IQR)	
rNRS 1. hour	2 (2.0)	2 (1.0)	0.287
rNRS 4. hour	1 (1.0)	1 (2.0)	0.744
rNRS 8. hour	1 (1.0)	1 (1.0)	0.760
rNRS 12. hour	1 (1.0)	1 (2.0)	0.908
rNRS 24. hour	1 (1.5)	1 (1.0)	0.621
dNRS 1. hour	3 (2.0)	3 (1.0)	0.369
dNRS 4. hour	3 (1.0)	3 (2.0)	0.535
dNRS 8. hour	3 (1.5)	2 (1.0)	0.852
dNRS 12. hour	1 (2.0)	2 (2.0)	0.327
dNRS 24. hour	1 (2.5)	2 (2.0)	0.839

Values are presented as median (IQR)
Group IV: Intravenous dexamethasone group; Group IF: Interfascial dexamethasone group, IQR: Interquartile range, rNRS: Resting numeric rating scale; dNRS: Dynamic numeric rating scale, *: p<0.05

Table 4. Comparison of postoperative side effects between the groups

	Group IV (n=33)		Group IF (n=39)		p
	n	%	n	%	
Nausea	6	18.2	16	41.0	**0.036***
Vomiting	2	6.1	3	7.7	0.786
Shoulder pain	7	21.2	12	30.8	0.359

Values are presented as number and percentage n (%)
Group IV: Intravenous dexamethasone group; Group IF: Interfascial dexamethasone group, *: p<0.05

J Acad Res Med 2026;16(1):10-16

because of the retrospective design and limited follow-up. Our analysis primarily focused on acute postoperative pain outcomes.

The use of intravenous dexamethasone for prophylaxis of nausea and vomiting is well established in the literature (29). It is also known to reduce surgical stress and inflammation, facilitate early patient mobilization, and, therefore, is included as standard practice in enhanced recovery after surgery protocols. In this study, patients who received intravenous dexamethasone showed a notably lower incidence of postoperative nausea than those who received interfascial dexamethasone, a finding attributed to its antiemetic effects. The incidence of vomiting was low and similar in both groups. The incidence of shoulder pain was higher in the IF group, although the difference was not statistically significant.

Study Limitations

A limitation of this study is that it was conducted at a single center. With a multicenter design and a larger sample size, more generalizable results could have been achieved. Another significant limitation is the retrospective design, which may have introduced bias despite inclusion of all eligible patients. Additionally, the relatively low volume of local anesthetic used for QLB might have affected block efficacy. The lack of standardized analgesic protocols, such as patient-controlled analgesia, is another methodological limitation. Chronic pain follow-up could not be conducted, limiting evaluation of long-term outcomes. Lastly, the absence of a control group represents another limitation, as only patients who received a block are systematically recorded in our analgesia follow-up forms.

CONCLUSION

This retrospective study showed that both interfascial (perineural) dexamethasone—administered as an adjuvant to the aQLB during LC—and systemic (intravenous) dexamethasone were similarly effective for postoperative analgesia. There were no significant differences between groups in total tramadol consumption or in time to first rescue analgesia; however, the interfascial dexamethasone group may offer a slight advantage in analgesic efficacy. However, further research with larger patient populations is needed to confirm this potential benefit. Notably, the incidence of postoperative nausea was significantly lower among patients who received systemic dexamethasone, suggesting that systemic administration may be preferable for patients at risk of postoperative nausea.

Ethics

Ethics Committee Approval: Ethical approval was granted by the Non-Interventional Research Ethics Committee of University of Health Sciences Türkiye, Gaziosmanpaşa Training and Research Hospital (approval no: 86, date: 18.06.2025).

Informed Consent: Retrospective study.

Footnotes

Author Contributions: Concept - S.Ş., D.G.M., V.D.; Design - S.Ş., D.G.M., V.D.; Data Collection and/or Processing - S.Ş., D.G.M., V.D.; Analysis and/or Interpretation - S.Ş., D.G.M., V.D.; Literature Search - S.Ş.; Writing - S.Ş.

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

Financial Disclosure: The authors report that no financial support was received for this study.

REFERENCES

1. Jiang B, Ye S. Pharmacotherapeutic pain management in patients undergoing laparoscopic cholecystectomy: a review. Adv Clin Exp Med. 2022; 31: 1275-88.
2. Feldheiser A, Aziz O, Baldini G, Cox BP, Fearon KC, Feldman LS, et al. Enhanced recovery after surgery (ERAS) for gastrointestinal surgery, part 2: consensus statement for anaesthesia practice. Acta Anaesthesiol Scand. 2016; 60: 289-334.
3. Bisgaard T. Analgesic treatment after laparoscopic cholecystectomy: a critical assessment of the evidence. Anesthesiology. 2006; 104: 835-46.
4. Elsharkawy H, El-Boghdadly K, Barrington M. Quadratus lumborum block: anatomical concepts, mechanisms, and techniques. Anesthesiology. 2019; 130: 322-35.
5. Pehora C, Pearson AM, Kaushal A, Crawford MW, Johnston B. Dexamethasone as an adjuvant to peripheral nerve block. Cochrane Database Syst Rev. 2017; 11: CD011770.
6. Zorrilla-Vaca A, Li J. Dexamethasone injected perineurally is more effective than administered ıntravenously for peripheral nerve blocks: a meta-analysis of randomized controlled trials. Clin J Pain. 2018; 34: 276-84.
7. Baeriswyl M, Kirkham KR, Jacot-Guillarmod A, Albrecht E. Efficacy of perineural vs systemic dexamethasone to prolong analgesia after peripheral nerve block: a systematic review and meta-analysis. Br J Anaesth. 2017; 119: 183-91.
8. Choi S, Rodseth R, McCartney CJ. Effects of dexamethasone as a local anaesthetic adjuvant for brachial plexus block: a systematic review and meta-analysis of randomized trials. Br J Anaesth. 2014; 112: 427-39.
9. Chong MA, Berbenetz NM, Lin C, Singh S. Perineural versus intravenous dexamethasone as an adjuvant for peripheral nerve blocks: a systematic review and meta-analysis. Reg Anesth Pain Med. 2017; 42: 319-26.
10. Sehmbi H, Brull R, Ceballos KR, Shah UJ, Martin J, Tobias A, et al. Perineural and intravenous dexamethasone and dexmedetomidine: network meta-analysis of adjunctive effects on supraclavicular brachial plexus block. Anaesthesia. 2021; 76: 974-90.
11. Bhatia N, Patial A, Ghai B, Dşpşka B, Attri SV, Gaba S, et al. European Society of Regional Anesthesia 19-0119. Comparison of systemic dexamethasone levels following its perineural versus intravenous administration: a randomized, double-blind study. Reg Anesth Pain Med. 2019; 44: A74-5.
12. Desai N, Pararajasingham S, Onwochei D, Albrecht E. Comparison of intravenous versus perineural dexamethasone as a local anaesthetic adjunct for peripheral nerve blocks in the lower limb: a meta-analysis and systematic review. Eur J Anaesthesiol. 2024; 41: 749-59.
13. Kadam VR, Ludbrook GL, Hewett P, Westley I. Plasma ropivacaine levels after ultrasound-guided erector spinae plane block and wound infiltration in laparoscopic colonic surgery - an observational study. Indian J Anaesth. 2022; 66: 231-2.
14. Kumar GD, Gnanasekar N, Kurhekar P, Prasad TK. A comparative study of transversus abdominis plane block versus quadratus lumborum block for postoperative analgesia following lower abdominal surgeries: a prospective double-blinded study. Anesth Essays Res. 2018; 12: 919-23.
15. Yousef NK. Quadratus lumborum block versus transversus abdominis plane block in patients undergoing total abdominal hysterectomy: a randomized prospective controlled trial. Anesth Essays Res. 2018; 12: 742-7.
16. Børglum J, Moriggl B, Jensen K, Lönnqvist PA, Christensen AF, Sauter A, et al. Ultrasound-guided transmuscular quadratus lumborum blockade. Br J Anaesth 2013; 111(Suppl).
17. Yetik F, Yilmaz C, Karasu D, Haliloğlu Dastan N, Dayioğlu M, Baytar Ç. Comparison of ultrasound-guided quadratus lumborum block-2 and quadratus lumborum block-3 for postoperative pain in cesarean section: a randomized clinical trial. Medicine (Baltimore). 2022; 101: e31844.
18. Mansour HS, Ali NS, Abdel Rahman MA. The effect of dexamethasone as an adjuvant in quadratus lumborum block to improves analgesia after laparoscopic cholecystectomy: controlled randomized study. Egypt J Anaesth. 2024; 40: 135-42.

J Acad Res Med 2026;16(1):10-16

19. Arafa SK, Elsayed AA, Hagras AM, Shama AAA. Pediatric postoperative pain control with quadratus lumborum block and dexamethasone in two routes with bupivacaine: a prospective randomized controlled clinical trial. Pain Physician. 2022; 25: E987-98.
20. De Bosscher K, Vanden Berghe W, Haegeman G. The interplay between the glucocorticoid receptor and nuclear factor-kappaB or activator protein-1: molecular mechanisms for gene repression. Endocr Rev. 2003; 24: 488-522.
21. Eker HE, Cok OY, Aribogan A, Arslan G. Management of neuropathic pain with methylprednisolone at the site of nerve injury. Pain Med. 2012; 13: 443-51.
22. Maagaard M, Albrecht E, Mathiesen O. Prolonging peripheral nerve block duration: current techniques and future perspectives. Acta Anaesthesiol Scand. 2025; 69: e70010.
23. An K, Elkassabany NM, Liu J. Dexamethasone as adjuvant to bupivacaine prolongs the duration of thermal antinociception and prevents bupivacaine-induced rebound hyperalgesia via regional mechanism in a mouse sciatic nerve block model. PLoS One. 2015; 10: e0123459.
24. Holmberg A, Hassellund SS, Draegni T, Nordby A, Ottesen FS, Gulestøl A, et al. Analgesic effect of intravenous dexamethasone after volar plate surgery for distal radius fracture with brachial plexus block anaesthesia: a prospective, double-blind randomised clinical trial. Anaesthesia. 2020; 75: 1448-60.
25. Matsuda K, Sasaki M, Baba H, Kamiya Y. Neuronal nitric oxide synthase suppression confers the prolonged analgesic effect of sciatic nerve block with perineural dexamethasone in postoperative pain model mice. J Pain. 2022; 23: 1765-78.
26. Yin W, Luo D, Mi H, Ren Z, Li L, Fan Z, et al. Rebound pain after peripheral nerve block: a review. Drugs. 2025; 85: 991-1002.
27. Singh NP, Makkar JK, Chawla JK, Sondekoppam RV, Singh PM. Prophylactic dexamethasone for rebound pain after peripheral nerve block in adult surgical patients: systematic review, meta-analysis, and trial sequential analysis of randomised controlled trials. Br J Anaesth. 2024; 132: 1112-21.
28. Niederberger E, Geisslinger G. The IKK-NF-kappaB pathway: a source for novel molecular drug targets in pain therapy? FASEB J. 2008; 22: 3432-42.
29. Lavand'homme P, Kehlet H. Benefits versus harm of intraoperative glucocorticoid for postoperative nausea and vomiting prophylaxis. Br J Anaesth. 2023; 131: 8-10.

DOI: 10.4274/jarem.galenos.2025.60252
J Acad Res Med 2026;16(1):17-23

Feasibility of the Speech Intelligibility Index in Turkish-speaking Adults with Sensorineural Hearing Loss

Atılım Atılgan[1], Murat Erinç[1], Cansu Tosyalı Salman[2], M. Tayyar Kalcıoğlu[3]

[1]İstanbul Medeniyet University Faculty of Health Sciences, Department of Audiology, İstanbul, Türkiye
[2]Göztepe Prof. Dr. Süleyman Yalçın City Hospital, Clinic of Audiology Unit, İstanbul, Türkiye
[3]İstanbul Medeniyet University Faculty of Medicine, Department of Otorhinolaryngology, İstanbul, Türkiye

Cite this article as: Atılgan A, Erinç M, Tosyalı Salman C, Kalcıoğlu MT. Feasibility of the speech intelligibility index in Turkish-speaking adults with sensorineural hearing loss. J Acad Res Med. 2026;16(1):17-23

ABSTRACT

Objective: The speech intelligibility index (SII) quantifies the proportion of speech that is audible to a listener. Its predictive accuracy is based on the acoustic and phonetic properties of the language it uses. Our study evaluated the feasibility of using the English-based SII for Turkish-speaking individuals with sensorineural hearing loss, comparing younger and older adults.

Methods: In this retrospective cross-sectional study, 161 participants (68 younger adults and 93 older adults) with bilateral sensorineural hearing loss were included. Hearing thresholds, word recognition scores (WRS), and SII values were analyzed. Non-parametric statistical analyses, including the Wilcoxon signed-rank test for paired comparisons, Mann-Whitney U test for group differences, Spearman's rank correlation for associations, and Fisher Z-test with bootstrap analysis to compare correlation coefficients were conducted to evaluate differences and relationships among audiological measures.

Results: Significant positive correlations were found between SII and WRS in both age groups (r=0.73 for younger adults; r=0.65-0.70 for older adults). Older adults had higher high-frequency thresholds and lower WRS ($p<0.001$), but no age-related differences were observed in SII or the pure-tone average.

Conclusion: The English-based SII can be used as a provisional tool for Turkish speakers, but age-related and frequency-specific variations highlight the need for a language-specific model.

Keywords: Speech intelligibility index, sensorineural hearing loss, age factors, Turkish language, word recognition, pure-tone average

INTRODUCTION

Speech intelligibility refers to a listener's ability to perceive speech under specific listening conditions and is often assessed as a measure of speech comprehension. It depends on the spectral and temporal characteristics of speech (1,2), the level of background noise (3), the dynamic range of the speech spectrum (4), the listener's hearing capacity (5), and the fundamental features of the target language (6). The speech intelligibility index (SII) quantifies weighted audibility across frequency bands at a given signal-to-noise ratio by using language specific band importance functions (BIF) and band audibility functions (7,8). Essentially, the amount of speech information accessible to the auditory system is determined by how much of the speech spectrum rises above the listener's hearing thresholds.

Language-specific differences determine the frequency weighting that forms the foundation of SII predictions. Languages vary in the extent to which they rely on particular bands, leading to distinct BIF profiles (6,9). For example, Korean places greater emphasis on lower bands than English and Mandarin do (10). Specifically for Turkish, agglutinative suffixation and vowel harmony increase the functional load of low-to-mid vowel energy, whereas English places greater weight on high-frequency consonantal cues (11-14). For instance, the high frequency /s/, which encodes plurality and possession in English, has no direct morphological counterpart in Turkish, which may shift band importance toward lower bands (10,15). This suggests that information may be concentrated in bands dominated by vowel energy rather than by high-frequency frication. Consistent with this linguistic profile, Turkish long-term average speech spectra (LTASS) indicate relatively elevated

ORCID IDs of the authors: A.A. 0000-0003-2944-6404; M.E. 0000-0003-2762-0173; C.T.S. 0000-0001-8574-8613; M.T.K. 0000-0002-6803-5467

Corresponding Author: Atılım Atılgan, Asst. Prof.;
E-mail: atilimatilgan@gmail.com
Received Date: 16.10.2025 **Accepted Date:** 26.12.2025
Epub: 05.02.2026
Publication Date: 28.04.2026

J Acad Res Med 2026;16(1):17-23

low-to-mid-frequency bands, providing an acoustic basis for reweighting band importance toward these frequencies (11). Taken together, these linguistic and acoustic features motivate a redistribution of band-importance toward lower frequencies and provide a principled rationale for deriving Turkish-specific BIFs.

The predictive validity of the SII is modulated by age-related factors. Although the SII demonstrates robust associations with speech recognition outcomes in both normal-hearing and hearing-impaired populations, prediction accuracy diminishes systematically in older listeners (16,17). This decline reflects age-related suprathreshold auditory deficits, particularly degraded temporal processing and difficulty understanding speech in noise (18,19). Notably, these age-dependent effects remain observable when language-specific weightings are applied, highlighting the need to account for both linguistic characteristics and age-related changes in auditory processing when making SII-based predictions (20).

In clinical practice, modern hearing aid analyzers compute SII with English-based parameters. Recent speech intelligibility studies have examined metrics for assessing communicative ability or simulated hearing loss in Turkish (21,22). However, these studies do not permit the development of a clinically usable model that predicts speech perception by estimating SII from audiometric data. Therefore, the aim of this preliminary study was to determine how the English-based SII differs among Turkish individuals and to explore its relationship with word recognition score (WRS), hearing thresholds, and pure tone averages (PTA) in Turkish-speaking older and younger adults.

METHODS

Participants

In this retrospective, cross-sectional study, data were obtained from the medical records of patients who visited the clinic between 2022 and 2023. Ethical approval for the study was obtained from the İstanbul Medeniyet University Göztepe Training and Research Hospital Clinical Research Ethics Committee (decision no: 2021/0596, date: 24.11.2021) before data collection.

A total of 202 patient records were initially reviewed. Of these, 41 records were excluded because they did not meet the inclusion criteria. The sample consisted of 161 individuals, including 80 females and 81 males, ranging in age from 16 to 95 years [mean age 63.5 years, standard deviation (SD)=18.6 years]. A total of 322 hearing test results (ears) were collected, including the pure tone average, speech recognition threshold (SRT), and WRS.

The participants were categorized into two age groups: younger adults (aged below 65 years) and older adults (aged 65 years and above). The younger adult group consisted of 68 participants (38 women, 30 men) with a mean age of 45.99 years (SD=14.63 years), while the older adult group included 93 participants (42 women, 51 men) with a mean age of 76.58 years (SD=6.84 years).

All participants were confirmed to have sensorineural hearing loss, with no individual exhibiting an air-bone gap of more than 10 decibel (dB) hearing level at any frequency. The inclusion criteria were as follows: absence of audiological findings suggestive of retrocochlear pathology, such as asymmetric hearing loss or disproportionately poor word-recognition scores relative to the degree of hearing loss; absence of conductive pathology; and hearing and behavioral thresholds obtained at all frequencies. Individuals who did not meet these criteria were excluded. Therefore, criterion-based convenience sampling was employed in this study.

Sample size calculation was performed using G*Power software to ensure sufficient statistical power for the analyses. The calculation was based on a two-tailed Fisher Z-test to compare two independent Pearson correlation coefficients, with an expected effect size q=0.6, an alpha error probability of 0.05, and a power of 0.95. The analysis indicated a required total sample size of 152 participants, with 76 per group. The actual sample size in this study (322 ears, corresponding to 161 participants) exceeds this requirement, ensuring robust statistical power for the planned analyses.

Procedures

For each ear of the patients, air conduction hearing thresholds at 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz, 6 kHz, and 8 kHz, four-frequency PTA (500 Hz-4 kHz), SRTs and WRS which are from using Turkish monosyllabic word recognition test (TMWRT) were obtained. TMWRT was a standardized, phonetically balanced single-syllable speech recognition test developed for Turkish (23).

To determine the SII scores, an Audioscan Verifit hearing aid analyzer was used. Air conduction hearing thresholds for each patient were entered into the software. Unaided SII values were calculated for each ear assuming a nominal speech input level of 65 dB sound pressure level (SPL). Calculations were performed using the Audioscan Verifit Speechmap module, which implements the ANSI/ASA S3.5-1997 1/3-octave band method with standard English BIF weights. The calculation excluded the 160 Hz band and did not incorporate masking effects. The SII scale ranges from 0.0 to 1.0.

Statistical Analysis

All statistical analyses were conducted using the R statistical software (version 4.3.1). Descriptive statistics provided an initial overview of data distribution. We assessed normality using the Shapiro-Wilk test; normality was confirmed only for hearing thresholds at 2,4,6, and 8 kHz. Therefore, non-parametric tests were used. Using the Wilcoxon signed-rank test, we compared, between ears, hearing thresholds at each frequency and PTA, WRS, SRT, and SII. Spearman's rank correlation was employed to assess the relationships between variables. The strengths of the correlation coefficients were interpreted according to Evans (24): $r<0.20$ as very weak, 0.20-0.39 as weak, 0.40-0.59 as moderate, 0.60–0.79 as strong, and $r\geq0.80$ as very strong. Mann-Whitney U tests were performed to examine differences in WRS, SRT, and PTA between younger and older adults. To investigate whether the correlation between the SII and WRS differs between younger and older adults and whether the relationships of both the SII and WRS with hearing thresholds across frequencies vary between

J Acad Res Med 2026;16(1):17-23

these age groups, Fisher Z-tests with bootstrap analysis were employed.

RESULTS

The study analyzed two groups of participants with hearing loss: younger adults (aged 32-64 years) and older adults (aged 65-95 years). The variables examined included hearing thresholds, PTA, SRTs, WRSs, and SII values. Analyses were conducted at the ear level. To assess the assumption of independence, we first tested for systematic right-left differences and found no statistically significant differences for SII, PTA, or WRS between ears ($p>0.05$). We then analyzed within-subject, ear-to-ear correlations and observed very strong correlations for SII, PTA, and WRS, with correlation coefficients (r) ranging from 0.85 to 0.95.

Hearing Thresholds Across Age Groups

Table 1 provides detailed descriptive statistics and compares hearing thresholds between groups across frequencies. Mann-Whitney U tests revealed significant age-related differences in the high-frequency range (4000-8000 Hz) in both ears ($p<0.05$), with increased thresholds in older adults. For example, at 8000 Hz, the threshold value was approximately 14 dB higher in the right ear ($p<0.001$) and 15 dB higher in the left ear ($p<0.001$), compared to the younger adult. In Figure 1, the mean hearing threshold values, with SDs, are plotted for both ears across frequency levels for younger and older adults.

PTA, SRT, and SII

The descriptive statistics for the audiological variables, including PTA, SRT, WRS, and SII values, are summarized in Table 2. The data are presented separately for younger and older adults, as well as for the right and left ears.

Comparison of Audiological Measures Between Ears and Age Groups

The Wilcoxon signed-rank test showed no difference between the right and left ears for hearing thresholds and PTA, WRS, SRT, and SII values ($p>0.05$). Mann-Whitney U tests were performed to examine differences in WRS, SRT, PTA, and SII between younger and older adults. WRS and SRT results revealed statistically significant differences. Younger adults demonstrated higher WRS (right ear: $p<0.001$; left ear: $p<0.001$) and lower SRT (right ear: $p<0.001$; left ear: $p<0.001$) than older adults. However, no significant differences were observed between younger and older adults in PTA or SII measures ($p>0.05$).

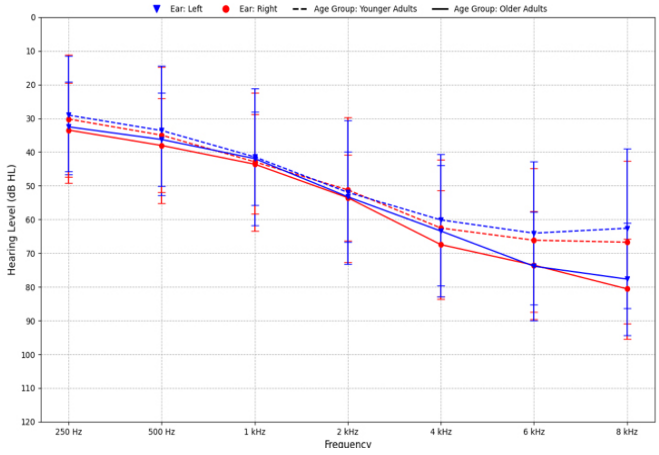

Figure 1. Mean and standard deviation of hearing thresholds by frequency and age group

Table 1. Hearing thresholds across frequencies: descriptive statistics and inter-group comparison

Frequency (Hz) Mean		Younger adults (n=68)						Older adults (n=93)						Between-group difference
		Median	SD	Min-max	95% CI lower	95% CI upper	Mean	Median	SD	Min-max	95% CI lower	95% CI upper	p-value	
250 Hz	R	30.22	25	19.07	5-90	25.51	34.99	33.49	30	13.92	10-85	30.75	36.50	0.037
	L	29.06	25	17.56	0-80	24.85	33.67	32.53	30	13.30	5-75	29.94	35.26	0.050
500 Hz	R	35.00	30	20.26	10-100	30.20	40	38.06	35	13.91	15-100	35.32	41.07	0.043
	L	33.62	30	19.15	5-90	29.77	38.67	36.29	35	13.83	10-70	33.38	39.08	0.073
1000 Hz	R	42.90	40	20.44	5-100	37.72	47.35	43.60	45	14.73	15-95	40.53	46.77	0.411
	L	41.45	40	20.24	10-105	36.98	46.61	41.94	40	13.85	15-80	39.14	44.73	0.379
2000 Hz	R	51.23	50	21.48	10-115	45.88	56.32	53.55	55	12.80	20-80	50.91	56.29	0.179
	L	51.96	50	21.30	10-115	47.79	57.94	53.28	55	13.40	25-90	50.43	56.02	0.236
4000 Hz	R	62.54	60	20.34	20-120	57.50	67.79	67.47	65	16.13	30-110	64.03	70.85	0.048
	L	60.14	60	19.52	15-120	56.17	65.43	60.14	60	19.52	15-120	63.44	69.30	0.005
6000 Hz	R	66.16	65	21.25	20-120	60.73	70.58	73.55	75	16.01	25-110	70.05	76.61	0.006
	L	64.06	65	21.15	15-120	60.14	69.70	73.87	75	16.12	35-110	70.64	77.25	0.002
8000 Hz	R	66.74	70	24.16	5-120	60.59	71.90	80.59	80	14.80	45-115	77.41	83.54	<0.001
	L	62.61	65	23.68	5-120	57.42	68.30	77.69	75	16.63	30-120	74.40	81.18	<0.001

R: Right ear, L: Left ear, Mean: Average, Median: Middle value, SD: Standard deviation, Min-max: Minimum-maximum values, CI: Confidence interval, p-value: Significance level

J Acad Res Med 2026;16(1):17-23

Table 2. Descriptive statistics of PTA, SRT, WRS, and SII values in younger and older adults

R		PTA		SRT		WRS		SII	
		L	R	L	R	L	R	L	
Younger adults	Mean	47.72	47.24	44.93	45.59	70.35	70.50	0.38	0.38
	Median	45.00	43.75	40.00	40.00	72.00	76.00	0.38	0.39
	SD	17.25	16.74	16.63	15.75	20.19	18.71	0.26	0.24
	Min-max	23.75-100	23.75-105.00	25.00-95.00	25.00-95.00	8.00-100.00	16.00-96.00	0.00-0.82	0.00-0.79
Older adults	Mean	50.67	49.45	52.37	50.91	59.53	60.13	0.30	0.31
	Median	51.25	48.75	50.00	50.00	60.00	64.00	0.27	0.31
	SD	11.93	10.95	13.68	13.12	17.70	18.20	0.20	0.19
	Min-max	27.50-87.50	27.50-80.00	25.00-95.00	20.00-85.00	0.00-92.00	0.00-92.00	0.00-0.73	0.00-0.72

R: Right ear, L: Left ear, Mean: Average, Median: Middle value, SD: Standard deviation, Min-max: Minimum-maximum values, PTA: Pure-tone average, SRT: Speech reception threshold, WRS: Word recognition score, SII: Speech intelligibility index

Correlations Between PTA, WRS, and SII

Spearman's rank correlation was used to examine the relationship between WRS and SII for both younger and older adults. A strong positive correlation was observed between WRS and SII in younger adults for the right ear (r=0.73, p<0.001) and the left ear (r=0.73, p<0.001). Similarly, older adults showed strong positive correlations for the right ear (r=0.65, p<0.001) and the left ear (r=0.70, p<0.001). The relationship between PTA and SII was also examined. In younger adults, very strong, statistically significant negative correlations between PTA and SII were observed in both the right (r=-0.95, p<0.001) and left (r=-0.95, p<0.001) ears. Older adults exhibited very strong negative correlations in both the right (r=-0.93, p<0.001) and left (r=-0.94, p<0.001) ears. Scatterplots with locally estimated scatterplot smoothing curves visually confirmed trends in the relationships between the SII and the WRS, and between the SII and the PTA. The relationship between SII and WRS (Figure 2) showed a positive trend, suggesting that higher SII values correspond to improved word recognition performance in both younger and older adults. In contrast, the relationship between SII and PTA (Figure 3) exhibited a negative, indicating that increased PTA values were associated with reduced SII.

Frequency-specific Correlations

Table 3 summarizes Spearman's rank-order correlation coefficients (r) between hearing thresholds at various frequencies and WRS and SII in both younger and older adults. The analysis revealed significant correlations across all examined frequencies, highlighting age-dependent differences in these relationships. For younger adults, hearing thresholds showed very strong negative correlations with WRS and SII, particularly at lower frequencies (e.g., 1000 Hz: r=-0.914 for WRS-left and r=-0.89 for SII-right; p<0.001). Similarly, older adults exhibited significant correlations,

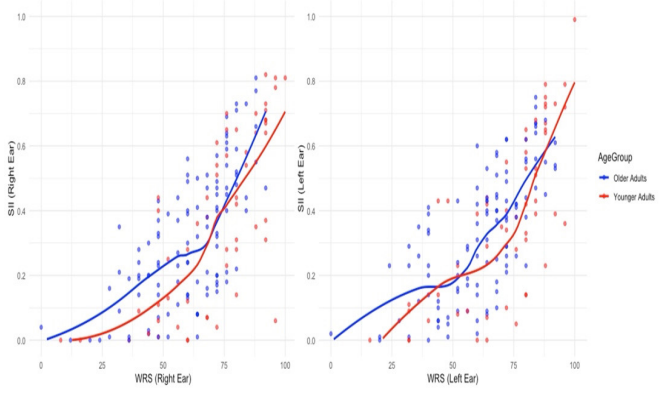

Figure 2. Relationship between SII and WRS with LOESS curves
SII: Speech intelligibility index, WRS: Word recognition score, LOESS: Locally estimated scatterplot smoothing

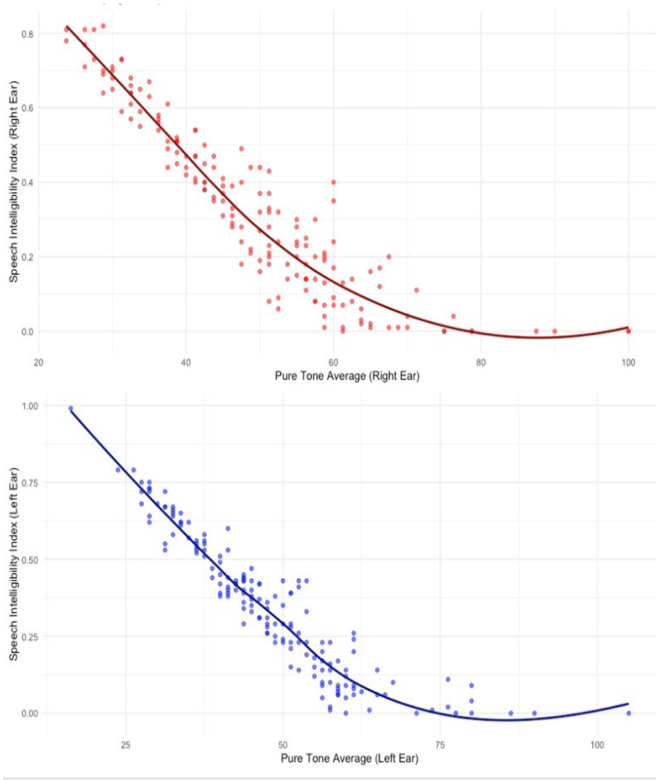

Figure 3. Relationship between SII and PTA with LOESS curves
SII: Speech intelligibility index, PTA: Pure tone averages, LOESS: Locally estimated scatterplot smoothing

J Acad Res Med 2026;16(1):17-23

Table 3. Spearman's rank-order correlations (r) between hearing thresholds at various frequencies and speech understanding measures

	Younger adults								Older adults							
	WRS				SII				WRS				SII			
	Right		Left		Right		Left		Right		Left		Right		Left	
	(r)	(p)	(r)	(p)	(r)	(p)	(r)	(p)	(r)	(p)	(r)	(p)	(r)	(p)	(r)	(p)
250 Hz	-0.699	<0.001	-0.694	<0.001	-0.65	<0.001	-0.64	<0.001	-0.75	<0.001	-0.7	<0.001	-0.7	<0.001	-0.68	<0.001
500 Hz	-0.780	<0.001	-0.755	<0.001	-0.74	<0.001	-0.72	<0.001	-0.919	<0.001	-0.87	<0.001	-0.88	<0.001	-0.82	<0.001
1000 Hz	-0.920	<0.001	-0.914	<0.001	-0.89	<0.001	-0.88	<0.001	-0.921	<0.001	-0.938	<0.001	-0.89	<0.001	-0.92	<0.001
2000 Hz	-0.866	<0.001	-0.853	<0.001	-0.84	<0.001	-0.81	<0.001	-0.825	<0.001	-0.776	<0.001	-0.8	<0.001	-0.77	<0.001
4000 Hz	-0.514	<0.001	-0.592	<0.001	-0.47	<0.001	-0.56	<0.001	-0.486	<0.001	-0.367	<0.001	-0.45	<0.001	-0.35	<0.001
6000 Hz	-0.431	<0.001	-0.498	<0.001	-0.42	<0.001	-0.45	<0.001	-0.268	0.009	-0.313	0.002	-0.25	0.013	-0.29	0.009
8000 Hz	-0.406	<0.001	-0.439	<0.001	-0.39	<0.001	-0.42	<0.001	-0.098	0.348	-0.231	0.026	-0.1	0.340	-0.2	0.045

WRS: Word recognition score, SII: Speech intelligibility index, r: Spearman's correlation coefficient, Right: Right ear, Left: Left ear, p: p-value

with notable differences at higher frequencies where correlations were generally weaker (e.g., at 8000 Hz: r=-0.406 for WRS-right in younger adults versus r=-0.098 in older adults, p<0.05). Figure 4 illustrates the frequency-specific Spearman correlations for WRS and the SII across younger and older adults.

Comparison of Correlation Strength Between Younger and Older Adults

The Fisher Z-test was conducted to compare the strength of correlations between the SII and WRS across younger and older adults, examining whether the relationship differed significantly between the two age groups. Because of the non-normal distribution of our data, a bootstrap analysis was also performed to estimate non-parametric confidence intervals (CI) for differences between correlations. The results revealed no statistically significant differences in the correlation between the SII and WRS for younger and older adults in either the right ear (Z=0.86, p>0.05) or the left ear (Z=0.30, p>0.05). When comparing WRS and SII between younger and older adults across frequency-specific hearing thresholds, the Fisher Z-test identified significant differences at 500 Hz for both WRS (Z=3.3, p=0.005) and SII (Z=2.613, p=0.045). However, the bootstrap analysis for this frequency did not support this finding, as the 95% CI included zero (CI: -0.2528 to 0.3097). For the remaining frequencies, the differences between the groups were also not statistically significant (p>0.05).

DISCUSSION

In this study, we examined the feasibility of using the English-based SII among Turkish-speaking individuals with sensorineural hearing loss the results showed significant correlations between SII and WRS, and between SII and PTA and hearing thresholds at individual frequencies, in both younger and older adults.

Hearing Loss and Speech Intelligibility

A strong negative correlation between SII and PTA was observed. Correlations between the thresholds (250-2000 Hz) and both WRS and SII were particularly high (>0.75) but decreased with

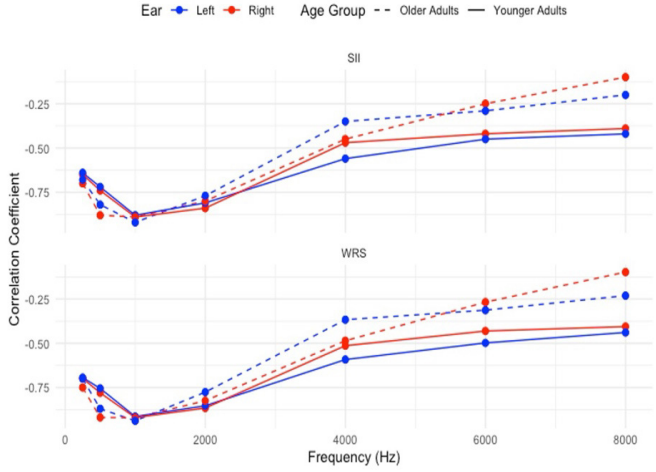

Figure 4. Frequency-specific spearman correlations of WRS and SII

WRS: Word recognition score, SII: Speech intelligibility index

increasing frequency. This is consistent with studies highlighting the critical role of this frequency range in SII and the association between this frequency range and elevated BIFs (15,25,26). Notably, the impact of language-specific characteristics on results is highlighted by differences in BIFs across languages (6,9,27). Similar to the differences observed in BIF, LTASS studies have revealed language-specific features, suggesting that frequency distributions may be influenced by the language's phonetic characteristics (11,28). These linguistic characteristics indicate that Turkish relies more heavily on low- to mid-frequency bands and less on high-frequency fricative information (e.g., /s/-like sounds). This interpretation is consistent with the frequency-specific patterns observed in our data.

Word Recognition and Speech Intelligibility

The findings of our study indicate that the English-based SII, despite lacking Turkish-specific BIF, retains robust sensitivity as a general measure of audibility. A strong positive correlation between SII and Turkish WRS in younger and older adults is

consistent with previous studies confirming the predictive power of SII for speech recognition in both normal-hearing and hearing-impaired individuals (13,16,29). We also found strong relationships between WRS and frequency-specific thresholds, which suggest that SII derived from English may provide clinically meaningful estimates in the Turkish-speaking population. Although this confirms its utility as a functional baseline, the generalizability of English-based standards remains questionable. The concentration of information in the lower band in Turkish, driven by its LTASS and agglutinative structure, suggests frequency-dependent differences in SII-performance predictions compared with English. Factors such as Turkish vowel harmony, agglutinative structure, vowel-to-consonant frequency ratios, and phonetic characteristics may contribute to variations in SII values across specific frequency bands (11). Studies in other languages have shown that language-specific SII adaptations improve predictive accuracy (6,20,27). Although the English-based SII provides a functional baseline, we believe that developing Turkish BIFs would be a logical next step to enhance the precision of this metric within clinical protocols.

Effects of Age

Our results indicate that older adults exhibited significantly higher SRTs and significantly lower WRSs than younger adults. These findings are consistent with previous studies showing that progressive hearing loss and age-related changes in cognitive processes negatively affect speech-recognition performance (16-18). However, no significant differences in the PTA or the SII were observed between younger and older adults. One possible explanation for the lack of significant differences between the two groups is the homogeneity of participants' hearing loss and the reliance of PTA and SII results on average hearing thresholds. PTA and SII mainly reflect audibility; thus, compensatory effects across frequencies might mask the age-related differences in these summary measures. While the SII effectively quantifies audibility for speech intelligibility prediction, it does not fully account for suprathreshold deficits associated with aging, such as reduced temporal resolution and impaired speech-in-noise performance (17,30,31). A high SII value may therefore fail to capture these limitations. Consequently, the SII should be regarded primarily as an audibility index rather than a complete measure of intelligibility.

In both younger and older adults, strong positive correlations were observed between WRS and SII, and strong negative correlations between PTA and SII. However, the Fisher Z-test showed no statistically significant differences between the correlation coefficients of these measures specifically between PTA and either WRS or SII across the groups at any tested frequency. Although not statistically significant, the correlation coefficient for 500 Hz hearing thresholds was higher in older adults. However, this trend decreased with increasing frequency. These observations indicate that age-related changes in hearing, particularly in critical frequency regions, may contribute to reduced predictive accuracy of the SII in older adults. Previous literature suggests that the predictive accuracy of the SII may deteriorate with age (20,32), especially under challenging listening conditions (18). Future studies should consider the impact of these age-related differences on the development of Turkish-derived SII models.

Clinical Implications and Future Directions

The studies of English-based SII conducted with non-English-speaking participants are limited. One study by Figueiredo et al. (33) compared unaided and aided SII values in Portuguese-speaking children with respect to the degree and configuration of hearing loss. They found that the average thresholds at 2000 Hz and 500 Hz were significant predictors of SII 65 values. In another study, the same researchers demonstrated how SII values varied at different input levels in a similar sample (34). Although SII is an important indicator based on sample data, the fact that they did not evaluate its relationship with speech perception creates uncertainty about the extent to which the results might differ for this Portuguese-speaking population. Nigri and Iório (32) assessed the relationship between verified aided SII, based on the desired sensation level (DSL) formula, and WRS at 65 dB SPL in Portuguese elderly individuals over 60 years of age and found a weak linear correlation. In contrast to our study, they found a weak relationship between aided SII and WRS. A possible explanation is that using the DSL v5 prescriptive method uniformly for all participants and matching amplification targets to real-ear measurements likely increased SII scores significantly. Thus, the SII scores became more similar across participants, regardless of individual differences in speech recognition. This produced a ceiling effect, thereby attenuating the observed relationship between SII and WRS. None of these studies discussed the linguistic dependency of SII and its potential impacts on results.

The integration of English-based SII values into audiological assessment and hearing aid analysis tools raises questions about their practical utility for non-English-speaking populations. Our study has demonstrated that these values can provide significant predictive utility for Turkish-speaking individuals. Nevertheless, the age-related differences in correlation coefficients at specific frequencies suggest that a linguistically adapted model could offer even greater predictive accuracy, thereby providing a strong rationale for establishing a Turkish-specific standard. Future research should therefore aim to develop a Turkish-specific model, primarily by deriving key language-specific parameters, such as BIF and its associated transfer function.

Study Limitations

This study has several limitations. First, English-based SII calculations were applied to Turkish participants, thereby ignoring Turkish-specific differences that may affect the SII. However, the aim of this preliminary study was to discuss situations that may arise from this discrepancy and help establish SII standards. Second, the use of a 65-year cut-off, though common in audiological research, may introduce bias because of unequal group sizes and heterogeneous aging trajectories. Third, SII values were obtained at a fixed speech level (65 dB SPL), limiting the generalizability to real-world listening conditions. Additionally, due to the retrospective nature of the

J Acad Res Med 2026;16(1):17-23

study design, heterogeneity among the obtained results may affect the outcomes despite the inclusion criteria we used. This should be taken into account. Lastly, analyses were conducted at the ear level to preserve clinically meaningful ear-specific variability. However, we acknowledge that ear level analyses may introduce within subject non-independence that can modestly reduce the effective sample size; therefore, p-values and CIs should be interpreted with this consideration in mind.

CONCLUSION

The SII, which is based on English, demonstrates clear, provisional usefulness for Turkish-speaking individuals and aligns well with WRS. However, reduced predictive accuracy among older adults and frequency-specific differences support the development of a Turkish-specific SII. Clinically, the current SII can be used for counseling and verification; however, caution is warranted in older adults and in high-frequency-weighted contexts.

Ethics

Ethics Committee Approval: Ethical approval for the study was obtained from the İstanbul Medeniyet University Göztepe Training and Research Hospital Clinical Research Ethics Committee (decision no: 2021/0596, date: 24.11.2021) before data collection.

Informed Consent: Retrospective study.

Footnotes

Author Contributions: Concept - A.A., M.E., M.T.K.; Design - A.A., M.E., M.T.K.; Data Collection and/or Processing - A.A., M.E., C.T.S.; Analysis and/or Interpretation - A.A., M.E., C.T.S., M.T.K.; Literature Search - A.A., M.E., C.T.S.; Writing - A.A.

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

Financial Disclosure: The authors report that no financial support was received for this study.

REFERENCES

1. Bernstein JG, Mehraei G, Shamma S, Gallun FJ, Theodoroff SM, Leek MR. Spectrotemporal modulation sensitivity as a predictor of speech intelligibility for hearing-impaired listeners. J Am Acad Audiol. 2013; 24: 293-306.
2. Monahan PJ, Schertz J, Fu Z, Pérez A. Unified coding of spectral and temporal phonetic cues: electrophysiological evidence for abstract phonological features. J Cogn Neurosci. 2022; 34: 618-38.
3. Amlani AM, Punch JL, Ching TY. Methods and applications of the audibility index in hearing aid selection and fitting. Trends Amplif. 2002; 6: 81-129.
4. Kringlebotn M. A graphical method for calculating the speech intelligibility index and measuring hearing disability from audiograms. Scand Audiol. 1999; 28: 151-60.
5. Souza PE, Yueh B, Sarubbi M, Loovis CF. Fitting hearing aids with the articulation index: impact on hearing aid effectiveness. J Rehabil Res Dev. 2000; 37: 473-81.
6. Wong LL, Ho AH, Chua EW, Soli SD. Development of the cantonese speech intelligibility index. J Acoust Soc Am. 2007; 121: 2350-61.
7. Hornsby BWY. The speech intelligibility index: what is it and what's it good for? Hear J. 2004; 57: 10.
8. Taghavi SMR, Mohammadkhani G, Jalilvand H. Speech intelligibility index: a literature review. یسانشی‌یاونش [Internet]. 2022 Jun 25 [cited 2024 Aug 31]; Available from: https://publish.kne-publishing.com/index.php/AVR/article/view/9861
9. Narne V, Tiwari N. Cross-language comparison of long-term average speech spectrum and dynamic range for three Indian languages and British English. Clin Arch Commun Disord. 2021; 6: 127-34.
10. Jin IK, Kates JM, Arehart KH. Does language matter when using a graphical method for calculating the speech intelligibility index? J Am Acad Audiol. 2017; 28: 119-26.
11. Yüksel M, Gündüz B. Long term average speech spectra of Turkish. Logoped Phoniatr Vocol. 2018; 43: 101-5.
12. Arik E. An experimental study of Turkish vowel harmony. Pozn Stud Contemp Linguist. 2015; 51: 359-74.
13. Studebaker GA, Sherbecoe RL. Frequency-importance and transfer functions for recorded CID W-22 word lists. J Speech Hear Res. 1991; 34: 427-38.
14. Oh YM, Pellegrino F, Coupé C, Marsico E. Cross-language comparison of functional load for vowels, consonants, and tones. In: Interspeech 2013 [Internet]. ISCA: ISCA; 2013. Available from: http://dx.doi.org/10.21437/interspeech.2013-662
15. Studebaker GA, Sherbecoe RL, Gilmore C. Frequency-importance and transfer functions for the Auditec of St. Louis recordings of the NU-6 word test. J Speech Hear Res. 1993; 36: 799-807.
16. Sherbecoe RL, Studebaker GA. Audibility-index predictions of normal-hearing and hearing-impaired listeners' performance on the connected speech test. Ear Hear. 2003; 24: 71-88.
17. Schum DJ, Matthews LJ, Lee FS. Actual and predicted word-recognition performance of elderly hearing-impaired listeners. J Speech Hear Res. 1991; 34: 636-42.
18. Souza PE, Boike KT, Witherell K, Tremblay K. Prediction of speech recognition from audibility in older listeners with hearing loss: effects of age, amplification, and background noise. J Am Acad Audiol. 2007; 18: 54-65.
19. Tremblay KL, Piskosz M, Souza P. Effects of age and age-related hearing loss on the neural representation of speech cues. Clin Neurophysiol. 2003; 114: 1332-43.
20. 20. Zeng J, Peng J, Xiang S. Chinese speech intelligibility and speech intelligibility index for the elderly. Speech Commun. 2024; 160: 103072.
21. Ariöz U, Günel B. Evaluation of hearing loss simulation using a speech intelligibility index. Turk J Elec Eng Comp Sci. 2016; 24: 4193-207.
22. Palaz H, Bicil Y, Kanak A, Doğan MU. New Turkish intelligibility test for assessing speech communication systems. Speech Commun. 2005; 47: 411-23.
23. Mungan Durankaya S, Serbetcioglu B, Dalkılıç G, Gürkan S, Kirkim G. Development of a Turkish monosyllabic word recognition test for adults. J Int Adv Otol. 2014; 10: 172-80.
24. Evans JD. Straightforward statistics for the behavioral sciences. Florence, AL: Duxbury Press; 1995. p.624.
25. Pavlovic CV. Derivation of primary parameters and procedures for use in speech intelligibility predictions. J Acoust Soc Am. 1987; 82: 413-22.
26. French NR, Steinberg JC. Factors governing the intelligibility of speech sounds. J Acoust Soc Am. 1947; 19: 90-119.
27. Jin IK, Lee J, Lee K, Kim J, Kim D, Sohn J, et al. The band-importance function for the Korean standard sentence lists for adults. J Audiol Otol. 2016; 20: 80-4.
28. Byrne D, Dillon H, Tran K, Arlinger S, Wilbraham K, Cox R, et al. An international comparison of long-term average speech spectra. J Acoust Soc Am. 1994; 96: 2108-20.
29. Humes LE, Kidd GR. Speech recognition for multiple bands: Implications for the Speech Intelligibility Index. J Acoust Soc Am. 2016; 140: 2019.
30. Cox RM, Alexander GC, Rivera IM. Comparison of objective and subjective measures of speech intelligibility in elderly hearing-impaired listeners. J Speech Hear Res. 1991; 34: 904-15.
31. Hargus SE, Gordon-Salant S. Accuracy of speech intelligibility index predictions for noise-masked young listeners with normal hearing and for elderly listeners with hearing impairment. J Speech Hear Res. 1995; 38: 234-43.
32. Nigri LF, Iório MCM. Study of correlation between speech intelligibility index (SII) and speech recognition percentage index. Distúrb Comun. 2019; 31: 33-43.
33. Figueiredo RS, Mendes B, Cavanaugh MC, Novaes B. Hearing loss classification by degree, configuration and relationship with amplified speech intelligibility index (SII). Codas. 2016; 28: 687-96.
34. Figueiredo R de SL, Mendes B, Cavanaugh MCV, Deperon TM, Novaes B. Speech intelligibility index (SII) and variation of speech signal intensity in hearing impaired children. Audiol, Commun Res [Internet]. 2019 Jan 31 [cited 2022 Apr 9]. Available from: https://www.scielo.br/j/acr/a/SFdDBcmkvcshpgJm3Ch7JpK/?lang=pt

DOI: 10.4274/jarem.galenos.2026.54366
J Acad Res Med 2026;16(1):24-28

Evaluation of the Association Between Preeclampsia Severity and Hematological and Biochemical Parameters

Havva Betül Bacak, Enes Serhat Coşkun, Süleyman Salman, Serkan Kumbasar, Batuhan Bulut

University of Health Sciences Türkiye, Gaziosmanpaşa Training and Research Hospital, Clinic of Obstetrics and Gynecology, İstanbul, Türkiye

Cite this article as: Bacak HB, Coşkun ES, Salman S, Kumbasar S, Bulut B. Evaluation of the association between preeclampsia severity and hematological and biochemical parameters. J Acad Res Med. 2026;16(1):24-28

ABSTRACT

Objective: Preeclampsia is a multisystemic disorder characterized by inflammation, platelet dysfunction, and target-organ involvement. Early identification of disease severity is crucial for improving maternal and fetal outcomes. This study aimed to evaluate the association between preeclampsia severity and hematological, biochemical, and derived inflammatory parameters.

Methods: This retrospective study included 135 pregnant women diagnosed with preeclampsia between 2020 and 2025 at a single tertiary center. Patients were classified as having mild or severe preeclampsia based on American College of Obstetricians and Gynecologists criteria. Hematological markers [hemoglobin, white blood cell (WBC), platelet, mean platelet volume (MPV)], biochemical parameters [aspartate aminotransferase (AST), alanine aminotransferase (ALT), creatinine], and derived indices [neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), platelet-to-neutrophil ratio (PNR)] were recorded. Group comparisons were performed using Student's t-test, Mann-Whitney U test, or Fisher's exact test as appropriate. Statistical significance was set at $p<0.05$.

Results: Severe preeclampsia was associated with significantly higher hemoglobin ($p=0.019$), WBC ($p=0.005$), neutrophil counts ($p=0.004$), and AST, ALT, and creatinine levels ($p=0.003$, $p<0.001$, $p=0.002$, respectively). Platelet counts were lower in severe cases ($p=0.0411$). No significant differences were found in MPV ($p=0.833$), NLR ($p=0.614$), or PLR ($p=0.109$). PNR was significantly higher in the mild group ($p=0.001$), whereas NLR and PLR did not differ significantly.

Conclusion: Severe preeclampsia is characterized by intensified inflammation, increased platelet consumption, and more pronounced organ involvement. Among the inflammatory indices, PNR differed significantly between severe and mild disease groups, whereas NLR and PLR did not differ significantly. Combined assessment of hematological and biochemical markers may provide a more comprehensive and reliable approach for evaluating preeclampsia severity. PNR may represent a potential supportive parameter associated with disease severity; however, further studies, including diagnostic performance analyses, are required.

Keywords: Preeclampsia, disease severity, platelet-to-neutrophil ratio, inflammatory markers

INTRODUCTION

Preeclampsia is a multisystemic and heterogeneous syndrome affecting approximately 3-5% of pregnancies and is one of the leading causes of maternal and perinatal morbidity and mortality (1). The clinical spectrum ranges from mild cases to severe forms accompanied by hepatic and renal dysfunction, hematologic abnormalities, and neurological manifestations. Therefore, early and accurate assessment of disease severity is critically important for both maternal and fetal prognosis (1).

The most widely accepted mechanism in the etiopathogenesis of preeclampsia involves abnormal trophoblast invasion and insufficient spiral artery remodeling, leading to placental hypoperfusion and ischemia (2,3). This hypoxic process triggers the release of anti-angiogenic factors, oxidative stress products, and proinflammatory mediators from the placenta into the maternal circulation. These substances induce systemic endothelial dysfunction, vasoconstriction, and activation of coagulation pathways, thereby shaping the clinical presentation of the disease (3,4). Consequently, the magnitude of the inflammatory response becomes traceable through alterations in hematological and biochemical parameters (4,5).

This systemic inflammatory response during pathogenesis leads to measurable alterations in peripheral blood cells. Neutrophil activation, a decrease in lymphocyte count, and impaired platelet function constitute the hematologic reflections of the preeclamptic process (5,6). Therefore, complete blood count parameters are considered inexpensive and reproducible indicators that reflect

ORCID IDs of the authors: H.B.B. 0000-0001-8516-5493; E.S.C. 0000-0002-6166-5458; S.S. 0000-0001-7090-6105; S.K. 0000-0003-4775-1592; B.B. 0000-0001-8564-5876

Corresponding Author: Havva Betül Bacak, MD;
E-mail: hbbacak90@hotmail.com

Received Date: 24.11.2025 **Accepted Date:** 17.03.2026
Epub: 08.04.2026
Publication Date: 28.04.2026

J Acad Res Med 2026;16(1):24-28

the severity of inflammation. Derived inflammatory indices—particularly the neutrophil-to-lymphocyte ratio (NLR) and platelet-to-lymphocyte ratio (PLR)—have been reported to play a significant role in the pathophysiology of preeclampsia (6-8). Specifically, elevated NLR is associated with neutrophil-mediated oxidative stress and endothelial activation, whereas increased PLR may reflect platelet activation and the vascular response (7,8).

Furthermore, parameters associated with platelet function—such as mean platelet volume (MPV)—and newly derived indices, including the platelet-to-neutrophil ratio (PNR), have also been proposed to correlate with disease severity (8,9). However, evidence regarding the clinical utility of these ratios in distinguishing the severity of preeclampsia remains limited.

Biochemical parameters reflect the manifestation of the systemic inflammatory response in target organ injury. Elevations in aspartate aminotransferase (AST) and alanine aminotransferase (ALT) indicate hepatic microcirculatory disturbance and endothelial damage, whereas increased creatinine levels reflect reduced renal perfusion and glomerular endotheliosis (10,11). The fact that serum creatinine—which physiologically decreases during pregnancy—shows a significant rise in preeclampsia further supports the renal involvement of this systemic process (11).

Evaluating hematological parameters [hemoglobin, white blood cell (WBC), platelet, MPV] and derived inflammatory ratios (NLR, PLR, PNR), together with biochemical markers (AST, ALT, creatinine), enables comprehensive characterization of the inflammation-platelet activation-organ involvement axis of the disease. The present study aimed to assess the associations between hematological and biochemical parameters and disease severity in cases of mild and severe preeclampsia.

METHODS

Study Design and Ethical Approval

This retrospective study involved reviewing the medical records of pregnant women diagnosed with preeclampsia and followed between 2020 and 2025 in the Department of Obstetrics and Gynecology at University of Health Sciences Türkiye, Gaziosmanpaşa Training and Research Hospital. Ethical approval for the study was obtained from the University of Health Sciences Türkiye, University of Health Sciences Türkiye, Gaziosmanpaşa Training and Research Hospital, Clinical Research Ethics Committee (decision no: 17, date: 21.03.2018), and the scope of the approval was revised on 15 October 2025 to update the research period to 2020-2025. All procedures were carried out in accordance with the principles of the Declaration of Helsinki.

Inclusion and Exclusion Criteria

Pregnant women diagnosed with preeclampsia after the 20th week of gestation who had complete laboratory data were included in the study. Exclusion criteria consisted of chronic hypertension, diabetes mellitus, hematologic disorders, chronic renal or hepatic insufficiency, multiple pregnancy, and active infection. Based on these criteria, a total of 135 patients were included in the study.

Patient Grouping

The diagnosis of preeclampsia was established according to criteria defined by international guidelines. Patients were categorized into two groups—mild preeclampsia and severe preeclampsia—according to the classification of the American College of Obstetricians and Gynecologists (ACOG). Blood pressure values, proteinuria levels, and clinical findings were used solely for grouping purposes; therefore, they were neither included in the analytical evaluations nor reported in the tables.

Collected Variables

Demographic data (age, gestational week, and mode of delivery) and hematological parameters (hemoglobin, WBC, platelet count, MPV, MCV, neutrophils, lymphocytes, monocytes, eosinophils, and basophils) were recorded. In addition, derived inflammatory indices—including the NLR, PLR, and PNR—were calculated. As part of the biochemical assessment, AST, ALT, and creatinine levels were evaluated.

Statistical Analysis

Continuous variables were presented as mean ± standard deviation or median [interquartile range (IQR)] according to data distribution. Normality was assessed using the Shapiro-Wilk test, and homogeneity of variances was evaluated with Levene's test. For variables with a normal distribution, the Student's t-test was used; for those not normally distributed, the Mann-Whitney U test was applied. Categorical variables were expressed as numbers and percentages. Because expected cell counts fell below 5 in some categories, Fisher's exact test was preferred for comparing delivery modes. A p-value <0.05 was considered statistically significant. Based on the post-hoc power analysis performed using data from Mohamed and Ali (10), the current sample size was determined to provide approximately 80% statistical power for the analyses.

RESULTS

A total of 135 patients were included in the study, comprising 82 mild and 53 severe preeclampsia cases. The two groups were similar in terms of maternal age [median (IQR): 30.00 (25.00-33.00) vs. 28.00 (24.00-34.75) years, p=0.535), gravidity [median (IQR): 2.00 (1.00-4.00) vs. 3.00 (2.00-4.00), p=0.089], and parity [median (IQR): 2.00 (1.00-2.00) vs. 2.00 (1.00-2.00), p=0.159). However, gestational age at delivery was significantly lower in the severe preeclampsia group compared with the mild preeclampsia group [median (IQR): 36.00 (36.00138.00) vs. 38.00 (37.00139.00) weeks, p<0.001]. The mode of delivery did not differ significantly between the severe and mild preeclampsia groups; cesarean section was the predominant route in both (96.2% vs. 93.9%, respectively; p=0.710) (Table 1).

The comparison of hematological and biochemical parameters is presented in Table 2. Hemoglobin levels were significantly higher in the severe preeclampsia group compared with the mild group (p=0.019). WBC values were also significantly elevated in the severe group (p=0.005), whereas platelet counts were significantly

J Acad Res Med 2026;16(1):24-28

higher in the mild group (p=0.041). Among the biochemical parameters, the levels of AST, ALT, and creatinine were significantly higher in the severe preeclampsia group (p=0.003, p<0.001, and p=0.002, respectively). No significant difference in MPV values was observed between the groups (p=0.833).

Inflammatory cell counts and derived indices are presented in Table 3. Neutrophil counts were significantly higher in the severe preeclampsia group than in the mild group (p=0.004). There was no significant difference between the groups in terms of lymphocyte counts (p=0.535). NLR and PLR values were similar between the two groups, with no statistically significant differences (p=0.614 and p=0.109, respectively). In contrast, PNR levels were significantly higher in the mild preeclampsia group (p=0.001).

DISCUSSION

In this study, hematological, biochemical, and derived inflammatory indices were compared between mild and severe preeclampsia cases, and the biological manifestations of the disease along the inflammation-platelet activation-organ involvement axis were evaluated. The findings indicate that, as the severity of preeclampsia increases, marked deterioration occurs in both hematological and biochemical profiles.

The significantly higher hemoglobin levels observed in the severe preeclampsia group are consistent with the fundamental pathophysiological mechanisms of the disease. Placental hypoperfusion and widespread endothelial dysfunction lead to intravascular fluid loss and a reduction in plasma volume, resulting

Table 1. Demographic and obstetric characteristics of the study groups

Variable	Severe (n=53)	Mild (n=82)	p-value
Age (years)	30.00 (25.00-33.00)	28.00 (24.00-34.75)	0.535
Gravidity	2.00 (1.00-4.00)	3.00 (2.00-4.00)	0.089
Parity	2.00 (1.00-2.00)	2.00 (1.00-2.00)	0.159
Gestational age (weeks)	36.00 (36.00-38.00)	38.00 (37.00-39.00)	<0.001
Mode of delivery - CS	51 (96.2%)	77 (93.9%)	0.710*
Mode of delivery - NSD	2 (3.8%)	5 (6.1%)	-

Continuous variables are presented as medians (IQR, 25th-75th percentiles), and categorical variables are presented as numbers and percentages. Because the distributions were non-normal, the Mann-Whitney U test was used for between-group comparisons
*: The comparison of the mode of delivery was performed using Fisher's exact test, CS: Cesarean section, NSD: Normal spontaneous delivery

Table. 2. Hematological and biochemical laboratory parameters

Parameter	Severe (n=53)	Mild (n=82)	p-value
Hemoglobin (g/dL)	12.00±1.45	11.39±1.49	0.019
WBC (10^9/L)	12.10 (9.90-14.30)	9.91 (9.00-12.57)	0.005
Platelet (10^9/L)	204.06±72.45	229.56±65.93	0.041
MPV (fL)	10.00±1.08	9.96±1.14	0.833
AST (IU/L)	24.00 (19.00-71.00)	20.90 (16.00-26.85)	0.003
ALT (IU/L)	16.00 (10.00-50.25)	10.00 (8.25-17.75)	<0.001
Creatinine (mg/dL)	0.57 (0.51-0.71)	0.52 (0.48-0.59)	0.002

Hemoglobin, platelet count, and MPV were compared using Student's t-test; all other parameters were compared using the Mann-Whitney U test due to their non-normal distribution. Parameters analyzed with Student's t-test are presented as mean ± standard deviation, whereas those analyzed with the Mann-Whitney U test are presented as median (interquartile range, 25th-75th percentile)
WBC: White blood cell, MPV: Mean platelet volume, AST: Aspartate aminotransferase, ALT: Alanine aminotransferase

Table 3. Inflammatory cell counts and derived indices

Parameter	Severe (n=53)	Mild (n=82)	p-value
Neutrophils (10^3/μL)	8.60 (6.71-10.30)	7.09 (5.85-8.68)	0.004
Lymphocytes (10^3/μL)	2.07 (1.66-2.58)	2.07 (1.64-2.44)	0.535
NLR	3.00 (2.00-5.00)	3.00 (2.00-4.00)	0.614
PLR	93.00 (70.00-135.00)	108.50 (86.00-139.75)	0.109
PNR	24.00 (16.00-33.00)	29.00 (25.00-38.00)	0.001

All variables are presented as medians (interquartile range, 25th-75th percentiles). All parameters in this table were compared using the Mann-Whitney U test because the data were not normally distributed
NLR: Neutrophil-to-lymphocyte ratio, PLR: Platelet-to-lymphocyte ratio, PNR: Platelet-to-neutrophil ratio

J Acad Res Med 2026;16(1):24-28

in relative hemoconcentration and consequently elevated hemoglobin levels (1). This mechanism is considered part of the vasoconstrictive and hypovolemic response frequently seen in severe preeclampsia. However, findings in the literature regarding the relationship between hemoglobin levels and preeclampsia severity are inconsistent; while some studies have reported no significant association (10), others have indicated that hemoglobin levels do not parallel the clinical presentation (12). The increase in hemoglobin observed in our study aligns with the volume contraction seen in severe cases, suggesting that this parameter may provide supportive information for assessing disease severity.

The marked elevation of WBC and neutrophil counts in the severe preeclampsia group reflects the systemic inflammatory component of the disorder. Neutrophil activation, oxidative stress generation, and endothelial injury play central roles in the pathogenesis (2,4). Neutrophilia has been associated with preeclampsia severity in numerous studies, consistent with the findings of our study (5,10,13). The wide standard deviation of neutrophil counts in the mild group may indicate variability in individual inflammatory responses. The absence of a significant difference in lymphocyte counts suggests that the neutrophil-dominant inflammatory response may not have resulted in pronounced lymphocyte suppression.

The significantly lower platelet counts observed in the severe group may be explained by platelet consumption secondary to coagulation activation in preeclampsia. Endothelial injury, exposure of the subendothelial matrix, and increased microthrombus formation lead to the withdrawal of platelets from the circulation (3,8). This finding is consistent with reports in the literature indicating that thrombocytopenia is associated with disease severity (8,14,15).

The absence of a significant difference in MPV levels aligns with the conflicting findings reported in the literature. While some studies have suggested that MPV may indicate increased platelet activation, others have demonstrated that it is not a reliable discriminator (7,14). Therefore, MPV alone may not serve as a dependable marker.

The markedly elevated AST, ALT, and creatinine levels observed in the severe preeclampsia group represent biochemical evidence of endothelial injury and target-organ involvement. Hepatocellular impairment and microcirculatory disturbance lead to increases in hepatic enzymes, whereas glomerular endotheliosis and reduced filtration contribute to elevated creatinine levels (1,10,11). Our findings are largely consistent with the literature reporting more pronounced renal and hepatic dysfunction in severe preeclampsia (11,12).

The absence of significant differences in NLR and PLR values between the two groups suggests these ratios may be influenced by population characteristics, the timing of sampling, and individual variability in inflammatory burden. Although several studies have reported increased NLR and PLR levels in severe preeclampsia (6,7,13), other series have found no significant differences (4,9). Moreover, evidence indicates that the performance of derived inflammatory indices in distinguishing disease severity is not consistently reliable (16,17). Our findings align with the latter group of studies, suggesting that NLR and PLR may have limited value in differentiating severe preeclampsia.

One of the most notable findings of this study is that PNR values were significantly higher in the mild group compared with the severe group. Since PNR represents the ratio of platelet count to neutrophil count, it simultaneously reflects platelet consumption and neutrophil elevation. Although data on PNR in the literature are limited, it has been suggested that this index may decrease in conditions where inflammation and platelet activation increase concurrently (14,18). Our results indicate that PNR values differed significantly between mild and severe preeclampsia groups, suggesting a potential association with disease severity. However, since diagnostic performance analyses such as ROC/AUC were not performed in the present study, the true discriminatory value of PNR cannot be determined. Further prospective studies are required to evaluate its diagnostic accuracy.

Moreover, the ease of calculating PNR from routine complete blood count parameters enhances its applicability in clinical practice.

Overall, the severe preeclampsia group demonstrated markedly increased inflammation (WBC, neutrophils), enhanced platelet consumption (reduced platelet count, decreased PNR), and more pronounced organ involvement (AST, ALT, creatinine). These findings support the progressive, multisystemic nature of preeclampsia. The lack of significant differences in NLR and PLR reflects the inter-population variability of these biomarkers (19,20) and suggests that PNR may represent a potentially useful adjunct parameter. A combined assessment of hematological and biochemical markers may offer a more comprehensive and reliable approach for determining disease severity.

This study has several notable strengths. First, all patients were followed at the same center, and laboratory measurements were performed under a single standardized protocol, which reduced biological and technical variability and thereby enhanced the internal validity of the results. The classification of mild and severe preeclampsia according to ACOG criteria further strengthened the accuracy of group stratification. Additionally, the simultaneous evaluation of hematological, biochemical, and derived inflammatory indices provided a comprehensive approach—integrating inflammation, platelet activation, and organ involvement—and enabled a multidimensional assessment of disease severity. Furthermore, the analysis of PNR as a severity marker, despite the limited data in the literature, represents one of this study's original contributions.

Study Limitations

This study has several limitations. Due to its retrospective design, the study's ability to establish causal relationships is limited, and missing records may introduce potential bias. The sample size—particularly the relatively small number of patients in the severe

preeclampsia group—may restrict the statistical power of the study. Measuring laboratory parameters at a single time point may not fully capture the dynamic nature of the inflammatory response. Despite strict exclusion criteria, residual confounding related to subclinical inflammation or unmeasured variables cannot be completely ruled out. Although gestational age was included as a baseline characteristic, the absence of multivariable analysis limits the ability to fully adjust for potential confounders. Additionally, the single-center design and a lack of evaluation across populations with different demographic characteristics limit the generalizability of the findings.

CONCLUSION

Severe preeclampsia is characterized by markedly increased inflammation, enhanced platelet consumption, and more pronounced organ involvement. While NLR and PLR may not always be sufficient to distinguish disease severity, PNR differed significantly between mild and severe preeclampsia groups and may serve as a potential adjunct marker. However, its clinical diagnostic value should be confirmed in larger prospective studies. A combined assessment of hematological and biochemical markers may provide a more comprehensive and reliable approach for determining preeclampsia severity. Supporting these findings with larger sample sizes and prospective study designs will strengthen their applicability in clinical practice.

Ethics

Ethics Committee Approval: Ethical approval for the study was obtained from the University of Health Sciences Türkiye, Gaziosmanpaşa Training and Research Hospital, Clinical Research Ethics Committee (decision no: 17, date: 21.03.2018).

Informed Consent: Retrospective study.

Footnotes

Author Contributions: Concept - H.B.B., S.K.; Design - H.B.B., S.S.; Data Collection and/or Processing - E.S.C., B.B.; Analysis and/or Interpretation - S.S., S.K.; Literature Search - E.S.C., S.K.; Writing - H.B.B.

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

Financial Disclosure: The authors report that no financial support was received for this study.

REFERENCES

1. Mol BWJ, Roberts CT, Thangaratinam S, Magee LA, de Groot CJM, Hofmeyr GJ. Pre-eclampsia. Lancet. 2016; 387: 999-1011.
2. Roberts JM, Hubel CA. The two stage model of preeclampsia: variations on the theme. Placenta. 2009; 30(Suppl A): S32-7.
3. Redman CW, Sargent IL. Placental stress and pre-eclampsia: a revised view. Placenta. 2009; 30(Suppl A): S38-42.
4. Çintesun E, Incesu Çintesun FN, Ezveci H, Akyürek F, Çelik Ç. Systemic inflammatory response markers in preeclampsia. J Lab Physicians. 2018; 10: 316-9.
5. Bozdağ H, Demircivi Bör E, Akdeniz E. The predictive value of total leukocyte count and leukocyte differential for severe preeclampsia. Perinatal J. 2018; 26: 25-31.
6. Kang Q, Li W, Yu N, Fan L, Zhang Y, Sha M, et al. Predictive role of neutrophil-to-lymphocyte ratio in preeclampsia: a meta-analysis including 3982 patients. Pregnancy Hypertens. 2020; 20: 111-8.
7. Yücel B, Ustun B. Neutrophil to lymphocyte ratio, platelet to lymphocyte ratio, mean platelet volume, red cell distribution width and plateletcrit in preeclampsia. Pregnancy Hypertens. 2017; 7: 29-32.
8. Temur M, Taşgöz FN, Çift T, Serpim G, Üstünyurt E. Role of platelet indices in prediction of preeclampsia. Ginekol Pol. 2021; 92: 792-6.
9. Tokgöz Çakır B, Aktemur G, Karabay G, Şeyhanlı Z, Çetin S, Filiz AA, et al. Evaluation of platelet indices and inflammation markers in preeclampsia. J Clin Med. 2025; 14: 1406.
10. Mohamed RA, Ali IA. Role of neutrophil/lymphocyte ratio, uric acid/albumin ratio and uric acid/creatinine ratio as predictors to severity of preeclampsia. BMC Pregnancy Childbirth. 2023; 23: 763.
11. Nguyen LM, Crants SA, Giordano CM, Zuckerwise LC, Grace MR. Association between pregnancy-related changes in serum creatinine and preeclampsia diagnosis. Arch Gynecol Obstet. 2024; 310: 1975-80.
12. Özkavak H, Ateş S, Özkavak E, Kaya AE, Çağlar T, Budak MS. Severity of disease in preeclampsia and its relationship with routine hematological parameters. Zeynep Kamil Med J. 2020; 51: 145-50.
13. Örgül G, Aydın Haklı D, Özten G, Fadiloğlu E, Tanacan A, Beksaç MS. First trimester complete blood cell indices in early and late onset preeclampsia. Turk J Obstet Gynecol. 2019; 16: 112-7.
14. Freitas LG, Alpoim PN, Komatsuzaki F, Carvalho Md, Dusse LM. Preeclampsia: are platelet count and indices useful for its prognostic? Hematology. 2013; 18: 360-4.
15. Udeh PI, Olumodeji AM, Kuye-Kuku TO, Orekoya OO, Ayanbode O, Fabamwo AO. Evaluating mean platelet volume and platelet distribution width as predictors of early-onset pre-eclampsia: a prospective cohort study. Matern Health Neonatol Perinatol. 2024; 10: 5.
16. Çetin Arslan H, Yılmaz N, Arslan K. Prognostic value of systemic immune-inflammation index in early- and late-onset preeclampsia. Comprehensive Medicine. 2025; 17: 124-9.
17. Gao LN, Yan D, Liu XH, Chen D, Guo H, Liu J. Association of systemic inflammatory biomarkers (NLR, MLR, PLR, SII, SIRI) with preeclampsia-related kidney injury: a retrospective observational study. J Inflamm Res. 2025; 18: 12725-38.
18. Bayram F, Karaşin SS. The predictive role of neutrophil-lymphocyte ratio, platelet lymphocyte ratio, and other complete blood count parameters in eclampsia and HELLP syndrome. J Surg Med. 2021; 5: 1202-5.
19. Demir Çendek B, Ağbal T, Akay A, Can İbanoğlu M, Engin Üstün Y. The performance of various serum parameters in blood during the first trimester in the early detection of pre-eclampsia. J Health Sci Med. 2024; 7: 451-8.
20. Göksun A, Çelik E, Kaymak Ç, Akdeniz E, Acar DB, Üstüner İ, et al. The role of inflammatory biomarkers in hypertensive disorders of pregnancy. J Exp Clin Med. 2021; 38: 521-6.

DOI: 10.4274/jarem.galenos.2026.57070
J Acad Res Med 2026;16(1):29-35

Multidimensional Data Visualization to Evaluate the Post-treatment Prognosis of Patients Presenting with Methanol Intoxication: Chernoff Faces

Büşra Dilara Dayıoğlu[1], Ejder Saylav Bora[2], Mehmet Göktuğ Efgan[2], Mustafa Agah Tekindal[3], Elif Kaymaz[3]

[1]University of Health Sciences Türkiye, Kartal Dr. Lütfi Kırdar City Hospital, Department of Emergency Medicine, İstanbul, Türkiye
[2]İzmir Katip Çelebi University Faculty of Medicine, Department of Emergency Medicine, İzmir, Türkiye
[3]İzmir Katip Çelebi University Faculty of Medicine, Department of Biostatistics, İzmir, Türkiye

Cite this article as: Dayıoğlu BD, Bora ES, Efgan MG, Tekindal MA, Kaymaz E. Multidimensional data visualization to evaluate the post-treatment prognosis of patients presenting with methanol intoxication: Chernoff faces. J Acad Res Med. 2026;16(1):29-35

ABSTRACT

Objective: Methanol intoxication is life-threatening and often needs urgent intervention due to severe metabolic disturbances. This study aims to evaluate Chernoff faces (CF) as a visual tool for representing patient prognosis. It also aims to support rapid clinical decision-making in cases of methanol intoxication.

Methods: This retrospective study included 81 patients who presented to the emergency department with methanol intoxication. Key biochemical parameters—pH, bicarbonate (HCO_3), base excess, and creatinine—were recorded before and after dialysis. CF, a graphical method for displaying multidimensional data, were used to visualize these values. The technique illustrated both individual—and group—level changes in metabolic status.

Results: The cohort comprised 81 patients (6 females, 75 males; mean age: 52.9 years). After dialysis, significant improvements were observed: pH and HCO_3 levels increased, while blood urea nitrogen and creatinine levels decreased. Chernoff's facial features clearly depict these changes. Pre-dialysis patients exhibited distressed facial expressions, reflecting severe acidosis. Post-dialysis faces appeared more harmonious and "smiling", suggesting metabolic recovery. This visual shift was consistent across the entire group and in selected individuals. The method effectively and intuitively demonstrated clinical improvement.

Conclusion: CF provides a practical, visually intuitive method for monitoring metabolic recovery in methanol intoxication. By enabling rapid assessment of complex data, this tool enhances decision-making, particularly in time-sensitive settings like mass poisonings. Further studies will expand its use to other clinical contexts requiring the swift interpretation of multidimensional parameters.

Keywords: Methanol intoxication, Chernoff faces, multidimensional data, data visualization, emergency medicine

INTRODUCTION

Methanol is a clear, colourless, and odourless solvent widely used in industry as an antifreeze and cleaning agent; it is produced via wood distillation. It is often mistaken for ethanol and illegally added to cheap alcoholic beverages, leading to accidental and intentional poisonings. While ingestion is the most common route, transdermal and inhalation exposures have also been reported (1-4). Methanol is rapidly absorbed from the gastrointestinal tract, reaching peak serum levels within 30-60 minutes. Its toxicity stems from its metabolites-formaldehyde and formic acid. Formic acid inhibits mitochondrial oxidative phosphorylation, leading to high-anion-gap metabolic acidosis (5). Clinically, methanol poisoning progresses through three phases: an early phase with non-specific symptoms (nausea, headache, vomiting), a latent phase with few or no symptoms, and a late phase marked by severe metabolic acidosis, visual disturbances (blurred vision, photophobia, diplopia), and central nervous system depression (6,7).

Treatment involves stabilization, antidotal therapy, and extracorporeal elimination. Airway, breathing, and circulation are secured with oxygen and IV fluids. Antidotes (ethanol or fomepizole) inhibit alcohol dehydrogenase, thereby preventing the formation of toxic metabolites. A loading dose followed by a

ORCID IDs of the authors: B.D.D. 0000-0001-5668-1501; E.S.B. 0000-0002-2448-2337; M.G.E. 0000-0002-0794-1239; M.A.T. 0000-0002-4060-7048; E.K. 0000-0003-2631-3067

Corresponding Author: Ejder Saylav Bora, Assoc. Prof.;
E-mail: ejdersaylav@gmail.com
Received Date: 05.08.2025 **Accepted Date:** 17.03.2026
Publication Date: 28.04.2026

J Acad Res Med 2026;16(1):29-35

maintenance infusion aims to achieve serum levels of 100-150 mg/dL. Fomepizole is given at 15 mg/kg IV every 12 hours, adjusted by response and labs to minimize side effects (8,9).

In severe cases, hemodialysis rapidly eliminates methanol and its metabolites while correcting acidosis (10). Folate and sodium bicarbonate (HCO_3) support formatted metabolism and acid-base balance (11). In the busy emergency department, clinicians must evaluate multiple biochemical parameters to guide treatment. Complex interactions may not be easily understood using numerical data. Alternative visualization methods can help interpret multidimensional data. Chernoff faces (CF), invented by Chernoff (12) in 1973, are widely used to visualize complex datasets by mapping multiple variables to facial features such as eye size, mouth curvature, and nose length. This technique, developed for general data analysis, has been used in psychology, social sciences, engineering, and finance to find patterns, clusters, and anomalies in large datasets (13). CF enables medical researchers to visually represent clinical data and compare multiple variables in an intuitive, interpretable format.

Computational algorithms generate CF. Each facial feature represents a variable in a multidimensional dataset. Any change in the 18 parameters alters the facial structure; therefore, each face visually encodes a unique, 18-dimensional data point. For datasets with fewer variables, missing dimensions are filled with fixed values to maintain consistency (12). Stata, Python, MATLAB, and R can create CF. These tools make complex datasets visual. Trends and relationships are easier to identify than when presented in tables or as numerical data. This method helps researchers and clinicians interpret large-scale data more easily and holistically.

This study examined CF as a multidimensional data visualization tool in emergency medicine, particularly for interpreting metabolic parameters and guiding clinical management. The parameters and facial features of the original CF are presented in Table 1 and Figure 1.

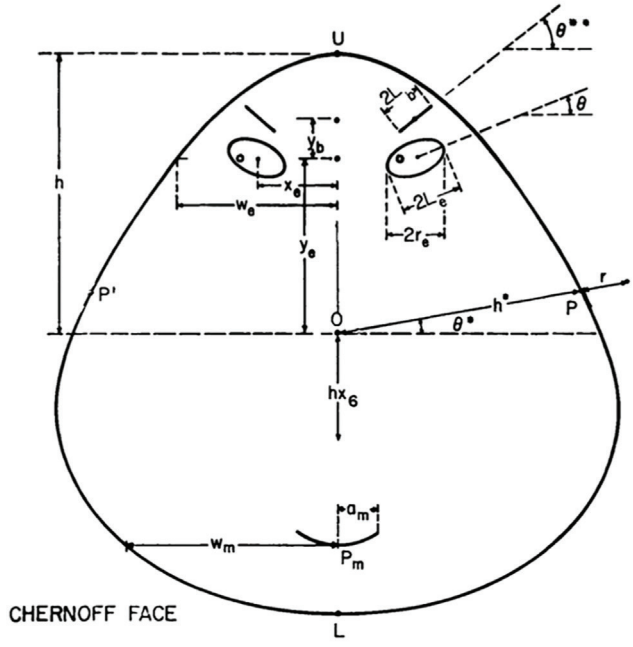

Figure 1. The original Chernoff face

Table 1. Parameters and facial features represented in the original Chernoff face

Parameter (X)	Feature	Description
*X*1	Face width (h*)	Horizontal width of the face, low values correspond to narrower faces
*X*2	Ear level (θ*)	Horizontal level of the ears based on distance from the face center
*X*3	Face height (h)	Vertical height of the face
*X*4	Upper face eccentricitys	Roundness or ellipticity of the upper half of the face
*X*5	Lower face eccentricity	Roundness or ellipticity of the lower half of the face
*X*6	Nose length	Length of the nose, high values indicate longer noses
*X*7	Mouth center (P_m)	Vertical position of the mouth
*X*8	Mouth curvature	Curvature of the mouth, indicating positive or negative expressions
*X*9	Mouth length	Width of the mouth
*X*10	Height of eye center (Ye)	Vertical position of the eyes
*X*11	Eye separation (*X*e)	Distance between the two eyes
*X*12	Eye slant (θ)	Horizontal slant of the eyes
*X*13	Eye eccentricity	Ovality or roundness of the eyes
*X*14	Eye length (Le)	Size of the eyes
*X*15	Pupil position	Position of the pupil within the eye
*X*16	Eyebrow height (Yb)	Vertical position of the eyebrows
*X*17	Eyebrow angle (θ**)	Angle of the eyebrows
*X*18	Eyebrow length	Length of the eyebrows

h*: A normalized (standardized) measure representing facial width. It expresses a relative (proportional) value rather than an absolute length,
θ*: An angular parameter indicating the position of features (e.g., ear level) relative to the facial center,
P_m: A normalized parameter representing the vertical position of the mouth center,
** (double asterisk): A derived angular measurement, particularly used for parameters such as eyebrow angle that define facial expressions

J Acad Res Med 2026;16(1):29-35

METHODS

Study Design

This research was conducted as a retrospective, observational study and was approved by the İzmir Katip Çelebi University Ethics Committee on March 21, 2024, under decision number 0175. It was conducted in accordance with ethical and scientific standards and was approved unanimously by the committee.

Study Population

Patients aged 18 years and older with methanol intoxication who received dialysis between November 1, 2015, and September 1, 2022 were included. Eligible patients were identified in the hospital database using international classification of diseases codes F10 (alcoholic mental and behavioral disorders) and T51.1 (methanol toxicity). Confirmation was obtained by reviewing nephrology and internal medicine consultations. Of the 92 patients initially identified, 6 were excluded due to missing data and 5 were excluded because they did not receive dialysis. For the final study, clinical and laboratory data from 81 patients were extracted from records and stored in a structured database.

Variables

The collected data included demographics (age, gender) and mode of hospital arrival (ambulance or self-admission).

Clinical symptoms: blurred vision, altered mental status, suspected alcohol intake, nausea/vomiting, syncope, dyspnea

Laboratory parameters: blood urea nitrogen (BUN), creatinine, sodium, potassium, chloride, calcium, pH, partial pressure of carbon dioxide ($PaCO_2$), HCO_3, lactate, ethanol, base excess, amylase, total bilirubin, international normalized ratio (INR).

These laboratory values were mapped to different facial features in a multidimensional visualization using CF.

Statistical Analysis

Data were analyzed using IBM SPSS Statistics (Standard Concurrent User, v. 27; IBM Corp., Armonk, New York, USA). Descriptive statistics were presented as the number of units (n), percentage (%), mean ± standard deviation, median, minimum, and maximum values. The normality of the numerical variables was assessed using the Shapiro-Wilk normality test. Values that met the prerequisites for parametric tests were analyzed using the paired t-test; otherwise, the Wilcoxon signed-rank test was used. A p-value <0.05 was considered significant. Each laboratory parameter was mapped to a facial feature for CF visualizations. Assignments included: face—height (BUN), width ($PaCO_2$), and shape (lactate); mouth—height (sodium, HCO_3) and curvature (smile; pH); eye—height (base excess) and width (ethanol); hair—height (potassium, chloride, calcium); nose—height (creatinine, amylase); and ear—width (total bilirubin) and height (INR). This variable-feature mapping is shown in Table 2.

Data visual data visualization was performed in RStudio version 2024.04.2 using the "aplpack" package for CF representation. This study used CF to visualize pre- and post-dialysis laboratory parameters in patients with methanol intoxication. Representations of metabolic changes after treatment made it easy to understand these changes.

RESULTS

This study included 81 patients: 6 women and 75 men. The average age was 52.9 years, ranging from 18 to 77. The emergency department received 38 walk-ins and 43 ambulance arrivals. Blurred vision (20 patients) and altered mental status (37 patients) were the most common symptoms. Overall, 54.3% of patients required intensive care. 12.3% of ED patients died (exitus), while 13.6% were discharged. The descriptive statistics for the patients are presented in Table 3.

Post-dialysis laboratory values showed a significant decrease in BUN and creatinine levels ($p<0.01$). While sodium levels remained unchanged ($p=0.815$), potassium levels significantly decreased ($p<0.001$). Treatment resulted in a significant increase in pH, indicating an improvement in metabolic acidosis ($p<0.001$). Lactate levels decreased slightly after treatment ($p=0.053$). Additionally, levels of base excess, HCO_3, INR, and total bilirubin significantly improved ($p<0.001$). Laboratory values and data visualization before and after dialysis are presented in Table 4 and Figure 2.

Data visualization and CF were generated from the average laboratory data of all patients, collected before and after dialysis. The CF technique effectively demonstrated the variations in biochemical parameters pre- and post-dialysis. The pH levels associated with the "smile" feature of the faces increased significantly post-treatment, from 7.04±0.23 to 7.17±0.24

Table 2. Laboratory parameters mapped to Chernoff face features in our study

Chernoff face parameter	Corresponding laboratory value
Face height	BUN
Face width	$PaCO_2$
Face structure	Lactate
Mouth height	Sodium
Mouth width	HCO_3
Smile	pH
Eye height	Base deficit
Eye width	Ethanol
Hair height	Potassium
Hair width	Chloride
Hair style	Calcium
Nose height	Creatinine
Nose width	Amylase
Ear width	Total bilirubin
Ear height	INR

BUN: Blood urea nitrogen, INR: International normalized ratio, $PaCO_2$: Partial pressure of carbon dioxide, HCO_3: Bicarbonate

J Acad Res Med 2026;16(1):29-35

(p<0.001). The change was visually evident in the post-treatment facial appearance, with a more pronounced smile, signifying improvement in metabolic acidosis. BUN levels decreased from 16.33±14.29 to 14.71±12.05 (p=0.010), indicating a corresponding reduction in face height in the CF. This visual alteration indicated an improvement in renal function. Creatinine levels demonstrated a significant reduction (from 1.33±0.52 to 1.21±0.56, p=0.004), which correlated with a decrease in nose height in the visual representation. HCO_3 levels, correlated with mouth width, increased from 11.31±6.88 to 15.55±7.31 (p<0.001), indicating a wider mouth in post-treatment faces, which visually reflected the correction of metabolic acidosis. Ethanol levels, indicated by eye width, decreased from 65.22±103.44 to 52.17±89.63 (p=0.028), resulting in narrower eyes in post-treatment faces. The base deficit, associated with eye height, improved significantly from -17.87±9.85 to -7.35±8.26 (p<0.001), resulting in elevated eye positions in the post-treatment faces. Potassium levels, correlated with hair height, decreased from 4.94±1.16 to 4.23±1.19 (p<0.001), resulting in shorter hair in the visual representation. representation.

Total bilirubin levels, indicated by ear width, rose from 0.66±0.62 to 1.01±0.75 (p<0.001), leading to an increase in ear width in post-treatment subjects.

The pre-dialysis CF depicted features consistent with impaired metabolic and renal function, including elongated facial features and reduced pH. In contrast, the post-dialysis facial appearance demonstrates significant changes: a shorter, broader structure; a noticeable smile indicating pH normalization; and wider eyes suggesting improved metabolic status.

Table 3. Demographic and clinical features of all patients

Variables	n	%
Gender		
Female	6	7.4
Male	75	92.6
Age		
$\bar{x}$ ± SD	52.90±14.33	
M (min-max)	56 (18-77)	
Admission type		
Walk-in	38	46.9
Ambulance	43	53.1
Presenting symptoms		
Blurred vision	20	24.7
Altered mental status	37	45.7
Suspected alcohol intake	16	19.8
Nausea/vomiting	3	3.7
Syncope	2	2.5
Dyspnea	3	3.7
Mortality		
Exitus	18	22.2
Survive	63	77.8
Outcome		
Exitus	10	12.3
Intensive care	44	54.3
Inpatient ward	12	14.8
Left against medical advice	4	4.9
Discharged	11	13.6

$\bar{x}$: Mean, SD: Standard deviation, M: Median

Case Examples

CF is shown in four randomly selected patients to demonstrate their clinical features. To ensure impartial representation of the study population, these patients were selected without selection criteria. The comparison of pre- and post-dialysis biochemical

Table 4. Laboratory values before and after dialysis

Variable	Measurement times		Test statistic	
	Before	After	Test value	p-value
BUN	13 (9)	11 (7.50)	2.593	0.010¥
Creatinine	1.24 (0.63)	1.11 (0.69)	2.898	**0.004¥**
Sodium	137.33±5.59	137.23±5.01	0.235	0.815£
Potassium	4.80 (1.60)	3.90 (1.40)	5.080	**<0.001¥**
Chloride	102.0 (6.50)	103.0 (6.50)	0.869	0.385¥
Calcium	8.78+0.86	8.66+0.88	1.524	0.131£
pH	7.10 (0.37)	7.30 (0.36)	5.144	**<0.001¥**
$PaCO_2$	32.5 (21.05)	34.10 (14.05)	1.505	0.132¥
HCO_3	8.90 (7.60)	16.90 (13.50)	5.143	**<0.001¥**
Lactate	4.30 (8.20)	3.0 (7.05)	1.936	0.053¥
Ethanol	10.0 (76.0)	10.0 (25.0)	2.194	**0.028¥**
Base excess	-21.0 (11.50)	-6.0 (11)	6.538	**<0.001¥**
Amylase	75.0 (57.50)	68.0 (80)	1.715	0.086¥
Total bilirubin	0.46 (0.50)	0.84 (0.97)	4.480	**<0.001¥**
INR	0.97 (0.14)	1.04 (0.23)	4.061	**<0.001¥**

Data are expressed as mean ± standard deviation or median (interquartile range)
¥: Wilcoxon signed-rank test, £: Paired samples t-test, $PaCO_2$: Partial pressure of carbon dioxide, HCO_3: Bicarbonate

J Acad Res Med 2026;16(1):29-35

parameters in the case series and CF, based on individual data from four selected patients, is shown in Table 5 and Figure 3.

In the first patient, the pre-dialysis facial appearance was tense and asymmetrical, reflecting severe metabolic derangement. Elevated lactate (15.0) and $PaCO_2$ (36.0) distorted the facial structure, while extreme acidosis (pH 6.60) created a pronounced frown. Post-dialysis, normalized values (lactate 1.2, pH 7.49, HCO_3 25.1) produced a more symmetrical, smiling facial appearance, visually confirming an improved acid-base balance.

Similarly, the second patient's face narrowed after $PaCO_2$ dropped from 50.7 to 31.6, and the frown softened as pH improved from 6.8 to 7.3. An increase in HCO_3 (5.9 to 15.3) and a reduction in lactate (13.6 to 8.7) further normalized facial features. These changes reflected respiratory and metabolic recovery.

The third patient showed subtle but meaningful improvements: pH increased from 7.33 to 7.37, base excess from -3 to 3, and HCO_3 from 21.4 to 25.5; a sharp drop in ethanol (286 to 10) and narrowing of the pupils, indicating detoxification after methanol poisoning. The patient initially presented with a severely distorted facial appearance due to extreme acidosis (pH 6.60, base deficit -27). After treatment, facial balance improved; however, residual acidosis (pH 7.19, base deficit -11) left the mouth relatively narrow. Overall, CF offered a unique visual summary of patients' clinical progression. Changes in smile, eye position, and facial symmetry corresponded to key biochemical improvements, enabling rapid and intuitive assessment of treatment effectiveness.

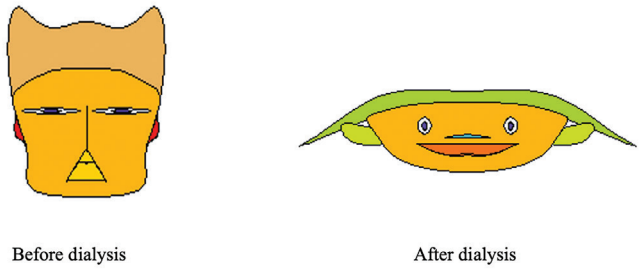

Figure 2. Chernoff face derived from pre- and post-dialysis mean data

DISCUSSION

Methanol poisoning, a rare but serious medical condition, is caused by industrial or illegal alcohol consumption (1,2). Outbreaks can strain healthcare systems. Managing complex metabolic disturbances during crises requires timely and accurate clinical decision-making. Methanol intoxication causes rapid fluctuations in pH, base deficit, HCO_3, and lactate, making real-time assessment crucial for patient outcomes.

CF has been studied and applied in epidemiology, environmental sciences, and social sciences since Hermann Chernoff introduced it in 1973, but its use in clinical medicine remains limited. One study has compared them to star plots and profile graphs (14). Some researchers have shown that CF elicits rapid visual awareness, but data analysis has mostly focused on detecting psychological states and on depicting heart rate and other physiological data in evaluations of adolescent social anxiety, enabling non-experts to accurately discern patterns and deviations from established norms (15-18). In this study, we aim to understand not only patients' central state but also their metabolic state upon arrival at emergency services. Because the facial results are binary (white or black), more facial-point variables are needed to better understand the severity of patients. In a 1992 study by Phillipou (19),

Table 5. Comparison of pre- and post-dialysis biochemical parameters in case series

Parameter	Patient 1		Patient 2		Patient 3		Patient 4	
	Before	After	Before	After	Before	After	Before	After
BUN	12	12	21.0	21.0	10	12	8	8
$PaCO_2$	36.0	29.7	50.7	31.6	41.5	48.3	53.9	43.6
Lactate	15.0	1.2	13.6	8.7	2.1	1.4	8	5.5
Sodium	140	134	138.0	139.0	143	144	135	132
HCO_3	3.8	25.1	5.9	15.3	21.4	25.5	10.6	15.6
pH	6.60	7.49	6.8	7.3	7.33	7.37	6.60	7.19
Base excess	-26	-1	-24.2	-10.7	-3	3	-27	-11
Ethanol	10	10	10	10	286	10	397	10
Potassium	4.7	3.7	5.4	3.2	3.7	3.8	5.9	3.6
Chloride	105	105	105	105	99	104	102	104
Calcium	9.1	8.7	9.3	9.3	9.1	9.4	7.5	6.5
Creatinine	1.29	1.62	1.8	1.8	0.81	0.82	1.53	1.45
Amylase	283	181	78	78	77	63	175	116
Total bilirubin	0.85	1.00	0.4	0.4	0.31	0.50	0.18	0.18
INR	1.07	0.97	1.0	1.0	0.93	0.95	0.97	1.41

BUN: Blood urea nitrogen, INR: International normalized ratio, $PaCO_2$: Partial pressure of carbon dioxide, HCO_3: Bicarbonate

J Acad Res Med 2026;16(1):29-35

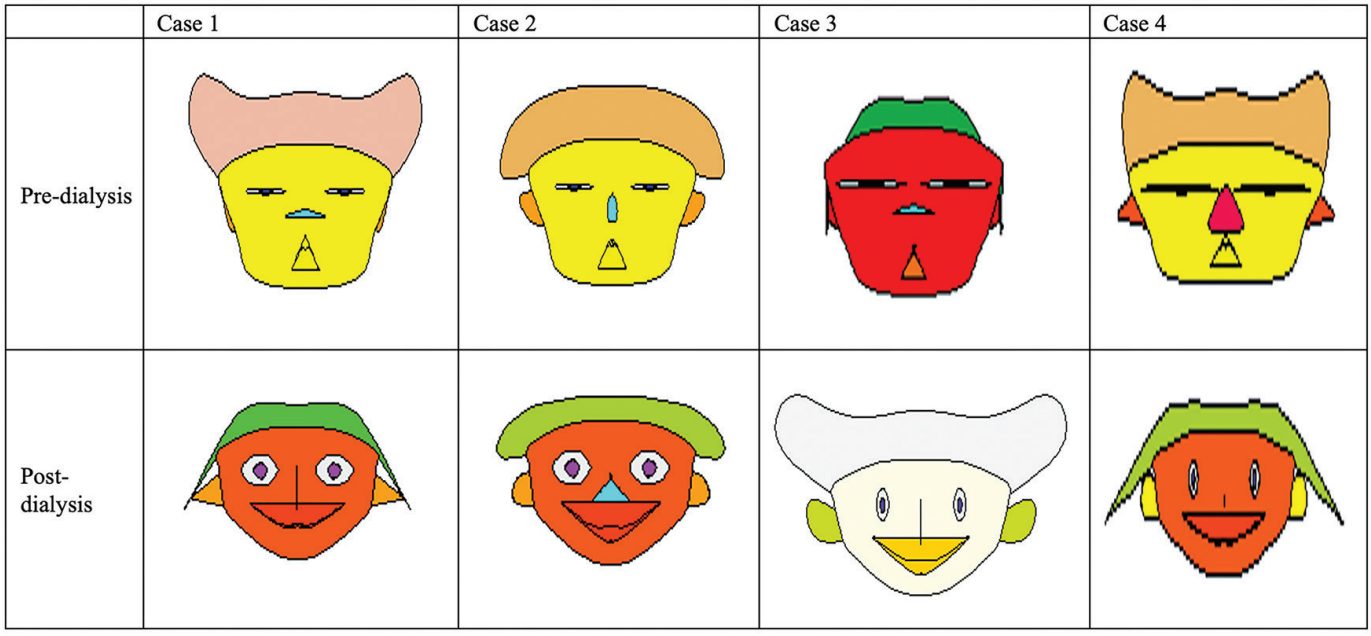

Figure 3. Chernoff faces based on individual data from four selected patients

CF was used to visually represent capillary blood glucose levels in diabetic individuals, facilitating rapid assessment of glucose regulation and temporal trends. This method enables both patients and healthcare providers to comprehend complex data more effectively than with conventional charts or tables. In studies by He et al. (20) and Shi et al. (21), the authors claim that machine learning, specifically RNN-LSTM, can precisely forecast the onset of acute kidney disease in patients with sepsis-associated acute kidney injury. Moreover, flexible machine learning techniques are widely used for predicting acute kidney injury, yet more complex deep learning models are emerging (22). As CF is a machine learning starter model, and among its advantages, rapid results are the most important, it can be used in emergency services in triage areas for multiple illnesses, such as methanol intoxication CF has been used to visually represent cancer incidence rates and to interpret large-scale health data in epidemiology (23). Lott and Durbridge (24) demonstrated how CF could be employed to track metabolic and biochemical changes in critically ill patients, thereby facilitating the interpretation of complex laboratory trends. In orthodontic research, CF has been utilized to visualize morphological changes before and after treatment (25). In this study, the pre- and post-treatment data of patients who presented to the emergency department with methanol intoxication were visualized, and their potential contributions to clinical decision-making were examined. The results indicated that facial visualizations generated from key laboratory parameters commonly monitored in methanol intoxication could serve as effective tools to support clinical decision-making before and after treatment. Although this study focused specifically on CF—one of the most widely used methods of data visualization—the findings underscore the broader value of data visualization in enhancing diagnostic and therapeutic processes. Previous studies have reported that CF is an effective tool for displaying trends in laboratory data and assisting in the recognition of abnormalities in acute care settings (24,26,27). Considering advances in technology and artificial intelligence, it is increasingly apparent that data visualization can be used more effectively. The conclusion that data visualization may be a useful tool in acute conditions such as methanol intoxication, which require urgent treatment and close monitoring, is therefore particularly significant. More research has examined which facial features are most perceptually informative (27,28). Eye size and mouth width are more intuitively recognized by observers, making them effective for visually encoding critical variables, while nose length and ear width had less impact on information retention (27). This study linked metabolic markers to distinct facial features, such as pH (smiling expression), HCO_3 (mouth width), base deficit (eye height), and ethanol (eye width), highlighting treatment-related changes. Pre- and post-dialysis average laboratory values for all patients showed a striking difference in CF values. Post-treatment, the face showed improved metabolic stability, whereas the pre-treatment face showed severe metabolic derangement. Individual case analyses showed clear correlations between laboratory improvements and facial transformations, supporting the clinical use of CF.

Study Limitations

CFSs have limitations despite their benefits. As the number of variables increases, facial representation becomes more complex, potentially causing cognitive overload. Chernoff noted that exceeding 18 variables could reduce interpretability and reduce the technique's ability to convey nuanced details (12). Graphical representations simplify data interpretation, but they may obscure numerical subtleties and require additional analytical methods for high-precision evaluations. Integrating CF into

J Acad Res Med 2026;16(1):29-35

clinical applications requires appropriate software and technical infrastructure, which may pose logistical and financial challenges for healthcare institutions. The retrospective single-center design of our study limits it. Future research should validate these findings in larger, prospective cohorts to investigate the potential of CF for real-time clinical decision-making. Researchers and clinicians can assign facial features to parameters based on their study, thereby enabling highly adaptable data representation. This flexibility may limit study comparability due to the lack of a standardized feature-variable mapping method. Each study must explicitly define its parameter assignments, as a map legend does, because there is no universal framework. Implementation in clinical settings may require additional training to ensure that healthcare professionals accurately interpret visualized data.

CONCLUSION

CF visualizes multivariate data by mapping variables to facial features, thereby facilitating pattern recognition and comparison. They work well in clinical and emergency medicine. We found that CF may benefit emergency medicine, particularly during mass-casualty events in which rapid patient assessment is essential. It visualizes complex metabolic changes to improve triage, prioritize treatment, and monitor patients. More standardization and validation of CF are needed to help healthcare professionals manage critical cases.

Ethics

Ethics Committee Approval: The study was approved by the İzmir Katip Çelebi University Ethics Committee on March 21, 2024, under decision number 0175.

Informed Consent: This research was conducted as a retrospective study.

Footnotes

Author Contributions: Concept - E.S.B., M.G.E.; Design - E.S.B.; Data Collection and/or Processing - M.A.T., E.K.; Analysis and/or Interpretation - M.G.E.; Literature Search - M.A.T., E.K.; Writing - E.S.B., M.G.E.

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

Financial Disclosure: The authors report that no financial support was received for this study

REFERENCES

1. Bennett IL Jr, Cary Fh, Mitchell GL Jr, Cooper MN. Acute methyl alcohol poisoning: a review based on experiences in an outbreak of 323 cases. Medicine (Baltimore). 1953; 32: 431-63.
2. Barceloux DG, Bond GR, Krenzelok EP, Cooper H, Vale JA; American Academy of Clinical Toxicology Ad Hoc Committee on the Treatment Guidelines for Methanol Poisoning. American Academy of Clinical Toxicology practice guidelines on the treatment of methanol poisoning. J Toxicol Clin Toxicol. 2002; 40: 415-46.
3. Kruse JA. Methanol poisoning. Intensive Care Med. 1992; 18: 391-7.
4. Hizarci B, Erdoğan C, Karaaslan P, Unlukaplan A, Oz H. Transdermal methyl alcohol intoxication: a case report. Acta Derm Venereol. 2015; 95: 740-1.
5. Bitar ZI, Ashebu SD, Ahmed S. Methanol poisoning: diagnosis and management. A case report. Int J Clin Pract. 2004; 58: 1042-4.
6. Brent J. Fomepizole for ethylene glycol and methanol poisoning. N Engl J Med. 2009; 360: 2216-23.
7. Hayreh MS, Hayreh SS, Baumbach GL, Cancilla P, Martin-Amat G, Tephly TR, et al. Methyl alcohol poisoning III. Ocular toxicity. Arch Ophthalmol. 1977; 95: 1851-8.
8. Wedge MK, Natarajan S, Johanson C, Patel R, Kanji S. The safety of ethanol infusions for the treatment of methanol or ethylene glycol intoxication: an observational study. CJEM. 2012; 14: 283-9.
9. Bestic M, Blackford M, Reed M. Fomepizole: a critical assessment of current dosing recommendations. J Clin Pharmacol. 2009; 49: 130-7.
10. Bouchard J, Lavergne V, Roberts DM, Cormier M, Morissette G, Ghannoum M. Availability and cost of extracorporeal treatments for poisonings and other emergency indications: a worldwide survey. Nephrol Dial Transplant. 2017; 32: 699-706.
11. Lim CS, Bryant SM. Forgoing the folate?-Contemporary recommendations for methanol poisoning and evidence review. Am J Ther. 2016; 23: e850-4.
12. Chernoff, H. The use of faces to represent points in K-dimensional space graphically. J Am Stat Assoc. 1973; 68: 361-8.
13. Çelik S, Akdamar E. Big data and data visualization [Büyük veri ve veri görselleştirme]. Akademik Bakış Dergisi. Uluslararası Hakemli Sosyal Bilimler E-Dergisi. 2018; 65: 253-64. Turkish.
14. Lee MD, Reilly RE. An empirical evaluation of Chernoff faces, star glyphs, and spatial visualisations for binary data. InVis.au. 2003; 24: 1-10.
15. Carp J, Park J, Polk TA, Park DC. Age differences in neural distinctiveness revealed by multi-voxel pattern analysis. Neuroimage. 2011; 56: 736-43.
16. De Los Reyes A, Augenstein TM, Aldao A, Thomas SA, Daruwala S, Kline K, et al. Implementing psychophysiology in clinical assessments of adolescent social anxiety: use of rater judgments based on graphical representations of psychophysiology. J Clin Child Adolesc Psychol. 2015; 44: 264-79.
17. Akbulut S, Çelebi N. Diagnosis and prediction of Alzheimer's disease by Chernoff faces. In: 2023 Innovations in Intelligent Systems and Applications Conference (ASYU). 2023; 303-8.
18. De Los Reyes A, Aldao A, Qasmieh N, Dunn EJ, Lipton MF, Hartman C, et al. Graphical representations of adolescents' psychophysiological reactivity to social stressor tasks: reliability and validity of the Chernoff face approach and person-centered profiles for clinical use. Psychol Assess. 2017; 29: 422-34.
19. Phillipou G. Graphical display of capillary blood glucose data using Chernoff faces. Diabet Med. 1992; 9: 293-4.
20. He J, Lin J, Duan M. Application of machine learning to predict acute kidney disease in patients with sepsis associated acute kidney injury. Front Med (Lausanne). 2021; 8: 792974.
21. Shi J, Han H, Chen S, Liu W, Li Y. Machine learning for prediction of acute kidney injury in patients diagnosed with sepsis in critical care. PLoS One. 2024; 19: e0301014.
22. Vagliano I, Chesnaye NC, Leopold JH, Jager KJ, Abu-Hanna A, Schut MC. Machine learning models for predicting acute kidney injury: a systematic review and critical appraisal. Clin Kidney J. 2022; 15: 2266-80.
23. Rahu M. Graphical representation of cancer incidence data: Chernoff faces. Int J Epidemiol. 1989; 18: 763-7.
24. Lott JA, Durbridge TC. Use of Chernoff faces to follow trends in laboratory data. J Clin Lab Anal. 1990; 4: 59-63.
25. Battagel JM. Chernoff faces: an orthodontic application. Br J Orthod. 1995; 22: 135-44.
26. Dunn EJ, Aldao A, De Los Reyes A. Implementing physiology in clinical assessments of adult social anxiety: a method for graphically representing physiological arousal to facilitate clinical decision-making. J Psychopathol Behav Assess. 2015; 37: 587-96.
27. Yurtsever G, Arıkan C, Acar H, Sorgun O, Bora ES. Methanol poisoning in the emergency department: a retrospective study. J Health Sci Med 2022; 5: 949-53.
28. Soete G, Corte W. On the perceptual salience of features of Chernoff faces for representing multivariate data. Applied Psychological Measurement. 1985; 9: 275-80.

DOI: 10.4274/jarem.galenos.2026.34713
J Acad Res Med 2026;16(1):36-42

Effectiveness of the Mini-clinical Evaluation Exercise in Otorhinolaryngology Training

Sezen Milli Avtan

University of Health Sciences Türkiye, Hamidiye Faculty of Medicine, Department of Medical Education, İstanbul, Türkiye

Cite this article as: Milli Avtan S. Effectiveness of the mini-clinical evaluation exercise in otorhinolaryngology training. J Acad Res Med. 2026;16(1):36-42

ABSTRACT

Objective: This study aimed to evaluate the impact of the mini-clinical evaluation exercise (mini-CEX) on clinical competence and educational outcomes in otorhinolaryngology (ENT) training across different educational levels.

Methods: A systematic review and meta-analysis was conducted in accordance with Preferred Reporting Items for Systematic reviews and Meta-analyses 2020 guidelines. PubMed/MEDLINE, Scopus, Web of Science, and Google Scholar were searched for studies assessing the mini-CEX in ENT education. Eligible studies included pre-post or observational designs reporting quantitative outcomes. Random-effects meta-analysis was used to calculate standardized mean differences (SMD) with 95% confidence intervals (CIs). Risk of bias was assessed using NIH and Newcastle-Ottawa tools.

Results: Five studies, including 506 trainees and over 2,600 mini-CEX encounters, were included. Mini-CEX implementation resulted in a large, statistically significant improvement in overall clinical competence (SMD=1.29; 95% CI 1.01-1.57; $p<0.001$), with moderate heterogeneity ($I^2=52\%$). Improvements were observed across all competency domains, with the largest effects in physical examination and history-taking skills. The assessment was feasible within routine clinical practice, with high levels of learner engagement and faculty acceptance.

Conclusion: Mini-CEX is an effective and practical formative assessment tool for improving clinical competence in ENT training. Its integration into competency-based curricula may support sustained improvements in clinical performance.

Keywords: Mini-CEX, otorhinolaryngology, ENT education, workplace-based assessment, clinical competence, meta-analysis

INTRODUCTION

Medical education requires students to learn clinical abilities which include surgical competencies as essential elements for their development (1). Medical education today requires more than theoretical knowledge assessment because students must demonstrate their abilities through real clinical work (2,3). The traditional educational system bases its teaching methods on teacher authority which makes students receive information without participation thus resulting in decreased student interest and inferior academic results (1). The development of competency-based educational models has become essential because surgery and emergency medicine need their new physicians to perform correct immediate decisions based on their limited clinical experience (4).

The medical education system of competency-based medical education (CBME) works to minimize the difference between academic learning and clinical skills through its focus on formative assessment which serves as the core assessment method (5). The assessment system includes workplace-based assessments (WPBA) tools that allow direct observation of trainees during their work in real clinical environments and help overcome the weaknesses of conventional assessment approaches. The mini-clinical evaluation exercise (mini-CEX) stands as one of the most commonly used assessment tools which the American Board of Internal Medicine developed through its mini-CEX program (1,6). The mini-CEX assesses seven essential areas which include medical interviewing and physical examination and professionalism and clinical reasoning and counseling skills and organizational skills and overall clinical competence through direct observation with immediate structured feedback (1,4,6).

The literature contains substantial evidence that mini-CEX implementation leads to improved clinical competence across various medical specialties. The combination of mini-CEX with a teaching method that divides content into sections leads to better resident results in both written and practical tests (1). The combination of mini-CEX with checklist-based assessment tools in emergency medicine and intensive care rotations has shown

ORCID ID of the author: S.M.A. 0000-0003-2528-8760

Corresponding Author: Sezen Milli Avtan, Asst. Prof.;
E-mail: sezenavtan@gmail.com

Received Date: 04.03.2026 **Accepted Date:** 10.04.2026
Epub: 13.04.2026
Publication Date: 28.04.2026

J Acad Res Med 2026;16(1):36-42

success in enhancing both diagnostic precision and procedural adherence (4). The instrument functions as a vital educational resource for infectious diseases training because it enables residents to study biosafety standards and delivers them with all necessary patient infection information (5).

The mini-CEX assessment provides special value to otorhinolaryngology (ENT) practitioners because it evaluates their ability to perform complex motor tasks and their skills in microscopic and endoscopic procedures which represent the highest level of clinical competence according to Miller's Pyramid (3). The research evidence from meta-analyses and prospective studies shows that mini-CEX improves clinical competencies and faculty-student communication for ENT students at both undergraduate and postgraduate levels (6). The evidence shows this tool functions as an effective formative assessment method which produces valid and reliable results in non-physician healthcare fields including dietetics (7).

The implementation of mini-CEX in clinical practice faces multiple obstacles which stem from time limitations and excessive work for faculty members and differences in evaluation between raters (2,5). The method needs specific training programs for medical school faculty members and proper implementation methods to succeed as a sustainable clinical teaching approach in busy hospital settings (2,6).

The objective of this meta-analysis is to assess how the mini-CEX affects clinical competence, skill development, and educational outcomes in ENT residency training programs. The research investigates how mini-CEX affects clinical competence in ENT education at three educational stages: undergraduate, internship, and residency programs and their impact on six competency areas: history taking, physical examination, clinical decision-making, communication, professionalism, and organization/efficiency. The research will use an extensive analytical method to study secondary educational results including feasibility and educator-learner satisfaction. The research results will help create better, more sustainable educational systems that support CBME reform initiatives.

METHODS

Study Design and Reporting Standards

The research design consisted of a systematic review and meta-analysis that assessed the effectiveness of mini-CEX in ENT training programs. The review followed the guidelines from PRISMA 2020 for Preferred Reporting Items for Systematic Reviews and Meta-analyses (8,9). The researchers created a protocol before starting their research, while following all established criteria for participant selection, measurement procedures, and data processing techniques.

Eligibility Criteria

The research team selected studies for evaluation based on their assessment of mini-CEX applications in ENT training programs. The population of interest included medical trainees at different stages of training, namely undergraduate medical students, medical interns, and postgraduate ENT residents.

The research focused on using the mini-CEX as an assessment tool that healthcare providers used to evaluate their performance during clinical work in ENT facilities providing outpatient, inpatient, and emergency services.

The research included baseline data from before-after studies, as well as standard assessment methods that used comparison groups.

The research team obtained its main results through changes in mini-CEX scores and global competency-rating assessments, which showed how mini-CEX implementation affected clinical competence. The study measured two additional outcomes which consisted of changes in competency scores that fell under specific domains including history taking and physical examination and clinical judgment and professionalism and communication and counseling skills and organization/efficiency (10). The research team monitored two main outcome variables: trainee and teacher contentment with mini-CEX implementation, and the length of assessment and feedback sessions and the number of mini-CEX encounters per trainee.

The research included all studies that used prospective and retrospective observational methods, and interventional educational designs with pre-post structures and control-group options. The research team eliminated all studies that were case reports, narrative reviews, editorials, conference abstracts lacking measurable quantitative information or that focused on subjects other than ENT training.

Information Sources and Search Strategy

The research team performed an extensive search of the PubMed/MEDLINE, Scopus, Web of Science, and Google Scholar databases. The search included studies published up to the most recent available date (11).

The following keywords and medical subject headings were used in various combinations: the search terms included "mini-CEX" OR "Mini Clinical Evaluation Exercise" AND "Otorhinolaryngology" OR "ENT" AND "medical education" OR "clinical competence" OR "formative assessment".

The research team conducted a manual review of reference lists of relevant articles to identify additional eligible studies.

Study Selection

The reference management software received all retrieved records, which allowed researchers to eliminate duplicate entries from the dataset. Title and abstract screening was performed independently by two reviewers to assess potential eligibility. The research team performed a full-text evaluation of all articles that were potentially relevant to determine which studies met the established inclusion criteria. Reviewers at both stages needed to reach a common understanding through discussion of all points.

J Acad Res Med 2026;16(1):36-42

Data Extraction

Data extraction was conducted independently by two reviewers using a predefined, standardized data-extraction form. The research team extracted the following data from each included study: the authors' publication year and the country and location where the study was conducted. The research design of this study followed a specific method described by the authors. Participants were categorized into specific groups based on their training level and demographic characteristics. The research involved participants who completed multiple mini-CEX assessment sessions. The assessment included multiple domains that were evaluated using specific scoring systems. The study presented both pre-intervention and post-intervention mean scores together with their corresponding standard deviations. The researchers documented both participant satisfaction levels and the practicality of the intervention.

The research team used established statistical methods to convert medians, ranges, and graphical data into means and standard deviations for quantitative analysis.

Risk of Bias Assessment

Two reviewers used validated tools that meet the requirements of educational intervention research to conduct independent assessments of the methodological quality of the included studies. The National Institutes of Health Quality Assessment Tool for Before-After Studies assessed pre-post design studies, and the Newcastle-Ottawa Scale, with educational research modifications, assessed observational studies.

The assessment results enabled researchers to determine the risk of bias for each study, which they classified into three categories: low, moderate, and high risk of bias. The team members discussed and reached agreement on the final risk-of-bias judgments for all studies.

Statistical Analysis

This research used a random-effects model in the meta-analysis to account for expected differences between studies arising from their methodological and educational variations. The research team combined continuous data using standardized mean differences (SMDs; Hedges' g), including 95% confidence intervals (CIs) to enable comparison across studies despite varying scoring systems.

The assessment of statistical heterogeneity used Cochran's Q test to identify variations, while the I^2 statistic evaluated the degree of heterogeneity, with higher values indicating more substantial differences. The study conducted subgroup analyses based on training level when sufficient data were available to analyze three trainee groups: undergraduate medical students, medical interns, and postgraduate residents.

The researchers performed sensitivity analyses to determine how each study affected the combined effect estimates by removing one study at a time. Assessment of publication bias was conducted by visual inspection of funnel plots when sufficient studies were available to perform the evaluation.

All statistical analyses were conducted using standard meta-analysis software and statistical significance was defined as a two-sided p-value <0.05.

Subgroup analyses were planned based on training level (undergraduate, internship, and residency) when sufficient data were available.

As this study utilized data from previously conducted research, no additional ethical approval was required for the present analysis.

RESULTS

The study selection process is illustrated in the flow diagram (Figure 1).

Study Selection and Included Studies

The research team conducted a systematic screening and eligibility assessment and selected five studies that assessed the effectiveness of the mini-CEX for ENT training. The research team selected these five studies for their meta-analysis because they met all the criteria established for inclusion. The research included participants who were medical trainees at various levels of their education, including 444 undergraduate medical students, 50 ENT interns, and 12 ENT postgraduate residents. The study included 506 trainees who participated in more than 2600 mini-

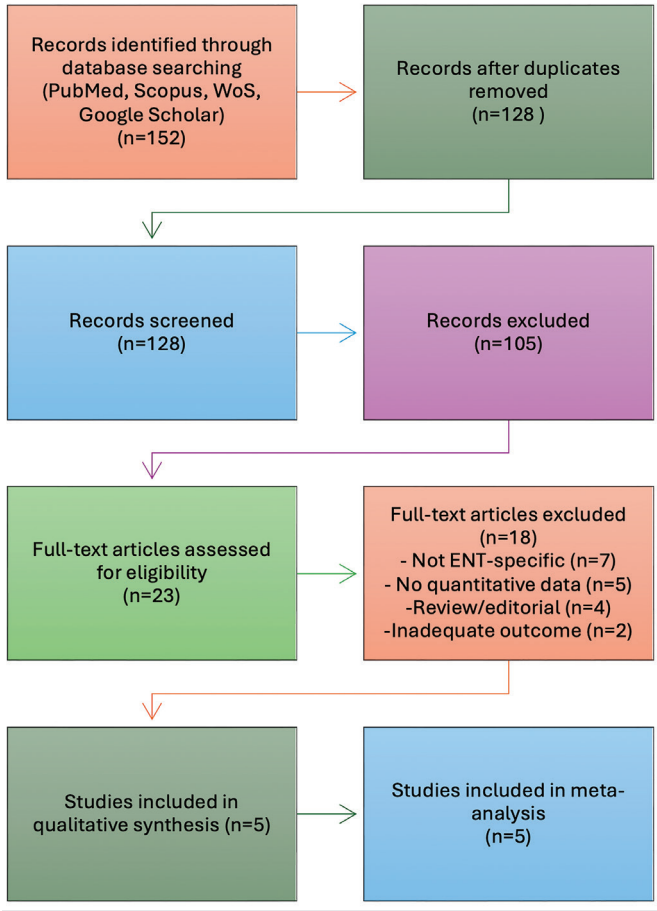

Figure 1. Flow diagram of the study selection process
WoS: Web of Science, ENT: Otorhinolaryngology

J Acad Res Med 2026;16(1):36-42

CEX assessments that took place in outpatient, inpatient, and emergency departments.

Study Characteristics

The included studies were published between 2015 and 2025 and originated in South Asia, primarily in India and Pakistan. The mini-CEX assessment was evaluated in three research studies that focused on undergraduate medical education programs, while two additional studies examined ENT interns and ENT postgraduate residency training. The research design used in all studies was a pre-post educational assessment that incorporated mini-CEX as either a formative or a WPBA tool. The mini-CEX evaluation required trainees to participate in two to six patient interactions, each followed by immediate structured feedback (Table 1).

Primary Outcome: Overall Clinical Competence

All included studies demonstrated a statistically significant improvement in overall clinical competence following mini-CEX implementation. Random-effects meta-analysis revealed a large effect size favoring mini-CEX (SMD=1.29; 95% CI 1.01-1.57; $p<0.001$). Between-study heterogeneity was moderate ($I^2=52\%$), reflecting variations in trainee level, assessment scales, and educational settings.

Subgroup Analysis by Training Level

Subgroup analysis was planned to compare the impact of mini-CEX across different training levels (undergraduate students, interns, and residents). However, due to the limited number of included studies and insufficient reporting of stratified outcome data, a formal subgroup meta-analysis was not feasible.

Descriptive comparisons suggest that the magnitude of improvement was greater for interns and postgraduate trainees than for undergraduate students, as reflected by larger absolute score changes. These findings should be interpreted with caution and require confirmation in future studies with more homogeneous and stratified datasets.

Pre-post Improvement in Overall Competence Scores

Across studies reporting quantitative pre-post data, improvements in overall competence scores were consistently observed. The magnitude of improvement varied by scoring scale and trainee level, with the largest absolute gains observed in intern and postgraduate cohorts (Table 2).

Domain-specific Competency Outcomes

Pooled analyses of domain-specific outcomes demonstrated significant improvements across all assessed competency domains. The largest effects were observed in physical examination and history-taking skills, while professionalism and organizational efficiency showed moderate but consistent gains (Table 3).

Feasibility and Acceptability Outcomes

Mini-CEX was consistently reported to be feasible and well accepted by both trainees and faculty in all included studies. The mean duration of a single mini-CEX encounter was between 11 to 16 minutes, with an additional 5 to 10 minutes allocated for structured feedback. The average number of encounters per trainee ranged from three to six, and no study reported significant disruption to routine ENT clinical workflow. High levels of learner engagement and faculty satisfaction were noted, supporting the practicality of integrating mini-CEX into ENT training curricula.

As shown in Table 4, implementation of the mini-CEX was associated with statistically significant improvements in clinical competence among ENT interns and residents. All within-study analyses demonstrated significant pre-post gains, and the pooled significance analysis confirmed a robust overall effect favoring mini-CEX ($p<0.001$) (Table 4).

DISCUSSION

The study employs a systematic review and meta-analysis to demonstrate that mini-CEX implementation results in substantial, statistically significant improvements in clinical competence among ENT trainees across educational stages. This study confirmed earlier educational research showing that direct observation with structured feedback produces the best learning outcomes for students acquiring clinical competencies in real-world practice. The research provides quantitative evidence of improvements in ENT education that exceed those of previous studies by presenting precise results for the specific field of ENT education that requires advanced psychomotor and endoscopic examination skills.

The physical examination and history-taking competencies demonstrated the greatest improvement through domain-level assessments because these fundamental skills enable doctors to make correct diagnoses and choose appropriate procedures in ENT practice. The educational value of moderate gains in

Table 1. Characteristics of studies included in the meta-analysis

Author (year)	Country	Training level	Sample size	Study design	Mini-CEX encounters	Outcome measure
Sowmya et al. (3) (2025)	India	ENT postgraduate	12	Prospective interventional	3	Total mini-CEX score
Sivaraman et al. (12) (2024)	India	Undergraduate (ENT)	149	Retrospective pre-post	≥5	Domain scores
Shafqat et al. (13) (2022)	Pakistan	Undergraduate	199	Prospective pre-post	2	Clinical performance
Saeed et al. (14) (2015)	Pakistan	Undergraduate	96	Longitudinal observational	~10/year	Skill progression
Gurumani et al. (15) (2025)	India	ENT interns	50	Prospective pre-post	3.2±0.6	Overall competence

ENT: Otorhinolaryngology, mini-CEX: Mini-clinical evaluation exercise

J Acad Res Med 2026;16(1):36-42

professionalism, communication, and organizational efficiency remains important because these competencies directly affect patient safety, healthcare quality, and interprofessional teamwork. The research findings validate the CBME model that uses workplace assessments to help students develop their skills through progressive learning.

Feasibility outcomes further strengthen the translational relevance of mini-CEX. The research showed that assessment time remained short enough to fit within standard medical practice, while participants in both groups expressed strong interest in the assessment. The research findings solve one major problem with WPBA, namely time consumption and demonstrate that proper implementation methods allow educational programs to become permanent parts of clinical training in busy medical facilities.

The research data show that statistical heterogeneity exists at a moderate level because different trainee levels, scoring systems, and educational settings produced these varying results. The results provide strong evidence because sensitivity analyses and all studies indicate the same direction of effect. The upcoming multicenter trials will determine the optimal number of patient interactions and the optimal feedback delivery method that support physician competence, through their implementation of mini-CEX protocols and their tracking of patient outcomes at different time points.

The study has several limitations that should be acknowledged. The review process becomes difficult because there are not enough ENT-specific studies, which makes it impossible to apply research findings across different population groups. The majority of included studies employed pre-post educational designs which lacked randomized controls thus creating potential bias because of participant development and additional training activities (3,12-15). Third, the studies conducted in South Asia show a geographic concentration, which reduces the generalizability of findings to other healthcare systems worldwide. Research results show that mini-CEX produces equivalent educational outcomes across diverse healthcare environments despite its limitations.

This meta-analysis demonstrates, based on statistical evidence, that mini-CEX improves ENT students' clinical skills and has potential for use in standard medical practice. The implementation of structured WPBA in ENT curricula provides an evidence-based approach to improving medical training through practical assessment methods that support competency-based education. The development of official mini-CEX implementation protocols should be the focus of future research for tracking surgeons' development and conducting location-based tests that will enhance the tool's effectiveness in surgical training.

Study Limitations

Several limitations of this study should be acknowledged. First, the number of eligible studies that specifically investigated the mini-CEX in ENT training was relatively limited. Consequently, the pooled analysis was based on a small number of studies, which may restrict the generalizability of the findings.

Second, most of the included studies employed pre-post educational designs without randomized control groups. Although these designs are commonly used in medical education research, they may introduce sources of bias, including maturation effects, concurrent educational interventions, and uncontrolled contextual factors that could influence observed improvements in clinical competence.

Third, the majority of the included studies were conducted in South Asian countries, particularly India and Pakistan. This geographical concentration may limit the external validity of the results, as

Table 2. Pre- and post- mini-CEX overall competence scores

Study (year)	Pre- mini-CEX mean ± SD	Post-mini-CEX mean ± SD	Mean difference
Gurumani et al. (15) (2025)	5.0±0.8	7.2±0.7	+2.2
Sowmya et al. (3) (2025)	36.1±3.2	47.4±2.8	+11.3
Saeed et al. (14) (2015)	6.78±1.3	7.49±1.4	+0.71

mini-CEX: Mini-clinical evaluation exercise, SD: Standard deviation

Table 3. Pooled standardized effects by competency domain

Competency domain	Pooled SMD	95% CI	Effect size interpretation
History taking	1.25	0.96-1.54	Large
Physical examination	1.38	1.06-1.70	Large
Clinical judgment	1.18	0.87-1.49	Large
Communication/ counseling	1.07	0.74-1.40	Moderate-large
Professionalism	0.91	0.57-1.25	Moderate
Organization/efficiency	1.02	0.70-1.34	Moderate-large

SMD: Standardized mean differences, CI: Confidence interval

Table 4. Statistical significance of mini-CEX-associated improvements in clinical competence in otorhinolaryngology training

Study (year)	Training level	Comparison	Statistical test	Test statistic	p-value
Gurumani et al. (15) (2025)	ENT interns (n=50)	5.0±0.8→7.2±0.7	Paired t-test	t (49)=20.6	<0.001
Sowmya et al. (3) (2025)	ENT residents (PGY-1, n=8)	36.1→47.4	Repeated-measures ANOVA	F=21.799	<0.00001
Pooled significance*	Interns+residents	-	Fisher's combined probability test	χ^2=34.7	<0.001

*Pooled significance was calculated by combining the p-values of the included studies using Fisher's combined probability test, ENT: Otorhinolaryngology, PGY: Program for first postgraduate year, ANOVA: Analysis of variance

J Acad Res Med 2026;16(1):36-42

educational structures, clinical training environments, and assessment cultures may differ across healthcare systems and regions.

Fourth, variability in training levels (undergraduate students, interns, and residents), mini-CEX implementation protocols, scoring scales, and the number of assessment encounters may have contributed to the moderate statistical heterogeneity observed in the meta-analysis. Although a random-effects model was used to account for such variability, these methodological differences should be considered when interpreting the pooled estimates.

Finally, while this review demonstrates improvements in clinical competence scores, most studies assess short-term educational outcomes. Long-term effects of mini-CEX implementation such as sustained competency development, impact on clinical performance over time, and potential effects on patient care outcomes remain insufficiently explored.

Future research should therefore include larger multicenter studies with standardized mini-CEX implementation protocols and longer follow-up periods to better evaluate the durability and broader educational impact of this assessment method in ENT training.

CONCLUSION

The findings across diverse medical and health disciplines, including surgery, pediatrics, emergency medicine, infectious diseases, ENT, and dietetics, consistently demonstrate that the mini-CEX is an effective and feasible formative assessment instrument. The present study indicates that the structured implementation of mini-CEX significantly enhances trainees' clinical competencies, with substantial improvements observed in medical interviewing, physical examination, and clinical reasoning skills.

A principal strength of the mini-CEX lies in its alignment with the principles of CBME and in its capacity to assess performance at the "does" level of Miller's Pyramid within authentic clinical contexts. Evidence suggests that the tool not only enhances practical performance but also correlates positively with outcomes in traditional summative examinations, indicating that the competencies and self-confidence acquired are transferable to formal assessment settings. Moreover, innovative adaptations such as integrating mini-CEX with segmented teaching strategies or structured checklists have demonstrated superior outcomes by effectively bridging the gap between theoretical knowledge acquisition and procedural standardization.

Despite high satisfaction levels reported by both faculty and trainees, the sustainability of mini-CEX implementation is challenged by systemic barriers, primarily time constraints in high-volume clinical environments and the need for structured faculty development. To overcome these limitations, institutions should consider adopting distributed assessment models that integrate evaluations into routine clinical workflows, in conjunction with digital scoring platforms, to minimize administrative burden.

The mini-CEX serves as a transformative mechanism for mentored clinical learning, promoting continuous student-faculty interactions and timely, constructive feedback. Although its construct validity has been supported, including in non-physician health disciplines such as dietetics, future research should prioritize multicenter trials with larger sample sizes to assess the long-term durability of competency gains and the direct impact of those gains on patient care outcomes. The systematic integration of mini-CEX into undergraduate and postgraduate curricula represents a critical step toward enhancing the quality of clinical education and preparing a highly competent healthcare workforce.

Ethics

Ethics Committee Approval: As this study utilized data from previously conducted research, no additional ethical approval was required for the present analysis.

Informed Consent: This study does not involve research conducted on human participants or animals. The manuscript is based on the analysis and synthesis of findings from previously published studies.

Footnotes

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

REFERENCES

1. Liu C, Ren M, Luo C, Asfandyar K, Liu J, Chen D, et al. Application of mini-CEX evaluation combined with segmented teaching approach in the training of surgical clinical practice skills for resident physicians. J Surg Educ. 2025; 82: 103497.
2. Dave H. Evaluating the utility of mini-CEX as a formative assessment tool for paediatric postgraduate education: a mixed-method educational interventional study from Gujarat, India. J Clin Diagn Res. 2025; 19: PC01-5
3. Sowmya S, Kumar NM, Kavyashree R, Ayub I. Mini clinical evaluation exercise (mini-CEX) as a formative assessment tool in ENT residency: a pilot study. Indian J Otolaryngol Head Neck Surg. 2025; 78.
4. Cao XG, Hu JX, Jian HC, Zhu XF, Meng H, Shao M. The combined application of mini-CEX and check-list scales in enhancing clinical competence among emergency and critical care residents: a comparative study. Front Med (Lausanne). 2025; 12: 1492206.
5. Wang X, Li H, Liu Y, Zhang D, Cai D, Zhong S. Enhancing clinical competency in infectious disease training: a longitudinal study of mini-CEX implementation for medical interns. Front Med (Lausanne). 2025; 12: 1582218.
6. Sharma R, Rana AK, Verma M, Singh AD. Implementing mini-clinical evaluation exercise (mini-CEX) and its efficacy as a formative assessment tool in the otorhinolaryngology (ORL) undergraduate medical program. Indian J Otolaryngol Head Neck Surg. 2025; 77: 5412-7.
7. Moe AK, Dahl H, Kvernenes M, Drotningsvik A, Rosendahl-Riise H. Validating an assessment and feedback instrument for use in dietetics education: construct validity of the mini clinical evaluation exercise (Mini-CEX). BMC Med Educ. 2025; 25: 1693.
8. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Int J Surg. 2021; 88: 105906.
9. Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ. 2021; 372: n160.
10. Norcini JJ, Blank LL, Duffy FD, Fortna GS. The mini-CEX: a method for assessing clinical skills. Ann Intern Med. 2003; 138: 476-81.
11. Kirkik D, Demircioglu S, Taş SK. TLR2 variants and *Helicobacter pylori:* revisiting a controversial link. Ups J Med Sci. 2025; 130.
12. Sivaraman G, Lakshmanan J, Alexander A, Mahalakshmy T, Raja K, Sabharisan P, et al. Use of mini-CEX as formative assessment tool in the

J Acad Res Med 2026;16(1):36-42

training of undergraduate medical students in ENT situation analysis and the way forward. Indian J Otolaryngol Head Neck Surg. 2024; 76: 2698-703.

13. Shafqat S, Tejani I, Ali M, Tariq H, Sabzwari S. Feasibility and effectiveness of mini-clinical evaluation exercise (mini-CEX) in an undergraduate medical program: a study from Pakistan. Cureus. 2022; 14: e29563.
14. Saeed N, Tariq N, Jaffery T. Mini-CEX (clinical evaluation exercise) as an assessment tool at Shifa College of Medicine. Rawal Medical Journal. 2015; 40: 220-4.
15. Gurumani S, Dilara K, Deepthi E, Arun Maran S, Subbu H, Amrisha A. Mini-clinical examination (mini-Cex) as a formative assessment tool for ENT interns at ANIIMS. Journal of Chemical Health Risks. 2025; 15: 2943-50.

DOI: 10.4274/jarem.galenos.2026.59455
J Acad Res Med 2026;16(1):43-53

Comparison of Nasalance Values Obtained with Nasometer-II and Praat-assisted Nasalance Meter

Mehmet Akif Kılıç[1], Muhammed Selman Babar[2], Göksu Tarakçı[2]

[1]Lokman Hekim İstanbul Hospital, Voice-Timbre-Speech-Language Unit, İstanbul, Türkiye
[2]Üsküdar University Faculty of Health Science, Department of Speech and Language Therapy, İstanbul, Türkiye

Cite this article as: Kılıç MA, Babar MS, Tarakçı G. Comparison of nasalance values obtained with Nasometer-II and Praat-assisted Nasalance Meter. J Acad Res Med. 2026;16(1):43-53

ABSTRACT

Objective: This study aimed to compare nasalance values obtained using the Praat-assisted Nasalance Meter (PANM) and the Nasometer II model (NMII) 6450 by examining the effects of hardware and software configurations and evaluating the strength of the linear relationships between measurements derived from these systems.

Methods: A total of 60 healthy participants (17 men, 43 women) aged 18-23 years were included. Acoustic signals were recorded and analyzed separately using PANM and NMII hardware and software configurations. PANM measurements were obtained using 80-1000 Hz and 300-750 Hz bandpass filters, whereas NMII measurements were obtained using a 300-750 Hz bandpass filter. Data were analyzed using Pearson correlation and two-way repeated-measures analysis of variance (ANOVA).

Results: Positive correlations were observed between nasalance scores obtained from different hardware and software configurations ($p<0.05$). Repeated-measures ANOVA revealed significant main effects of both software and hardware across speech materials. NMII software yielded higher nasalance scores than PANM software, whereas PANM hardware produced higher scores than NMII hardware in most speech materials. Similar patterns were observed across different types of speech materials.

Conclusion: The findings demonstrate that nasalance scores obtained from the PANM and NMII systems vary according to hardware and software configurations. Although measurements showed linear associations, the consistent differences observed across speech materials indicate that system-specific technical properties influence score magnitude. Therefore, hardware and software configurations should be carefully considered when interpreting and reporting nasalance scores in clinical and research settings.

Keywords: Velopharyngeal dysfunction, nasal resonance, nasalance measurement, Praat

INTRODUCTION

Nasal measurement is a widely used technique for clinical evaluation of nasal resonance disorders (1). The nasalance score obtained from this measurement is calculated by dividing the nasal acoustic energy by the sum of nasal and oral acoustic energies (also called "total acoustic energy"), and multiplying this by 100 (2). In this context, different systems have different software and hardware to measure nasalance [e.g., Nasometer 6200 (Kay Elemetrics Corp.), Nasometer II 6400 (KayPENTAX, Inc.), Nasometer II 6450 (KayPENTAX, Inc.), Nasometer 6500 (KayPENTAX, Inc.), NasalView (Tiger DRS, Inc.), OroNasal Nasality System (Glottal Enterprises, Inc.)] (1,3).

The aforementioned systems, which play an active role in the diagnostic and therapeutic processes of nasal resonance disorders are usually expensive and/or relatively difficult to access. This may be a challenge for clinicians and researchers dealing with nasal resonance disorders. Therefore, the Praat-assisted Nasalance Meter (PANM) system, which has free software, is easily accessible and requires only low-cost materials was developed by the first author and was standardized by Kılıç et al. (4).

The PANM system consists of hardware with earphones acting as microphones on both sides (upper and lower) of the separator plate to measure nasalance and software to calculate the nasalance score in a computer environment. The software of the PANM system is designed as the Praat (5) plugin and has an easy-to-use interface, which also runs on Praat after installation. Headphones are used instead of microphones to minimize costs and facilitate the accessibility of PANM hardware. Dynamic headphones use the electromagnetic field created by a vibrating diaphragm and a copper coil around a magnet. As a result, they have the same working principle as dynamic microphones. Thus,

ORCID IDs of the authors: M.A.K. 0000-0003-0706-7032; M.S.B. 0000-0001-7974-5774; G.T. 0000-0001-6123-8395

Corresponding Author: Mehmet Akif Kılıç, Prof. Dr.;
E-mail: makilic@yahoo.com
Received Date: 29.01.2026 **Accepted Date:** 13.04.2026
Publication Date: 28.04.2026

J Acad Res Med 2026;16(1):43-53

headphones, in addition to being cost-effective, can also function as microphones that allow minimal sound transmission between microphones, especially for nasalance measurement (6). Another important aspect of this research is that the pair of microphones used in this study is a matched pair. For this reason, the researchers tested their assumption that a matched pair of headphones can maintain this property when used as a microphone, and when the frequency response curves were analysed, they found that both microphones had a similar frequency response (Figure 1). In this context, it should not be forgotten that an audio interface capable of stereo recording is required to record the acoustic energy from the nose and mouth separately for the PANM system to work correctly.

The aim of this study was to compare nasalance values obtained using the Nasometer II and the PANM by examining the effects of hardware and software configurations on nasalance scores and by evaluating the strength of the linear relationships between measurements derived from these systems.

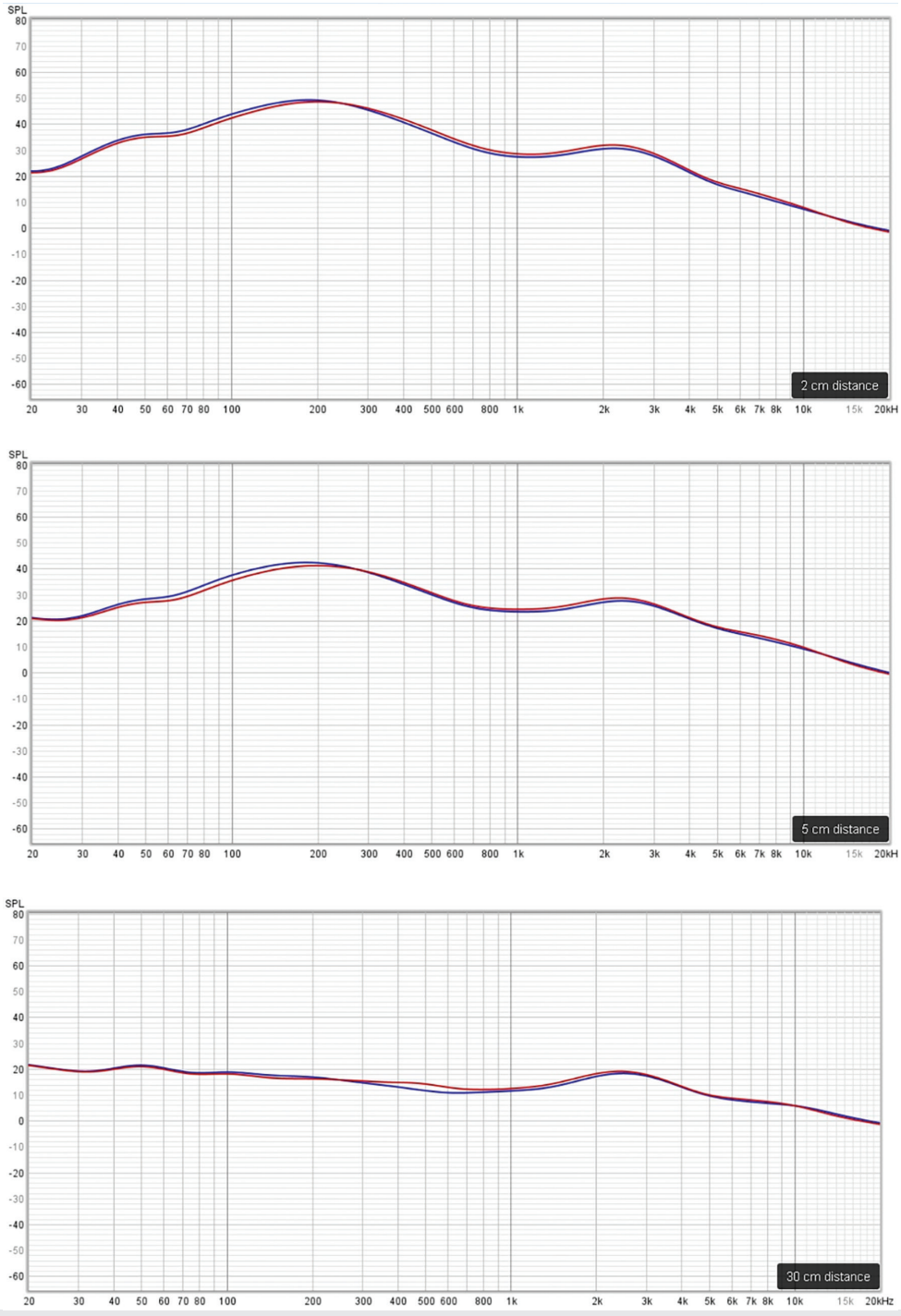

Figure 1. Frequency response curves of microphones at 2, 5 and 30 cm distance

J Acad Res Med 2026;16(1):43-53

METHODS

Participants

A total of 60 adults, consisting of 43 women and 17 men aged 18-23 years, were included in this study. The criteria for inclusion in the study were (1) not having hearing loss, (2) not having a physiological and anatomical anomaly that would impair speech and speech intelligibility, (3) not having a runny nose and/or nasal congestion during the study, and (4) not having a perceptually significant resonance disorder.

This study was conducted in accordance with the principles of the Declaration of Helsinki. Written informed consent was obtained from all participants prior to participation in the study. Ethical approval to conduct the study was obtained from the Non-Interventional Research Ethics Board (number: 61351342/2020-653, date: 31.12.2020) of Üsküdar University.

Data Collection Tools and Procedure

The necessary information about the programs and instructions for the PANM software were turned into an internet link, shared with the readers in this article (link: https://bit.ly/2Xt6o8x). Necessary documents related to PANM software and hardware are located in the file named "PANM Script". Nasalance score analyses can be easily performed with the "Simplified Nasalance Measurement" button on the Praat interface of the PANM Plugin, and the audio signal recorded with the "TextGrid for PANM," entered via the "Annotate" button, can be divided into target segments and analyzed collectively. Detailed information on this analysis is provided in Appendix 1.

An important distinguishing feature of PANM software from nasalance measurement batteries is that the band filter setting can be changed easily. Therefore, the filter settings used in this study were 80-1000 Hz according to researchers' recommendations (4,7) and the 300-750 Hz filter ranges were preferred for comparing the PANM software with the Nasometer II model (NMII) software. Consequently, two different bandpass-filter settings were used in this research, and the analyses were carried out on these two filter settings.

The PANM equipment comprises a headset (Philips SHE1350) with two microphones on a 3 mm-thick plexiglass plate. One side of the plate has a concave morphology that coincides with the mouth and nose, and an integrated plexiglass handle that separates the energy emanating from the mouth and nose. The headphones use a 3.5 mm jack and are positioned 20 mm from the concave edge of the plate. This distance was preferred because of the low sensitivity of the headphones. An Andrea USB-SA (Andrea Communications, Farmingdale, NY) audio interface was used to transfer the recorded sound to the computer environment. In addition, this enabled stereo recording and was compatible with the jack input of the headphones used. Another nasalance measurement tool used in this study is the NMII 6450 (KayPENTAX, Lincoln Park, NJ, USA). The NMII system utilizes unidirectional microphones in the frequency response range of 50 Hz-15 kHz, positioned at a distance of approximately 50 mm from the nose and mouth. The software has a fixed bandpass-filter range of 300-750 Hz which cannot be changed. Both measurement tools are illustrated in Figure 2.

Two different computers were used to create the frequency response curve of the Philips SHE1350 earphones and the Andrea USB-SA audio interface, and the right side of the reference monitor (speaker) was connected to the first computer. Afterward, the pink noise that was generated in the Audacity® (version 3.7.3; Audacity Team, USA) program was conveyed to the reference monitor via the same program. The microphone system that was intended for frequency response curve generation was connected to the second computer. The microphone system was anchored using a microphone stand and appropriate adapters, and the recordings were taken inside a double-walled audiometric cabin. After this, the microphone system was placed at distances of 2 cm, 5 cm, and 30 cm, respectively, so that it would be aligned with the center of the reference monitor, and pink noise was recorded for approximately 10 seconds. The RMS (-dB) values (left: 40.7; right: 40.7) of the recordings were obtained using the Audacity® (version 3.7.3; Audacity Team, USA) software. Nasalance measurements were analyzed using Praat and the PANM plugin (nasalance: 51%). Finally, using the REW - Room EQ Wizard Room Acoustics Software, the frequency response curve (Figure 1) was obtained at a resolution of 1 dB and the smoothing setting of 1/1.

The participants' voice recordings to be analyzed in PANM software were captured with the portable computer via Praat (version 6.2.22) in "PCM wav" format. The sound recordings to be analyzed in the NMII software were recorded in the ".wav" format using the software interface. All recordings were made in the soundproofed phonetic laboratory at Üsküdar University. They were recorded at 11025 Hz sampling rate and 16-bit stereo format. After the transcription of the speech material used (Figure 3) (a detailed explanation of this stage is provided in Appendix 1), "TextGrid" files were created for analysis on PANM software. Nasalance measurements were carried out after the preparations were completed. The recording and analysis procedures used in the study are summarized in Table 1.

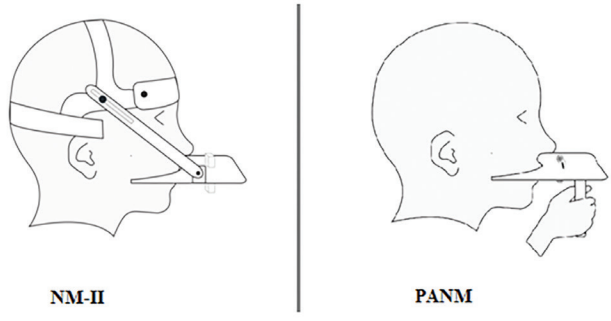

Figure 2. PANM and NMII equipment
PANM: Praat-assisted Nasalance Meter, NMII: Nasometer II model

J Acad Res Med 2026;16(1):43-53

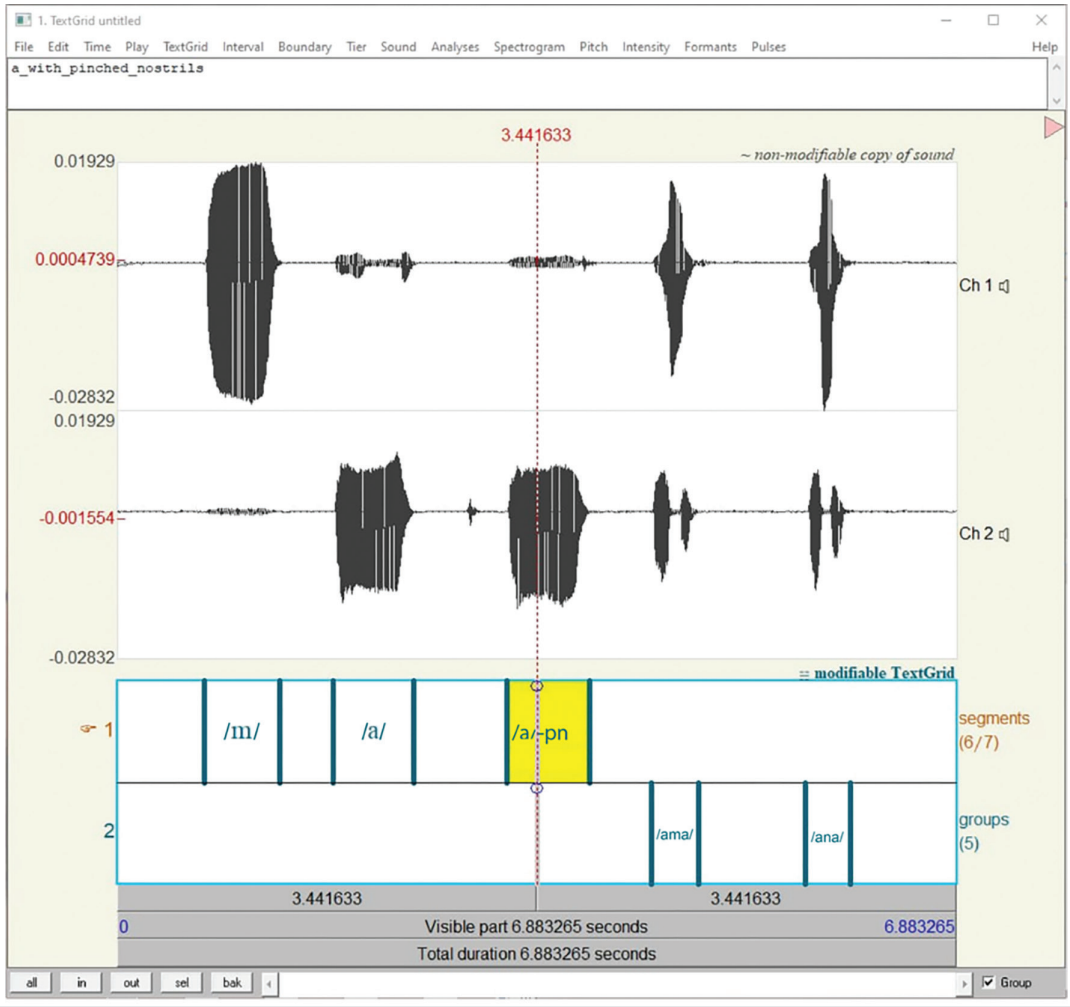

Figure 3. Creating a sample "TextGrid" file in the PANM system
PANM: Praat-assisted Nasalance Meter, pn: Pinched nostrils

Speech materials used for nasalance measurement are affected by linguistic differences. Therefore, the recording material used in this study is given in Table 2. Within the scope of the study, audio recordings were obtained using microphone systems from two different devices (NMI-II and PANM). To minimize potential order effects, the sequence of hardware use was counterbalanced. Accordingly, 30 randomly selected participants were first recorded using the NMI-II device and subsequently with the PANM device. For the remaining 30 participants, the order was reversed, with recordings obtained first using PANM and then NMI-II. In addition, the presentation order of the recording materials was counterbalanced to further control for potential order effects. The overall procedure is illustrated in Figure 4.

Statistical Analysis

Data analysis was performed using the Statistics Package for Social Science (SPSS 22.0, IBM, NY, USA). The data were first analyzed with the Shapiro-Wilk normative distribution test, and it was determined that the normal distribution assumption was met. The recording material's sounds, words, and sentences were coded A1, A2, B1, B2, C1, C2, C3, C4, and C5 (Table 2). A two-way repeated measures ANOVA was conducted for each speech material (A1-C5) to examine the main effects of hardware and software, as well as their interaction effect, on nasalance scores. In addition, Pearson correlation analysis was conducted to examine the linear relationship between the measurements obtained from different hardware and software systems.

Table 1. Recording and analysis procedures used in the study

	Hardware (H)	Software (S)
1	PANM-H	PANM-S (80-1000 Hz)
2	PANM-H	PANM-S (300-750 Hz)
3	PANM-H	NMII-S (300-750 Hz)
4	NMII-H	PANM-S (80-1000 Hz)
5	NMII-H	PANM-S (300-750 Hz)
6	NMII-H	NMII-S (300-750 Hz)

PANM: Praat-assisted Nasalance Meter, NMII: Nasometer II model

J Acad Res Med 2026;16(1):43-53

Table 2. Recording material used in the study		
Isolated phonems	**A1** Sustained consonant production (m)	
	A2 Sustained vowel production (a)	
	A3 Sustained vowel production with pinched nostrils (a)	
Words	**B1** (ɑmɑ)	**B2** (ɑnɑ)
Sentences	**Oral plossive**	**C1** (pɛtɛc kɯɾɯk tɑhtɑ kɑpɯjɯ kɑpɑt:ɯ)
	Oral plossive with pinched nostrils	**C2** (pɛtɛc kɯɾɯk tɑhtɑ kɑpɯjɯ kɑpɑt:ɯ)
	Oral sibilan	**C3** (sɛʃil sɯdʒɑk hɑvuzdɑ sɛs:izdʒɛ jyzdy)
	Nasal	**C4** (ɑn:ɛm emineje nin:i mɯɾɯɬdɑndɯ)
Passage	**C5** (dɑɪkɑpɯsɯndɑn bɑʃkɑ ɑjdɯnɬuk ɟiɾɛdʒɛc hiʃbiɪ jɛɾi oɬmɑjɑn dyc:ɑnɯndɑ tɛc bɑʃunɑ jedʒɛ ɟyndyz kɯvɯɬdʒɯmɬɑɪ sɑʃɑɾɑk ʃɑɬɯʃɑn kɔdʒɑ ɑli tɯpkɯ kɑfesɛ kɔnmuʃ tɛɾbijɛli biɪ ɑsɬɑnɯ ɑndɯɾɯjɔɪdu)	

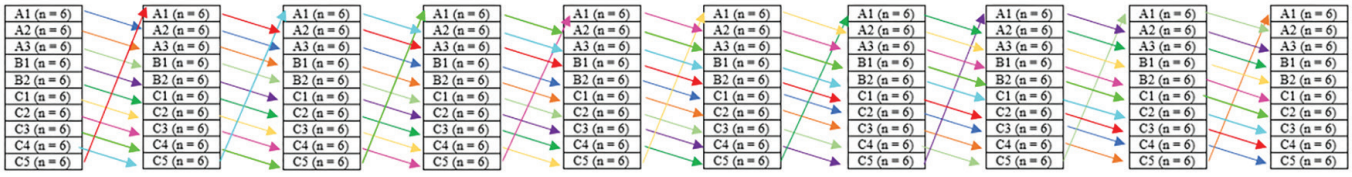

Figure 4. The order of recording acoustic materials

RESULTS

In this section, the results of the correlation and variance analyses will be presented respectively. The relationship between the results obtained in the correlation analysis was examined at two different filter settings: 80-1000 Hz and 300-750 Hz for the PANM software, and 300-750 Hz filter setting for the NMII software. Additionally, the relationship between the nasalance scores obtained when each software was in the filter range of 300-750 Hz was examined for variance analysis.

Correlation Analysis Results

Pearson correlation coefficients were calculated for the isolated phonemes A1, A2, and A3 to examine the relationship between nasalance scores obtained from PANM-H, PANM-S, NMII-H, and NMII-S hardware and software. Additionally, 80-1000 Hz and 300-750 Hz filter ranges were used for PANM-S. Finally, the correlations with the NMII-S, whose default setting was 300-750 Hz, were examined. The results are presented in Table 3.

When the correlations of nasalance scores analyzed with the PANM-S using an 80-1000 Hz bandpass filter and the NMII-S using a 300-750 Hz bandpass filter in Table 3 were examined for A1, there was a statistically significant positive correlation ($p<0.01$) between the measurements of "PANM-H with PANM-S - PANM-H with NMII-S" and "NMII-H with PANM-S - NMII-H with NMII-S" at a high level (r=0.867 and 0.918, respectively). In contrast, no statistically significant correlation was found between the "PANM-H with PANM-S - NMII-H with NMII-S" nasalance score measurements for A1 ($p>0.05$). For A2, there was a statistically significant positive correlation ($p<0.01$) between the measurements of "PANM-H with PANM-S - PANM-H with NMII-S", "NMII-H with PANM-S - NMII-H with NMII-S" and PANM-H with PANM-S - NMII-H with NMII-S" at a high level (r=0.866 and 0.894, respectively) and low (r=0.297). For A3, there was a statistically significant positive correlation ($p<0.01$) between the measurements of "PANM-H with PANM-S - PANM-H with NMII-S", NMII-H with PANM-S - NMII-H with NMII-S" and PANM-H with PANM-S - NMII-H with NMII-S" at the high, moderate and low levels, respectively.

When the nasalance scores analyzed with the PANM-S using a 300-750 Hz bandpass filter and the NMII-S using a 300-750 Hz bandpass filter in Table 3 were examined for A1, there was a moderately positive correlation (r=0.694, 0.663, 0.518, respectively) and a statistically significant relationship ($p<0.01$) between the measurements of "PANM-H with PANM-S - PANM-H with NMII-S", "NMII-H with PANM-S - NMII-H with NMII-S" and PANM-H with PANM-S - NMII-H with NMII-S". For A2, a statistically significant positive correlation ($p<0.01$) was found between the measurements of "PANM-H with PANM-S - PANM-H with NMII-S", NMII-H with PANM-S - NMII-H with NMII-S" and "PANM-H with PANM-S - NMII-H with NMII-S" at a high level (r=0.736 and 0.880, respectively) and at a moderate level (r=0.624), respectively. For A3, a statistically significant positive correlation ($p<0.01$) was found between the measurements of "PANM-H with PANM-S - PANM-H with NMII-S", "NMII-H with PANM-S - NMII-H with NMII-S" and "PANM-H with PANM-S - NMII-H with NMII-S" at a high level (r=0.805 and 0.826, respectively) and at a moderate level (r=0.533).

J Acad Res Med 2026;16(1):43-53

The Pearson correlation coefficient was calculated for B1 and B2 (words) to examine the relationship between the nasalance scores obtained from PANM-H, PANM-S, NMII-H, and NMII-S hardware and software. In addition, two different filter ranges, 80-1000 Hz and 300-750 Hz, were used for the PANM-S's software settings, and the correlations with the NMII-S, whose default setting was 300-750 Hz, were examined, respectively. The results are presented in Table 4.

When the correlation results of nasalance scores analyzed using the PANM-S with an 80-1000 Hz bandpass filter and using an NMII-S with a 300-750 Hz bandpass filter in Table 4 were examined for the B1, a statistically significant correlation ($p<0.01$) was found between "PANM-H with PANM-S - PANM-H with NMII-S" and "NMII-H with PANM-S - NMII-H with NMII-S" measurements at a high (r=0.822) and moderate (r=0.461) levels, respectively. No statistically significant correlation was found between the measurements of "PANM-H with PANM-S - NMII-H with NMII-S" for the B1 ($p>0.05$). For B2, a statistically significant positive correlation ($p<0.01$) was found between "NMII-H with PANM-S - NMII-H with NMII-S", "PANM-H with PANM-S - PANM-H with NMII-S" and "PANM-H with PANM-S - NMII-H with NMII-S" measurements at the moderate (r=0.760 and 0.654, respectively) and low (r=0.366) levels, respectively.

When the correlation of nasalance scores analyzed with the PANM-S using a 300-750 Hz bandpass filter and NMII-S using a 300-750 Hz bandpass filter, as shown in Table 4, were examined for the B1, there was a statistically significant ($p<0.01$) positive correlation (r=0.528, 0.413, and 0.662, respectively) between the measurements of "PANM-H with PANM-S - PANM-H with NMII-S", "PANM-H with PANM-S - NMII-H with NMII-S" and "NMII-H with PANM-S - NMII-H with NMII-S". For B2, there was a highly correlation (r=0.840) between "NMII-H with PANM-S - NMII-H with NMII-S" measurements and a moderately correlation (r=0.354, 0.644, respectively) between "PANM-H with PANM-S - PANM-H with NMII-S" and "PANM-H with PANM-S - NMII-H with NMII-S" measurements ($p<0.01$).

The Pearson correlation coefficient was calculated to examine the relationship between nasalance scores obtained from PANM-H, PANM-S, NMII-H, and NMII-S hardware and software for C1, C2, C3, C4, and C5 (sentences and passage reading). Additionally, two different filter ranges, 80-1000 Hz and 300-750 Hz, were used as the PANM-S's software settings, and their correlations with the NMII-S, whose default setting was 300-750 Hz, were examined, respectively. The results are presented in Table 5.

Table 3. Relationships between nasalance scores obtained from PANM-H, PANM-S, NMII-H and NMII-S hardware and software for A1, A2 and A3 (isolated phonemes)

Software: PANM-S with 80-1000 Hz filter and NMII-S with 300-750 Hz filter			
Correlation	**A1**	**A2**	**A3**
PANM-H with PANM-S - PANM-H with NMII-S	0.867**	0.866**	0.584**
PANM-H with PANM-S - NMII-H with NMII-S	0.238	0.297*	0.285*
NMII-H with PANM-S - NMII-H with NMII-S	0.918**	0.894**	0.839**
Software: PANM-S with 300-750 Hz filter and NMII-S with 300-750 Hz filter			
Correlation	**A1**	**A2**	**A3**
PANM-H with PANM-S - PANM-H with NMII-S	0.694**	0.736**	0.805**
PANM-H with PANM-S - NMII-H with NMII-S	0.518**	0.624**	0.533**
NMII-H with PANM-S - NMII-H with NMII-S	0.663**	0.880**	0.826**

*: p<0.05, **: p<0.01, A1: Sustained consonant /m/, A2: Sustained vowel /a/, A3: Sustained vowel /a/with pinched nostrils, H: Hardware, S: Software, PANM: Praat-assisted Nasalance Meter, NMII: Nasometer II model

Table 4. Relationship between nasalance scores obtained from PANM-H, PANM-S, NMII-H and NMII-S hardware and software for B1 and B2 (words)

Software: PANM-S with 80-1000 Hz filter and NMII-S with 300-750 Hz filter		
Correlation	**B1**	**B2**
PANM-H with PANM-S - PANM-H with NMII-S	0.822*	0.654*
PANM-H with PANM-S - NMII-H with NMII-S	0.212	0.366*
NMII-H with PANM-S - NMII-H with NMII-S	0.461*	0.760*
Software: PANM-S with 300-750 Hz filter and NMII-S with 300-750 Hz filter		
Correlation	**B1**	**B2**
PANM-H with PANM-S - PANM-H with NMII-S	0.528*	0.354*
PANM-H with PANM-S - NMII-H with NMII-S	0.413*	0.644*
NMII-H with PANM-S - NMII-H with NMII-S	0.662*	0.840*

*: p<0.05, B1: Word (ama), B2: word (ana), H: Hardware, S: Software, PANM: Praat-assisted Nasalance Meter, NMII: Nasometer II model

J Acad Res Med 2026;16(1):43-53

Table 5. The relationship between nasalance scores for C1, C2, C3, C4 and C5 (sentences and passage reading) obtained from PANM-H, PANM-S, NMII-H and NMII-S hardware and software

Software: PANM-S with 80-1000 Hz filter and NMII-S with 300-750 Hz filter					
Correlation	**C1**	**C2**	**C3**	**C4**	**C5**
PANM-H with PANM-S - PANM-H with NMII-S	0.879**	0.601**	0.876**	0.804**	0.776**
PANM-H with PANM-S - NMII-H with NMII-S	0.540**	0.264*	0.509**	0.200	0.088
NMII-H with PANM-S - NMII-H with NMII-S	0.853**	0.910**	0.862**	0.688**	0.670**
Software: PANM-S with 300-750 Hz filter and NMII-S with 300-750 Hz filter					
Correlation	**C1**	**C2**	**C3**	**C4**	**C5**
PANM-H with PANM-S - PANM-H with NMII-S	0.612**	0.865**	0.526**	0.474**	0.272*
PANM-H with PANM-S - NMII-H with NMII-S	0.660**	0.726**	0.545**	0.470**	0.500**
NMII-H with PANM-S - NMII-H with NMII-S	0.748**	0.853**	0.645**	0.787**	0.608**

*: p<0.05, **: p<0.01, C1: Oral plosive sentence, C2: Oral plossive with pinched nostrils sentence, C3: Oral sibilan sentence, C4: Nasal sentence, C5: Passage, H: Hardware; S: Software, PANM: Praat-assisted Nasalance Meter, NMII: Nasometer II model

When the correlation results of nasalance scores analyzed with the PANM-S using the 80-1000 Hz bandpass filter and NMII-S using the 300-750 Hz bandpass filter in Table 5 were examined for C1, there was a statistically significant positive correlation (p<0.01) between the measurements of "PANM-H with PANM-S - PANM-H with NMII-S", "NMII-H with PANM-S - NMII-H with NMII-S" and "PANM-H with PANM-S - NMII-H with NMII-S" at a high level (r=0.879, and 0.853, respectively) and at a moderate level (r=0.540), respectively. For C2, there was a statistically significant correlation between "NMII-H with PANM-S - NMII-H with NMII-S", "PANM-H with PANM-S - PANM-H with NMII-S", and "PANM-H with PANM-S - NMII-H with NMII-S" (p<0.01; p=0.043; p<0.01, respectively) measurements at high (r=0.910), moderate (r=0.601), and low (r=0.264) levels, respectively. For C3, there was a statistically significant positive correlation (p<0.01) between the "PANM-H with PANM-S - PANM-H with NMII-S", "NMII-H with PANM-S - NMII-H with NMII-S" and "PANM-H with PANM-S - NMII-H with NMII-S" measurements for high (r=0.876 and 0.862, respectively), and moderate (r=0.509, respectively) levels. It was determined that there was a statistically significant correlation (p<0.01) between the "PANM-H with PANM-S - PANM-H with NMII-S" and "NMII-H with PANM-S - NMII-H with NMII-S" measurements for C4 at high (r=0.804) and moderate (r=0.688) levels, respectively. No statistically significant correlation was found between the measurements of "PANM-H with PANM-S - NMII-H with NMII-S" for C4 (p>0.05). It was determined that there was a statistically significant positive correlation (p<0.01) between the "PANM-H with PANM-S - PANM-H with NMII-S" and "NMII-H with PANM-S - NMII-H with NMII-S" measurements for C5 at a moderate level (r=0.776 and 0.670, respectively). No statistically significant correlation was found between the measurements of "PANM-H with PANM-S - NMII-H with NMII-S" for C5 (p>0.05).

When the correlation results of nasalance scores analyzed with the PANM-S using a 300-750 Hz bandpass filter and NMII-S using a 300-750 Hz bandpass filter in Table 5 were examined, for C1, a statistically significant positive correlation (p<0.01) was found between the "NMII-H with PANM-S - NMII-H with NMII-S", "PANM-H with PANM-S - PANM-H with NMII-S" and "PANM-H with PANM-S - NMII-H with NMII-S" measurements, at the high (r=0.748) and moderate (r=0.612 and 0.660, respectively) levels. For C2, there was a statistically significant positive correlation (p<0.01) at the high level (r=0.865, 0.726, and 0.853, respectively) between the "PANM-H with PANM-S - PANM-H with NMII-S", PANM-H with PANM-S - NMII-H with NMII-S" and "NMII-H with PANM-S - NMII-H with NMII-S" measurements. A statistically significant correlation (p<0.05) was found between the measurements of "PANM-H with PANM-S - PANM-H with NMII-S", "PANM-H with PANM-S - NMII-H with NMII-S" and "NMII-H with PANM-S - NMII-H with NMII-S" for C3, which was moderately positive (r=0.526, 0.545, and 0.645, respectively). A statistically significant correlation (p<0.01) was found between "NMII-H with PANM-S - NMII-H with NMII-S", "PANM-H with PANM-S - PANM-H with NMII-S" and "PANM-H with PANM-S - NMII-H with NMII-S" measurements for C4, which was highly positive (r=0.787) and moderately positive (r=0.474 and 0.470, respectively). For C5, a statistically significant positive correlation (p=0.046; p<0.01; p<0.01, respectively) was found between moderate (r=0.500; 0.608, respectively) and low (r=0.272), and low (r=0.272) measurements of "PANM-H with PANM-S - PANM-H with NMII-S", "PANM-H with PANM-S - NMII-H with NMII-S" and "PANM-H with PANM-S - PANM-H with NMII-S".

Variance Analysis Results

In this section, the results obtained when each software was at the same filter setting (300-750 Hz) were included. A two-way repeated measures ANOVA was used to examine the differentiation of nasalance scores for A1, A2, and A3 (isolated phonemes) according to hardware and software. The results are presented in Table 6. In addition, descriptive statistics of nasalance score analyses performed at the PANM 80-1000 Hz bandpass-filter setting regarding the recording material used are included in Appendix 2.

Table 6 shows that nasalance scores for A1, A2, and A3 differed significantly according to software (p<0.01). In addition,

J Acad Res Med 2026;16(1):43-53

significant differences were observed according to hardware for A1 and A2 (p<0.01). No statistically significant interaction effect between hardware and software was found at any level (p>0.05). Examination of the mean values indicated that, for A1, A2, and A3, nasalance scores obtained using the NMII software were higher than those obtained using the PANM software. For A1 and A2, however, nasalance scores obtained using the PANM hardware were higher than those obtained using the NMII hardware.

A two-way repeated measures ANOVA was used to examine the differentiation of nasalance scores for B1 and B2 (words) according to hardware and software. In the software, a 300-750 Hz bandpass-filter set was used on both devices. The results are presented in Table 7.

When Table 7 was examined, it was seen that nasalance scores for B1 and B2 statistically significant difference (p<0.01) according to both hardware and software. It was found that nasalance scores for B1 did not show a statistically significant difference (p>0.05), and B2 showed a statistically significant difference (p=0.031), according to the hardware*software interaction effect. When the averages were examined, it was determined that the averages of nasalance scores obtained from the NMII software were higher than those obtained from the PANM software, and that the averages of the nasalance scores obtained from the PANM hardware were higher than those obtained from the NMII hardware, in common for B1 and B2. Examination of the interaction plot for B2 (Appendix 3) revealed a similar pattern in the hardware*software interaction.

A two-way repeated measures ANOVA was used to examine the differentiation of nasalance scores for C1, C2, C3, C4, and C5 (sentences and reading passage) according to hardware and software. For the software analyses, a bandpass-filter setting of 300-750 Hz was used on both devices. The results are presented in Table 8.

Table 6. Examining the differentiation of nasalance scores for A1, A2 and A3 (isolated phonemes) according to software and hardware

	Software (S)	Hardware (H)		SD		F	p
A1	PANM-S	PANM-H	96.75	1.68	Software	60.789	**<0.01**
	PANM-S	NMII-H	94.05	2.62	Hardware	111.909	**<0.01**
	NMII-S	PANM-H	97.87	1.19	Software*hardware	1.842	0.180
	NMII-S	NMII-H	95.55	1.61			
A2	PANM-S	PANM-H	25.64	16.06	Software	36.657	**<0.01**
	PANM-S	NMII-H	15.87	11.59	Hardware	29.834	**<0.01**
	NMII-S	PANM-H	32.50	20.00	Software*hardware	0.483	0.490
	NMII-S	NMII-H	21.39	12.52			
A3	PANM-S	PANM-H	5.27	5.18	Software	34.326	**<0.01**
	PANM-S	NMII-H	3.98	2.75	Hardware	3.780	0.057
	NMII-S	PANM-H	6.75	6.08	Software*hardware	0.433	0.513
	NMII-S	NMII-H	5.84	3.90			

A1: Sustained consonant /m/, A2: Sustained vowel /a/, A3: Sustained vowel /a/with pinched nostrils, PANM: Praat-assisted Nasalance Meter, NMII: Nasometer II model, SD: Standard deviation

Table 7. Examination of the differentiation of nasalance scores for B1 and B2 (words) according to hardware and software

	Software (S)	Hardware (H)		SD		F	p
B1	PANM-S	PANM-H	54.25	11.76	Software	225.152	<0.01
	PANM-S	NMII-H	36.80	11.32	Hardware	210.544	<0.01
	NMII-S	PANM-H	69.52	8.81	Software*hardware	0.093	0.761
	NMII-S	NMII-H	52.52	6.62			
B2	PANM-S	PANM-H	65.80	9.86	Software	88.700	<0.01
	PANM-S	NMII-H	48.14	11.96	Hardware	391.102	<0.01
	NMII-S	PANM-H	75.98	7.50	Software*hardware	4.911	0.031
	NMII-S	NMII-H	55.21	7.60			

B1: Word (ɑmɑ), B2: Word (ɑnɑ), PANM: Praat-assisted Nasalance Meter, NMII: Nasometer II model, SD: Standard deviation

J Acad Res Med 2026;16(1):43-53

Table 8. Investigation of the differentiation of nasalance scores for C1, C2, C3, C4 and C5 (sentences and reading passage) according to hardware and software

	Software (S)	Hardware (H)		SD		F	p
C1	PANM-S	PANM-H	13.54	6.89	Software	53.183	**<0.01**
	PANM-S	NMII-H	8.20	5.00	Hardware	47.478	**<0.01**
	NMII-S	PANM-H	18.18	8.09	Software*hardware	0.097	0.756
	NMII-S	NMII-H	12.62	4.65			
C2	PANM-S	PANM-H	2.79	2.55	Software	101.268	**<0.01**
	PANM-S	NMII-H	3.95	1.49	Hardware	7.191	**0.010**
	NMII-S	PANM-H	4.50	3.00	Software*hardware	17.239	<0.01
	NMII-S	NMII-H	4.91	1.37			
C3	PANM-S	PANM-H	15.86	8.38	Software	110.032	**<0.01**
	PANM-S	NMII-H	8.80	5.77	Hardware	69.566	**<0.01**
	NMII-S	PANM-H	24.37	9.01	Software*hardware	0.344	0.560
	NMII-S	NMII-H	17.80	6.07			
C4	PANM-S	PANM-H	69.07	9.14	Software	177.937	**<0.01**
	PANM-S	NMII-H	50.80	11.12	Hardware	361.800	**<0.01**
	NMII-S	PANM-H	80.75	5.25	Software*hardware	0.107	0.744
	NMII-S	NMII-H	62.86	5.19			
C5	PANM-S	PANM-H	39.07	9.72	Software	73.679	**<0.01**
	PANM-S	NMII-H	25.71	8.09	Hardware	205.550	**<0.01**
	NMII-S	PANM-H	49.50	9.54	Software*hardware	3.066	0.086
	NMII-S	NMII-H	33.93	4.20			

C1: Oral plosive sentence, C2: Oral plossive with pinched nostrils sentence, C3: Oral sibilan sentence, C4: Nasal sentence, C5: Passage, SD: Standard deviation

When Table 8 was examined, it was noted that nasalance scores for C1, C3, C4, and C5 statistically significant difference ($p<0.01$) based on hardware and software. Similarly, for C2, statistically significant differences were observed both between the software systems ($p<0.01$) and between the hardware systems ($p=0.010$). It was determined that there was no statistically significant difference for C1, C3, C4, and C5 according to the hardware-software interaction effect ($p>0.05$). C2 showed a statistically significant difference ($p<0.01$), according to the hardware*software interaction effect. When the averages were examined, the common result for C1, C2, C3, C4, and C5 was as follows: the nasalance score averages obtained from the NMII software were higher than those obtained from the PANM software. In parallel, the nasalance score averages obtained using NMII hardware were higher than those obtained using PANM hardware. Examination of the interaction plot for C2 (Appendix 3) revealed a similar pattern in the hardware*software interaction.

DISCUSSION

The nasalance parameter allows for the quantitative evaluation and description of an individual's nasal resonance characteristics. However, there are many nasal measurement systems in the literature and clinical use, and many studies have compared the different nasal resonance scores obtained using each of these systems (1,8-12). Therefore, this study aimed to compare the nasalance scores obtained from the PANM system, which is equipped with easily accessible and cost-effective materials, as well as free access to the plugin link related to the software, with the nasalance scores obtained from NMII.

The unchangeable bandpass-filter range (300-750 Hz) of the NMII system may have minimal analytical power for variations related to individual differences, phonetic inventory of different languages and/or resonance disturbance. Indeed, Awan (3) pointed out that signal filtering affects researchers' ability to perform acoustic analyses of nasal sound recordings. In addition, one of the essential advantages of the system is that the filter setting of the PANM system can be changed to determine nasal emission (13), which is acoustically characterized by wide-band noise. Unfortunately, publications examining the effect of manufacturer selections, such as microphone position and bandpass-filter range for signal analysis, on the nasalance value are also very limited in line with the available resources.

It is thought that the PANM system may have critical importance in the diagnosis, therapy and biofeedback processes of many different resonance disorders, mainly since different band filter settings can be used in the analyses performed by that system. Furthermore, the acoustic signal can be observed spectrally during the analysis and this signal can be segmented into background or

smaller units. As a first step in this study, the researchers performed a correlation analysis in which both software (PANM and NMII) had equal bandpass-filter settings (300-750 Hz), moreover, the correlations of nasalance scores obtained from the PANM adjusted to the 80-1000 Hz bandpass filter range suggested by researchers (4) and obtained from the NMII (300-750 Hz) system, whose filter settings were fixed by the manufacturers, were then examined.

To summarize, the results of the correlation analysis between the nasalance scores obtained using the PANM and NMII software and hardware and a 300-750 Hz bandpass filter for both systems were statistically significant positive correlation and regardless of the hardware used for all signals in the recording material (Table 2). In other words, when only the software systems were considered, linear relationships were observed between the nasalance scores. Additionally, the results show that the nasalance scores recorded using the PANM hardware and analyzed with the PANM software had a statistically significant positive correlation with the nasalance scores recorded with the NMII hardware and analyzed with the NMII software. In other words, when PANM and NMII were evaluated as integrated hardware-software systems, a linear relationship was observed between the nasalance scores obtained from the two systems, provided that the same filter settings were applied. In addition, when the bandpass-filter range for PANM was set to its default setting of 80-1000 Hz, there were some inconsistencies between both software systems for some acoustic signals in the recording materials. We believe that this was caused by different filter settings.

Correlation analyses demonstrated significant linear associations between nasalance scores obtained from different hardware and software configurations. However, covariation between measurements does not imply the absence of differences between them. Therefore, to determine whether mean-level differences existed between systems and to examine how these differences varied as a function of hardware and software factors, the results of the two-way repeated-measures analysis of variance were examined. When the results were evaluated in terms of software, NMII yielded higher nasalance scores than PANM across all acoustic recording materials, regardless of the hardware used. Although identical filter settings were applied, this difference may be attributable to variations in the underlying signal processing algorithms of the software systems. With respect to hardware, PANM produced higher nasalance scores than NMII across all acoustic materials (except C2), independent of the software used. The difference in hardware can be attributed to microphones having different transducer types (dynamic and condenser), frequency response curves, and sensitivity (14,15). Different microphone features have been reported to affect nasalance score results (16). The consistency of the variance patterns related to software and hardware across different task types (isolated sounds, words, and sentence-level materials) suggests that the observed findings reflect systematic differences rather than random variation. The observed difference in the C2 material may stem from the recording material's task-specific characteristics and does not appear to substantially affect the general pattern of findings.

In conclusion, different hardware and software configurations were found to have significant effects on nasalance scores, and these effects emerged consistently across various speech tasks/materials. These findings indicate that nasalance measurements may be influenced not only by filter settings but also by the signal processing approach of the software and the technical characteristics of the hardware. Accordingly, the specific configuration used should be taken into account when interpreting nasalance scores obtained from different systems. To enhance comparability across clinical and research settings, measurement conditions should be clearly reported, and each system should be evaluated within the context of its own technical specifications. In parallel with this, researchers who compared different systems used to obtain nasalance scores and/or other models related to the same system obtained similar results (1,8-12,17).

Study Limitations

Future studies should focus on calculating frequency response curves for the PANM hardware using mixed and matched-pair microphone configurations and investigating the effects of microphone distance on nasalance scores. In addition, the influence of different filter ranges and various speech or recording materials on nasalance measurements should be systematically examined to further refine the system's measurement properties.

CONCLUSION

The findings of this study indicate that nasalance scores obtained using NMII and PANM may differ significantly depending on hardware and software configurations. Although linear associations were observed between measurements, the magnitude of nasalance scores appears to be influenced by the technical characteristics of the systems used.

Accordingly, when interpreting nasalance scores in clinical and research contexts, the hardware and software characteristics of the measurement system should be clearly reported. Although the PANM system offers advantages in terms of accessibility and spectral view, the values obtained should be interpreted with caution.

Appendix 1-3: https://d2v96fxpocvxx.cloudfront.net/68ab204c-182b-49da-b227-bc7efe058632/content-images/fbcd6d7f-bf4e-47c4-a85a-802eba1e09b6.pdf

Ethics

Ethics Committee Approval: Ethical approval to conduct the study was obtained from the Non-Interventional Research Ethics Board (number: 61351342/2020-653, date: 31.12.2020) of Üsküdar University.

Informed Consent: Written informed consent was obtained from all participants prior to participation in the study.

J Acad Res Med 2026;16(1):43-53

Acknowledgment

The corrected version is: "The basis for this study is the second author's (M.S.B.) MSc thesis, which was supervised by the first author (M.A.K.). The third author (G.T.) supported the data collection and the writing process of this study".

Footnotes

Author Contributions: Concept - M.A.K., M.S.B., G.T.; Design - M.A.K., M.S.B., G.T.; Data Collection and/or Processing - M.A.K., M.S.B., G.T.; Analysis and/or Interpretation - M.A.K., M.S.B., G.T.; Literature Search - M.A.K., M.S.B., G.T.; Writing - M.A.K., M.S.B., G.T.

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

Financial Disclosure: The authors report that no financial support was received for this study.

REFERENCES

1. Bressmann T. Comparison of nasalance scores obtained with the Nasometer, the NasalView, and the OroNasal System. Cleft Palate Craniofac J. 2005; 42: 423-33.
2. KayPENTAX. Instruction manual: nasometer II, model 6450. Software Instruction Manual, issue C. KayPENTAX. 2010.
3. Awan SN. Analysis of nasalance: NasalView. Clinical Phonetics and Linguistics. London: Whurr-Publishers. 1997: 518-25.
4. Kılıç MA, Tüysüz O, Hanege FM, Paltura C. Praat-assisted Nasalance Meter: a low-cost nasalance measurement system for evaluation of nasal resonance disorders. Hamidiye Med J. 2021; 2: 116-21.
5. Boersma P, Weenink D. Praat: doing phonetics by computer [computer program]. Version 6.2.22. Amsterdam: University of Amsterdam; 2022. (cited 6 Oct 2022). Available from: https://www.praat.org.
6. Stewart J, Kohlberger M. Earbuds: a method for analyzing nasality in the field. Lang Doc Conserv. 2017; 11: 49-80.
7. Tüysüz, O. The effect of nasal surgery on nasal consonants (dissertation). İstanbul: İstanbul Medeniyet University, 2019.
8. Bressmann T, Klaiman P, Fischbach S. Same noses, different nasalance scores: data from normal subjects and cleft palate speakers for three systems for nasalance analysis. Clin Linguist Phon. 2006; 20: 163-70.
9. Awan SN, K Omlar, Christopher RW. Effects of computer system and vowel loading on measures of nasalance. J Speech Lang Hear Res. 2011; 54: 1284-94.
10. Awan SN, Virani A. Nasometer 6200 versus nasometer II 6400: effect on measures of nasalance. Cleft Palate Craniofac J. 2013; 50: 268-74.
11. Lewis KE, Watterson T. Comparison of nasalance scores obtained from the nasometer and the nasal view. Cleft Palate Craniofac J. 2003: 40; 40-5.
12. Watterson T, Lewis K, Brancamp T. Comparison of nasalance scores obtained with the nasometer 6200 and the nasometer II 6400. Cleft Palate Craniofac J. 2005; 42; 574-9.
13. Rollins M, Oren L. Effects of nasal emission and microphone placement on nasalance score during /s/. Proc Mtgs Acoust. 2020; 42: 060001.
14. Patel RR, Awan SN, Barkmeier-Kraemer J, Courey M, Deliyski D, Eadie T. Recommended protocols for instrumental assessment of voice: American Speech-Language-Hearing Association expert panel to develop a protocol for instrumental assessment of vocal function. Am J Speech Lang Pathol. 2018; 27: 887-905.
15. Svec JG, Granqvist S. Guidelines for selecting microphones for human voice production research. Am J Speech Lang Pathol. 2010; 19: 356-68.
16. Zajac DJ, Lutz RW, Mayo R. Microphone sensitivity as a source of variation in nasalance scores. J Speech Lang Hear Res. 1996; 39; 1228-31.
17. de Boer G, Bressmann T. Comparison of nasalance scores obtained with the nasometers 6200 and 6450. Cleft Palate Craniofac J. 2014; 51: 90-7.

DOI: 10.4274/jarem.galenos.2025.38981
J Acad Res Med 2026;16(1):54-55

Methodological Considerations Regarding the Development of the Tactical Medicine Knowledge and Awareness Scale: Implications for Future Validation Studies

Mehmet Necmeddin Sutaşır

University of Health Sciences Türkiye, Başakşehir Çam and Sakura City Hospital, Department of Emergency Medicine, İstanbul, Türkiye

Cite this article as: Sutaşır MN. Methodological considerations regarding the development of the tactical medicine knowledge and awareness scale: implications for future validation studies. J Acad Res Med. 2026;16(1):54-55

Keywords: Methodological considerations, tactical medicine, awareness scale, emergency medicine, validation studies

Dear Editor,

I'd like to thank Sarbay and his associates for rapidly addressing a significant gap in emergency medical education. The tactical medicine knowledge and awareness scale (TAMKA) is essential, particularly as tactical emergency medical support (TEMS) evolves in contemporary emergency care. Although I acknowledge the study's contribution, I assert that certain methodological aspects require scrutiny before the widespread implementation of this strategy. The internal consistency (Cronbach's α=0.808) is satisfactory; however, this number is derived from a single-center study with a limited sample size (n=131). The authors correctly assert that this strategy constrains generalizability across various geographic regions and emergency medicine training environments. The measure lacks significant evidence of discriminant or convergent validity, which the authors acknowledge are critical issues. Modern psychometric standards necessitate validation of the five qualities as distinct constructs prior to endorsing the instrument for evaluating various aspects of knowledge in tactical medicine (1). Additionally, no information is available regarding criterion validity. It is uncertain whether TAMKA ratings may predict a person's effectiveness in tactical operations, field decision-making, or, ultimately, patient outcomes. Such a demonstration is particularly useful for determining whether the scale accurately measures the aspects most relevant to tactical scenarios (2). It's also vital to discuss about issues with technology. Certain factor loadings are very low (item 20: λ=0.435 in factor 2), although certain content validity indices are close to important values (item 9: content validity ratio=0.54). A content validity coefficient of 0.632 is acceptable; however, it suggests that the scale may not cover all aspects of tactical medicine. There is no strong methodological rationale for the reduction from 55 initial items to 28 final items. Without a test-retest reliability assessment, it's difficult to determine whether the scale remains stable over time. This is crucial for a tool that teachers may use to test students individually to assess how well they perform. Despite these concerns, this research represents a significant initial step towards formalizing competency assessment in tactical medicine. The authors have identified a specific educational prerequisite and devised a technique grounded in expert consensus. I recommend that the authors regard this publication as a foundation for future improvement rather than a finished product. TAMKA's psychometric properties and clinical significance would be greatly improved by multicenter validation studies, thorough assessments of criterion-related and discriminant validity, evaluations of test-retest reliability, and analyses of correlations with external performance metrics. Not only would these efforts improve Turkish emergency medical systems, but they would also serve as a model for other nations if TEMS becomes recognized as a critical skill for emergency care (3,4).

Sincerely,

Footnotes

Financial Disclosure: The author report that no financial support was received for this study.

ORCID IDs of the author: M.N.S. 0000-0001-6472-5092

Corresponding Author: Mehmet Necmeddin Sutaşır, MD;
E-mail: drmehmetns@gmail.com

Received Date: 23.11.2025 **Accepted Date:** 05.12.2025
Epub: 22.01.2026
Publication Date: 28.04.2026

J Acad Res Med 2026;16(1):54-55

References

1. Polit DF, Beck CT. Nursing research: generating and assessing evidence for nursing practice. 10th ed. Philadelphia (PA): Wolters Kluwer; 2021.
2. Sarbay İ, Kanter E, Efgan MG, Kaymaz E, Özel Z, Tekindal MA. Development of tactical medicine knowledge and awareness scale: validity and reliability study. J Acad Res Med. 2025; 15: 1-9
3. Ulintz AJ, O'Connor LC, Heffern TM, Rowe J, Rollman JE, Wingrove G, et al. Mobile integrated health care and community paramedicine: a position statement and resource document of NAEMSP. Prehosp Emerg Care. 2025: 1-13.
4. Pap R, Lockwood C, Stephenson M, Simpson P. Indicators to measure prehospital care quality: a scoping review. JBI Database System Rev Implement Rep. 2018; 16: 2192-223.