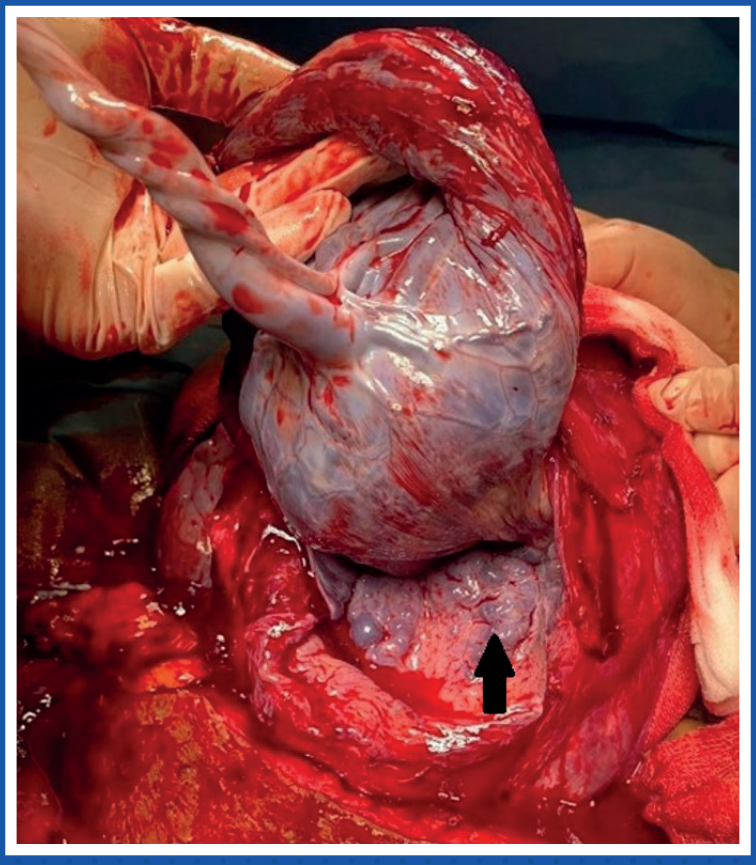
ISSN: 2979-9538

NORTHWESTERN MEDICAL JOURNAL

Year: 2025 ▪ Volume: 5 ▪ Issue: 4



Official Journal of Izzet Baysal Training and Research Hospital

NORTHWESTERN MEDICAL JOURNAL

Year: 2025 ▪ Volume: 5 ▪ Issue: 4

ISSN: 2979-9538

www.nwmedj.org

Official Journal of Izzet Baysal
Training and Research Hospital

Northwestern Medical Journal

ISSN: 2979-9538 (Online)

Owner
Owner on behalf of the Bolu İzzet Baysal Training and Research Hospital
Assoc. Prof. Süleyman Özsarı, MD
Director, Bolu İzzet Baysal Training and Research Hospital, Türkiye

Publication Type
International peer-reviewed journal

Publication Frequency and Language
Quarterly (January, April, July, October), English

Abstracting and Indexing
Northwestern Medical Journal indexed in TR Index, DOAJ, GALE, J-Gate, and TürkMedline.

Editor in Chief
Prof. Ahmet Ural, MD
Bolu Abant Izzet Baysal University, Faculty of Medicine, Department of Otorhinolaryngology, Head & Neck Surgery, Türkiye

Publisher
Izzet Baysal Training and Research Hospital

Publisher Address
Izzet Baysal Training and Research Hospital, Gölköy 14280 Bolu, Türkiye
Phone: +90 374 253 46 18 / +90 374 253 46 56
E-mail: nwmedj@gmail.com
Web: izzetbaysaleah.saglik.gov.tr

Publishing Services
Akdema Informatics and Publishing
Address: Kızılay Mah. Gazi Mustafa Kemal Bulvarı No: 23/8 06420 Çankaya/Ankara
Certificate number: 52576
E-mail: bilgi@akdema.com
Tel: +90 533 166 80 80
Web: www.akdema.com

You can reach all publication policies and author guidelines from **www.nwmedj.org**

Northwestern Medical Journal

Editorial Board

Northwestern Medical Journal

Advisory Board

Contents

Editorial

The Northwestern Medical Journal's editorial board takes a moment as 2025 draws to a close to consider a year that has reshaped the parameters of patient care, medical practice, and research. This year has been a celebration of human ingenuity and resiliency, from the busy clinics of urban hubs to the remote outreach initiatives addressing global health inequities. We invite our readers—clinicians, researchers, educators, and policymakers—to join us in this time of reflection in this last issue of Volume 5, as we celebrate accomplishments and map out new paths.

Significant developments over the last nine months highlight the progress of medicine. In addition to increasing early detection rates for diseases like sepsis and cancer, the broad use of AI in diagnostics has made high-fidelity imaging more accessible to those in situations with limited resources. We especially uphold the following ethical imperatives throughout our publication journey: protecting patient privacy, guaranteeing algorithmic equity, and encouraging interdisciplinary collaboration to stop the digital gap from escalating health disparities.

We see the full potential of precision medicine shining on the horizon in 2026. Gene-editing treatments, such as CRISPR-based treatments for uncommon genetic illnesses, are about to make the leap from research to clinical use. As wearable biosensors usher in an era of proactive, patient-empowered wellness, the development of telemedicine will further dissolve geographic borders. We promise to give these views more weight, and we are looking for proposals that consider "for whom and why" in addition to "what works." Our request is clear: send us your most audacious concepts, your most meticulous statistics, and your impactful anecdotes. We can turn evidence into equity if we work together.

As we celebrate this season, we would want to express our sincere gratitude to all of our readers, reviewers, and writers who help to make Northwestern Medical Journal an essential platform for medical discussion. May the year's end encourage us to move forward with bravery, kindness, and inquiry.

Sincerely,
Prof. Ahmet Ural, M.D.
Editor-in-chief

NORTHWESTERN MEDICAL JOURNAL

Northwestern Med J. 2025;5(4):201-208
DOI: 10.54307/NWMJ.2025.162

RESEARCH ARTICLE

HPV status and colposcopy: key predictors in cervical cancer and precancerous lesion diagnosis

Sıtkı Özbilgeç[1], Fatih Akkuş[2], Rıfat Şener[1], Emin Erhan Dönmez[3], Elif Ceyda Serin[4], Ali Acar[2]

[1]Konya City Hospital, Gynecologic Oncology Surgery, Konya, Türkiye
[2]Department of Perinathology, Faculty of Medicine, Necmettin Erbakan University, Konya, Türkiye
[3]Istanbul Training and Research Hospital, Gynecologic Oncology Surgery, Istanbul, Türkiye
[4]Konya City Hospital, Obstetrics and Gynecology, Konya, Türkiye

Cite as: Özbilgeç S, Akkuş F, Şener R, Dönmez EE, Serin EC, Acar A. HPV status and colposcopy: key predictors in cervical cancer and precancerous lesion diagnosis. Northwestern Med J. 2025;5(4):201-208.

ABSTRACT

Objective: This study aims to evaluate the diagnostic value of high-risk human papillomavirus (HR-HPV) positivity—particularly HPV type 16—together with cytological findings and colposcopic assessment in detecting cervical intraepithelial neoplasia (CIN) and cervical cancer.

Materials and Methods: We retrospectively analyzed 781 women who underwent colposcopic examination at a tertiary healthcare center between January 2020 and April 2024. HPV DNA results and cytological evaluations prior to colposcopy were reviewed. Histopathological outcomes from cervical biopsy, endocervical curettage (ECC), and probe curettage (P/C) were examined. Multivariate logistic regression was used to determine the predictors of CIN 2 or more severe lesions.

Results: HPV type 16 was found to be a significant independent predictor of CIN2+ lesions, with an odds ratio of 22.36 (p=0.002) compared to HPV-negative individuals. The model demonstrated statistical significance, and the area under the curve (AUC) was calculated as 0.679, indicating moderate diagnostic performance. Other HPV genotypes and unknown HPV status also showed a significant association with higher-grade lesions.

Conclusion: HR-HPV, and particularly HPV 16, is strongly associated with the presence of advanced cervical lesions. Women testing positive for HPV 16 should undergo close surveillance and timely colposcopic evaluation by specialists to enable early diagnosis and prevent progression to invasive cervical cancer.

Keywords: HPV 16, cervical cancer, colposcopy, CIN2+, precancerous lesion

Corresponding author: Sıtkı Özbilgeç **E-mail:** sozbilgec@yahoo.com
Received: 16.11.2024 **Accepted:** 14.02.2025 **Published:** 26.10.2025

INTRODUCTION

Cervical cancer remains one of the most common malignancies affecting women worldwide and ranks fourth among all cancers in terms of incidence (1). What makes this disease unique is that it has a well-defined and preventable cause—persistent infection with high-risk human papillomavirus (HR-HPV) (2). Thanks to advancements in vaccination and screening programs, the burden of cervical cancer has decreased in many countries; however, it continues to be a serious public health issue, especially in regions with limited access to preventive care (3).

The early stages of cervical cancer and precancerous lesions often present no symptoms, complicating timely detection. Symptoms such as abnormal vaginal discharge or postcoital bleeding typically appear in more advanced stages (4,5). Therefore, early detection methods are vital to identify and treat lesions before they progress (6).

Among the established risk factors, persistent HR-HPV infection—especially with types such as HPV 16 and 18—has been recognized as the principal cause of high-grade cervical intraepithelial neoplasia (CIN) and cervical cancer (7-10). Although most HPV infections resolve spontaneously, a small proportion may persist and eventually lead to malignant transformation (11,12). This underlines the importance of HPV-based screening strategies, in addition to traditional cytology (13-15).

Colposcopy serves as a bridge between screening and definitive diagnosis. It allows targeted biopsy of suspicious areas, increasing the diagnostic yield in patients with abnormal cytology or positive HPV tests. However, the accuracy of colposcopic evaluation can vary depending on the examiner's experience and the nature of the lesion, which makes it essential to combine clinical findings with objective markers like HPV status (16-18).

In this study, we aim to explore the diagnostic value of HR-HPV positivity—especially HPV 16—and its correlation with cytological results and colposcopic findings. We also investigate their collective utility in predicting high-grade lesions (CIN2+), which are crucial for timely intervention.

MATERIALS AND METHODS

A total of 781 patients who underwent colposcopy in the gynecological oncology clinic of a tertiary hospital between January 2020 and April 2024 were included in this study. Considering the cytological evaluation and HPV status of these patients prior to the colposcopy procedure, the results of their cervical biopsy, endocervical canal curettage, and probe curettage performed during colposcopy were statistically evaluated. The pathological diagnoses based on cervical biopsy were categorized as CIN I (mild dysplasia; 170 cases), CIN II (moderate dysplasia; 16 cases), CIN III (severe dysplasia and carcinoma in situ; 73 cases), and cervical cancer (6 cases) (19).

The patients who met the following criteria were included in this research:

- Detection of HR-HPV DNA and colposcopy;
- Subjective symptoms such as bleeding after sexual intercourse or increased vaginal discharge;
- Comprehensive clinical and imaging data;
- Age of 18 years or older and sexual experience;
- Patient agreement to participate in the current research.

The following criteria were used to exclude patients from the study:

- Administration of radiotherapy and/or chemotherapy;
- Pregnancy or lactation;
- Previous cervical surgery;
- Other gynecological malignancies;
- Infections caused by other viruses;
- Infectious lesions in the vagina;
- Autoimmune disorders;
- History of hysterectomy;
- Less than 24 hours since last sexual intercourse;
- Within 48 hours of using vaginal medication;
- Ongoing menstruation.

In this study, SPSS 26 was used for statistical analysis of the data. For continuous variables, the mean ± standard deviation was used to describe the distribution. Categorical variables were expressed as frequencies and percentages. The performance of age, HPV status, and single or multiple HPV infections in predicting the presence of CIN 2 and higher lesions on cervical biopsy was analyzed using logistic regression. The model's performance in predicting CIN 2 and above lesions was visualized by the area under the ROC curve (AUC). The statistical significance level was set at $p < 0.05$ and two-tailed.

This study was conducted in accordance with the ethical principles of the Declaration of Helsinki. Ethical approval was obtained from the Ethics Committee of Necmettin Erbakan University Faculty of Medicine (Approval No: 19293, Date: 17.05.2024). Since the study was retrospective in nature, all data were collected anonymously, and informed consent was deemed not required as per the ethics board's decision.

RESULTS

In this study, 781 patients who underwent colposcopy were analyzed. The mean age of the patients was 42.7 ± 9.48 years. The distribution of HPV positivity, additional HPV presence, smear results, cervical biopsy results, Endocervical Canal Curettage (ECC) results and Probe Curettage (P/C) results is presented in Table 1.

In Table 2, the distribution of cervical smear results according to HPV status is presented. The striking situation here is that HPV DNA positivity increases the cytological ASCUS status.

In Table 3, the distribution of cervical biopsy results according to HPV status is presented. It is noteworthy that HPV 16 DNA-positive patients constitute 20.3% of the CIN 2 and above lesions in cervical biopsy results.

In Table 4, the distribution of endocervical curettage (ECC) results according to HPV status is presented. Again, the association of HPV DNA 16 with lesions of CIN 2 and above was observed to increase.

In Table 5, the distribution of probe curettage (PC) results according to HPV status is presented.

Table 1. Clinical features and histopathological findings in a cohort of patients undergoing colposcopy

Variable (n=781)	**Mean ± SD, n (%)**
Age (year)	42.7 ± 9.48
HPV	
HPV 16	277 (35.5%)
HPV18	72 (9.2%)
Unknown	56 (7.2%)
Other	283 (36.2%)
Negative	93 (11.9%)
Presence of additional HPV	
HPV18	18 (2.7%)
Other	78 (11.9%)
None	560 (85.4%)
Smear Results	
AGC	1 (0.1%)
ASC-H	13 (1.7%)
ASCUS	199 (25.5%)
Unknown	81 (10.4%)
Inflammation	20 (2.6%)
HSIL	2 (0.3%)
LSIL	76 (9.7%)
Negative	378 (48.4%)
Insufficient	11 (1.4%)
Cervical Biopsy Results	
Not Taken	115 (14.7%)
Benign	401 (51.3%)
CIN1	170 (21.8%)
CIN2	16 (2.0%)
CIN3	73 (9.3%)
Malignant	6 (0.8%)
ECC	
Not Taken	309 (39.6%)
Benign	435 (55.7%)
CIN1	14 (1.8%)
CIN2	1 (0.1%)
CIN3	20 (2.6%)
Malignant	2 (0.3%)
PC	
Not Taken	700 (89.6%)
Benign	76 (9.7%)
Premalignant	4 (0.5%)
Malignant	1 (0.1%)

AGC: Atypical Glandular Cells; ASC-H: Atypical Squamous Cells-High; ASCUS: Atypical Squamous Cells Undetermined Significance; HSIL: High Grade Squamous Intraepithelial Lesion; LSIL: Low Grade Squamous Intraepithelial Lesion; CIN: Cervical Intraepithelial Neoplasia; ECC: Endocervical Canal Curettage; PC: Probe Curettage.

Table 2. Distribution of cervical smear results according to HPV status

Smear	HPV Status				
	HPV Negative	HPV 16	HPV 18	HPV Other	HPV Unknown
Inadequate	2 (2.2%)	7 (2.5%)	1 (1.4%)	1 (0.4%)	0 (0.0%)
Negative	54 (58.1%)	149 (53.8%)	44 (61.1%)	123 (43.5%)	8 (14.3%)
Inflammation	1 (1.1%)	10 (3.6%)	2 (2.8%)	7 (2.5%)	0 (0.0%)
AGC	0 (0.0%)	1 (0.4%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
ASCUS	15 (16.1%)	55 (19.9%)	18 (25.0%)	107 (37.8%)	4 (7.1%)
LSIL	17 (18.3%)	23 (8.3%)	2 (2.8%)	32 (11.3%)	2 (3.6%)
ASC-H	3 (3.2%)	6 (2.2%)	0 (0.0%)	4 (1.4%)	0 (0.0%)
HSIL	0 (0.0%)	1 (0.4%)	0 (0.0%)	1 (0.4%)	0 (0.0%)
Unknown	1 (1.1%)	25 (9.0%)	5 (6.9%)	8 (2.8%)	42 (75.0%)

AGC: Atypical Glandular Cells, ASC-H: Atypical Squamous Cells-High, ASCUS: Atypical Squamous Cells Undetermined Significance, HSIL: High Grade Squamous Intraepithelial Lesion, LSIL: Low Grade Squamous Intraepithelial Lesion.

Table 3. Distribution of cervical biopsy results according to HPV status

Cervical Bx	HPV Status				
	HPV Negative	HPV 16	HPV 18	HPV Other	HPV Unknown
Not received	13 (14.0%)	32 (11.6%)	12 (16.7%)	56 (19.8%)	2 (3.6%)
Benign	59 (63.4%)	123 (44.4%)	38 (52.8%)	143 (50.5%)	38 (67.9%)
CIN 1	20 (21.5%)	66 (23.8%)	16 (22.2%)	58 (20.5%)	10 (17.9%)
CIN 2	0 (0.0%)	6 (2.2%)	2 (2.8%)	7 (2.5%)	1 (1.8%)
CIN 3	1 (1.1%)	47 (17.0%)	3 (4.2%)	18 (6.4%)	4 (7.1%)
Malignant	0 (0.0%)	3 (1.1%)	1 (1.4%)	1 (0.4%)	1 (1.8%)

Bx: Biopsy, CIN: Cervical intraepithelial neoplasia.

Table 4. Distribution of endocervical curettage results according to HPV status

ECC	HPV Status				
	HPV Negative	HPV 16	HPV 18	HPV Other	HPV Unknown
Not received	38 (40.9%)	117 (42.2%)	11 (15.3%)	120 (42.4%)	23 (41.1%)
Benign	54 (58.1%)	140 (50.5%)	59 (81.9%)	151 (53.4%)	31 (55.4%)
CIN 1	0 (0.0%)	4 (1.4%)	1 (1.4%)	8 (2.8%)	1 (1.8%)
CIN 2	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (0.4%)	0 (0.0%)
CIN 3	1 (1.1%)	16 (5.8%)	0 (0.0%)	3 (1.1%)	0 (0.0%)
Malignant	0 (0.0%)	0 (0.0%)	1 (1.4%)	0 (0.0%)	1 (1.8%)

ECC: Endocervical curettage, CIN: Cervical intraepithelial neoplasia.

Table 5. Distribution of probe curettage results according to HPV status

Probe curettage	HPV Status				
	HPV Negative	HPV 16	HPV 18	HPV Other	HPV Unknown
Not Taken	76 (81,7%)	254 (91,7%)	68 (94,4%)	253 (89,4%)	49 (87,5%)
Benign	16 (17,2%)	22 (7,9%)	3 (4,2%)	29 (10,2%)	6 (10,7%)
Premalignant	1 (1,1%)	1 (0,4%)	0 (0,0%)	1 (0,4%)	1 (1,8%)
Malignant	0 (0,0%)	0 (0,0%)	1 (1,4%)	0 (0,0%)	0 (0,0%)

Table 6. Logistic multivariate regression analysis to predict CIN2 and above lesions

Predictors	Estimate	SE	Z	p	Odds Ratio	95% CI Lower	95% CI Upper
Intercept	-3.83	1.11	-3.42	0.001	0.02	0.002	0.194
Age	-0.01	0.01	-1.37	0.169	0.98	0.960	1.007
HPV status							
HPV 16- Negative	3.10	1.02	3.04	**0.002**	22.36	3.024	165.361
HPV18- Negative	2.08	1.09	1.90	0.057	8.07	0.940	69.346
Unknown- Negative	2.41	1.09	2.20	**0.027**	11.21	1.311	95.820
Other HPV- Negative	2.24	1.02	2.18	**0.029**	9.45	1.265	70.721
Multiple HPV inf. status							
Multi HPV inf. - Single HPV inf.	0.17	0.30	0.55	0.580	1.18	0.647	2.178

Logistic multivariate regression analysis to predict CIN2 and above lesions revealed several important predictors. Several model fit measures were used to assess the overall fit of the model. The deviation value was 541, the AIC (Akaike Information Criterion) value was 555 and the BIC (Bayesian Information Criterion) value was 587. Cox & Snell R^2 value was 0.0648 and Nagelkerke R^2 value was 0.0896. The chi-square (χ^2) value testing the overall goodness of fit of the model was 37.5 with 6 degrees of freedom and a p-value less than 0.001. These results show that the model is statistically significant and explains the data well (N=781). Age was not a significant predictor, with an odds ratio of 0.98 (95% CI: 0.960 to 1.007, p=0.169). Regarding HPV status, HPV 16 was a significant predictor with an odds ratio of 22.36 (95% CI: 3.024 to 165.361, p=0.002) compared to HPV Negative. Compared to HPV Negative, HPV 18 approached significance with an odds ratio of 8.07 (95% CI: 0.940 to 69.346, p=0.057). Compared to HPV Negative, unknown HPV was significant with an odds ratio of 11.21 (95% CI: 1.311 to 95.820, p=0.027). Other HPV types were also significant, with an odds ratio of 9.45 (95% CI: 1.265 to 70.721, p=0.029) compared to HPV Negative. Multiple HPV infections, compared to single HPV infection, were not significant with an odds ratio of 1.18 (95% CI: 0.647 to 2.178, p = 0.580) (Table 6). The model's performance in detecting CIN2 and above lesions is illustrated in Figure 1, with an AUC (Area Under the Curve) of 0.679, indicating a moderate level of diagnostic accuracy. The results of the logistic

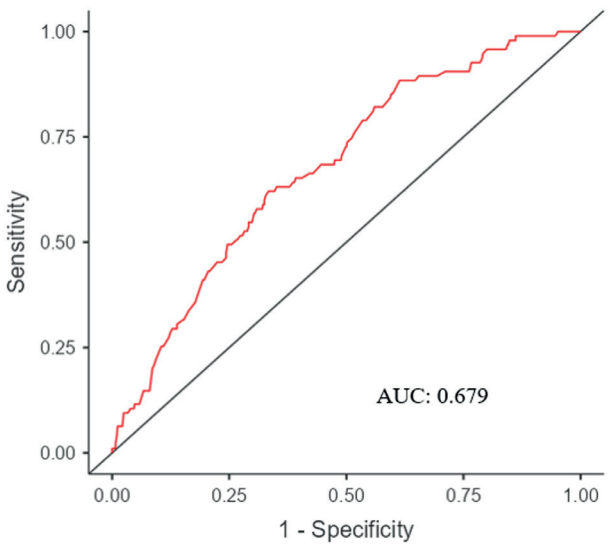


Figure 1. Performance of the logistic regression analysis model in predicting CIN2 and above lesions by ROC Curve

regression analysis show that the model is statistically significant ($p<0.05$) and the AUC value for detecting CIN2 and above lesions is 0.679. This value indicates a moderate level of diagnostic accuracy.

DISCUSSION

Previous studies have consistently demonstrated that high-risk HPV infections are closely linked to the development of cervical precancerous changes and invasive cervical cancer. Moreover, nearly all cervical cancer specimens have been found to contain high-risk HPV DNA, reinforcing its pivotal role in the carcinogenic process (20). Therefore, the present study focused on evaluating the prevalence of HR-HPV positivity across different categories of cervical lesions. Our findings revealed that HR-HPV positivity was significantly more frequent in CIN2+ lesions compared to CIN1 and benign biopsy outcomes. This observation aligns with findings reported in a previous study, which demonstrated a similar distribution of HR-HPV among high-grade lesions (21).

Notably, patients who tested positive for HPV 16 DNA had a 22.36-fold increased risk of developing CIN2 or more severe lesions compared to those without HPV infection.

Interestingly, individuals with unknown HPV DNA status were also found to have an 11.21 times greater likelihood of harboring CIN2+ lesions compared to HPV-negative counterparts. Salvadó et al. reported that HPV 16 infection, along with a history of HSIL cytology, were significant predictors for the persistence or advancement of CIN2 lesions (22). This finding is in agreement with the results of our study.

In another study by Tian et al., HPV status, somatic mutations, and copy number variations were used as important biomarkers for risk classification of CIN2+ lesions. Machine learning algorithms successfully classified the risk of lesions using these biomarkers (23). These findings support the approach of our study, which also utilized predictive modeling to estimate the risk of CIN2+ lesions prior to diagnostic procedures. The AUC value (0.679) calculated from our logistic regression model indicates a moderate diagnostic capability for identifying CIN2 and more advanced lesions. This result underscores the clinical relevance of HPV genotyping, especially in the context of triaging patients for further evaluation. It also highlights the need for close monitoring of patients with HPV 16 positivity, given the elevated risk levels associated with this genotype.

In a study by Quint et al., it was observed that most CIN lesions were linked to a single HPV genotype, even in the presence of multiple infections. This suggests that individual lesions may be driven predominantly by one high-risk type, which is consistent with our findings (24). The strong association between HPV 16 and CIN2+ lesions observed in our cohort further reinforces this conclusion.

This relationship holds particular significance for family physicians managing primary care and gynecologists working in secondary care settings. Identifying HPV 16-positive patients should prompt referral for colposcopic evaluation by experienced gynecologic oncology specialists. Our findings confirm that HPV 16 is strongly correlated with higher-grade cervical lesions (OR: 22.36, $p=0.002$). This highlights its importance as a clinical marker and underscores the necessity for vigilant follow-up strategies in this population.

Since cervical precancerous lesions are often asymptomatic and lack distinct visual features, early detection is particularly challenging. Colposcopy remains a valuable, non-invasive diagnostic tool that enhances biopsy precision and reduces the likelihood of diagnostic errors (25). When performed by trained specialists, colposcopy allows magnified visualization of cervical surface changes, facilitating targeted biopsies based on vascular and epithelial morphology. Previous research has shown that, despite its utility, colposcopy may sometimes yield inaccurate results due to variability in operator experience and interpretation (26-28). The subjective nature of the procedure can influence diagnostic accuracy, emphasizing the importance of integrating objective markers such as HPV typing into clinical workflows (29-31).

HR-HPV testing has become increasingly vital in cervical cancer screening due to its improved sensitivity compared to cytology. It also allows for risk

stratification in patients with ambiguous cytological findings, reduces the testing burden in HPV-negative individuals, and contributes to a more efficient screening process overall (32-34).

CONCLUSION

The combination of HR-HPV testing and colposcopy contributes significantly to the early detection of cervical cancer and its precancerous stages. Moreover, a higher HR-HPV positivity rate appears to correlate with increased lesion severity. A limitation of this study is its retrospective design and the variability in colposcopic assessment due to different gynecologic oncologists performing the procedures.

Ethical approval

This study has been approved by the Necmettin Erbakan University Faculty of Medicine Ethics Committee (approval date 17.05.2024, number 19293).

Author contribution

Surgical and Medical Practices: SÖ, RŞ; Concept: SÖ; Design: AA; Data Collection or Processing: ECS; Analysis or Interpretation: FA; Literature Search: SÖ; Writing: SÖ. All authors reviewed the results and approved the final version of the article.

Source of funding

The authors declare the study received no funding.

Conflict of interest

The authors declare that there is no conflict of interest.

REFERENCES

1. Singh D, Vignat J, Lorenzoni V, et al. Global estimates of incidence and mortality of cervical cancer in 2020: a baseline analysis of the WHO Global Cervical Cancer Elimination Initiative. Lancet Glob Health. 2023; 11(2): e197-206. [Crossref]
2. Li M, Du X, Lu M, et al. Prevalence characteristics of single and multiple HPV infections in women with cervical cancer and precancerous lesions in Beijing, China. J Med Virol. 2019; 91(3): 473-81. [Crossref]
3. Lin S, Gao K, Gu S, et al. Worldwide trends in cervical cancer incidence and mortality, with predictions for the next 15 years. Cancer. 2021; 127(21): 4030-9. [Crossref]
4. Shami S, Coombs J. Cervical cancer screening guidelines: An update. JAAPA. 2021; 34(9): 21-4. [Crossref]
5. Awolude OA, Oyerinde SO, Ayeni AO, Adewole IF. Human papillomavirus-based cervical precancer screening with visual inspection with acetic acid triage to achieve same-day treatments among women living with human immunodeficiency virus infection: test-of-concept study in Ibadan, Nigeria. Pan Afr Med J. 2021; 40: 48. [Crossref]
6. Zhao H, He Y, Fan B, Wang Y, Wu YM. Human papillomavirus E6E7 mRNA and TERC lncRNA in situ detection in cervical scraped cells and cervical disease progression assessment. Virol J. 2022; 19(1): 18. [Crossref]
7. Dong A, Xu B, Wang Z, Miao X. Survival-related DLEU1 is associated with HPV infection status and serves as a biomarker in HPV-infected cervical cancer. Mol Med Rep. 2022; 25(3): 77. [Crossref]
8. Lyu Y, Ding L, Gao T, et al. Influencing Factors of High-Risk Human Papillomavirus Infection and DNA Load According to the Severity of Cervical Lesions in Female Coal Mine Workers of China. J Cancer. 2019; 10(23): 5764-9. [Crossref]
9. Horn J, Denecke A, Luyten A, et al. Reduction of cervical cancer incidence within a primary HPV screening pilot project (WOLPHSCREEN) in Wolfsburg, Germany. Br J Cancer. 2019; 120(10): 1015-22. [Crossref]
10. Jary A, Teguete I, Sidibé Y, et al. Prevalence of cervical HPV infection, sexually transmitted infections and associated antimicrobial resistance in women attending cervical cancer screening in Mali. Int J Infect Dis. 2021; 108: 610-6. [Crossref]
11. Hu H, Zhao J, Yu W, et al. Human papillomavirus DNA, HPV L1 capsid protein and p16(INK4a) protein as markers to predict cervical lesion progression. Arch Gynecol Obstet. 2019; 299(1): 141-9. [Crossref]
12. Bedell SL, Goldstein LS, Goldstein AR, Goldstein AT. Cervical Cancer Screening: Past, Present, and Future. Sex Med Rev. 2020; 8(1): 28-37. [Crossref]
13. Chen L, Dong B, Zhang Q, et al. HR-HPV viral load quality detection provide more accurate prediction for residual lesions after treatment: a prospective cohort study in patients with high-grade squamous lesions or worse. Med Oncol. 2020; 37(5): 37. [Crossref]
14. Sargent A, Fletcher S, Bray K, Kitchener HC, Crosbie EJ. Cross-sectional study of HPV testing in self-sampled urine and comparison with matched vaginal and cervical samples in women attending colposcopy for the management of abnormal cervical screening. BMJ Open. 2019; 9(4): e025388. [Crossref]
15. Gunes AC, Ozgul N, Turkyılmaz M, et al. Evaluation of colposcopy after the addition of human papillomavirus testing to the Turkish cervical cancer screening program. Cancer Med. 2023; 12(24): 21751-60. [Crossref]

16. Hernández-López R, Lorincz AT, Torres-Ibarra L, et al. Methylation estimates the risk of precancer in HPV-infected women with discrepant results between cytology and HPV16/18 genotyping. Clin Epigenetics. 2019; 11(1): 140. [Crossref]

17. Okunade KS. Human papillomavirus and cervical cancer. J Obstet Gynaecol. 2020; 40(5): 602-8. [Crossref]

18. Santesso N, Mustafa RA, Schünemann HJ, et al. World Health Organization Guidelines for treatment of cervical intraepithelial neoplasia 2-3 and screen-and-treat strategies to prevent cervical cancer. Int J Gynaecol Obstet. 2016; 132(3): 252-8. [Crossref]

19. Fontham ETH, Wolf AMD, Church TR, et al. Cervical cancer screening for individuals at average risk: 2020 guideline update from the American Cancer Society. CA Cancer J Clin. 2020; 70(5): 321-46. [Crossref]

20. McCluggage WG, Judge MJ, Alvarado-Cabrero I, et al. Data Set for the Reporting of Carcinomas of the Cervix: Recommendations From the International Collaboration on Cancer Reporting (ICCR). Int J Gynecol Pathol. 2018; 37(3): 205-28. [Crossref]

21. Wang P, Gao D, Yu X, Zhu G. Value of high-risk human papillomavirus detection combined with colposcopy in the diagnosis of cervical cancer and precancerous lesions. Oncol Lett. 2024; 27(4): 185. [Crossref]

22. Salvadó A, Miralpeix E, Solé-Sedeno JM, et al. Predictor factors for conservative management of cervical intraepithelial neoplasia grade 2: Cytology and HPV genotyping. Gynecol Oncol. 2021; 162(3): 569-74. [Crossref]

23. Tian R, Cui Z, He D, et al. Risk stratification of cervical lesions using capture sequencing and machine learning method based on HPV and human integrated genomic profiles. Carcinogenesis. 2019; 40(10): 1220-8. [Crossref]

24. Quint W, Jenkins D, Molijn A, et al. One virus, one lesion--individual components of CIN lesions contain a specific HPV type. J Pathol. 2012; 227(1): 62-71. [Crossref]

25. Wittenborn J, Weikert L, Hangarter B, Stickeler E, Maurer J. The use of micro RNA in the early detection of cervical intraepithelial neoplasia. Carcinogenesis. 2020; 41(12): 1781-9. [Crossref]

26. Luo Q, Lang L, Han N, Liang L, Shen L, Zhang H. Prevalence and genotype distribution of high-risk human papillomavirus infection among women with cervical cytological abnormalities in Chongqing, China, 2014-2020. Diagn Cytopathol. 2021; 49(12): 1237-43. [Crossref]

27. Painter H, Erlinger A, Simon B, Morroni C, Ramogola-Masire D, Luckett R. Impact of cervicitis on performance of cervical cancer screening using HRHPV testing and visual evaluation in women living with HIV in Botswana. Int J Gynaecol Obstet. 2020; 151(1): 144-6. [Crossref]

28. Liu G, Sharma M, Tan N, Barnabas RV. HIV-positive women have higher risk of human papilloma virus infection, precancerous lesions, and cervical cancer. AIDS. 2018; 32(6): 795-808. [Crossref]

29. Vodicka EL, Chung MH, Zimmermann MR, et al. Estimating the costs of HIV clinic integrated versus non-integrated treatment of pre-cancerous cervical lesions and costs of cervical cancer treatment in Kenya. PLoS One. 2019; 14(6): e0217331. [Crossref]

30. Husaiyin S, Han L, Wang L, et al. Factors associated with high-risk HPV infection and cervical cancer screening methods among rural Uyghur women aged > 30 years in Xinjiang. BMC Cancer. 2018; 18(1): 1162. [Crossref]

31. Garza-Rodríguez ML, Oyervides-Muñoz MA, Pérez-Maya AA, et al. Analysis of HPV Integrations in Mexican Pre-Tumoral Cervical Lesions Reveal Centromere-Enriched Breakpoints and Abundant Unspecific HPV Regions. Int J Mol Sci. 2021; 22(6): 3242. [Crossref]

32. Dong B, Chen L, Lin W, et al. Cost-effectiveness and accuracy of cervical cancer screening with a high-risk HPV genotyping assay vs a nongenotyping assay in China: an observational cohort study. Cancer Cell Int. 2020; 20: 421. [Crossref]

33. Ferrera A, Valladares W, Cabrera Y, et al. Performance of an HPV 16/18 E6 oncoprotein test for detection of cervical precancer and cancer. Int J Cancer. 2019; 145(8): 2042-50. [Crossref]

34. Suwanthananon C, Inthasorn P. A comparison of the associations of Reid Colposcopic Index and Swede Score with cervical histology. J Obstet Gynaecol Res. 2020; 46(4): 618-24. [Crossref]

NORTHWESTERN MEDICAL JOURNAL

Northwestern Med J. 2025;5(4):209-221
DOI: 10.54307/NWMJ.2025.220

RESEARCH ARTICLE

Evaluation of ChatGPT-4.5 and DeepSeek-V3-R1 in answering patient-centered questions about orthognathic surgery: a comparative study across two languages

İpek Necla Güldiken[1], **Emrah Dilaver**[1]

[1]Department of Oral and Maxillofacial Surgery, Faculty of Dentistry, İstinye University, İstanbul, Türkiye

Cite as: Güldiken İN, Dilaver E. Evaluation of ChatGPT-4.5 and DeepSeek-V3-R1 in answering patient-centered questions about orthognathic surgery: a comparative study across two languages. Northwestern Med J. 2025;5(4):209-221.

ABSTRACT

Aim: Patients undergoing orthognathic surgery frequently seek online resources to better understand the procedure, risks, and outcomes. As generative artificial intelligence (AI) models are increasingly integrated into healthcare communication, it is essential to evaluate their ability to deliver accurate, comprehensive, and readable patient information.

Methods: This study conducted a comparative assessment of two large language models (LLMs)—ChatGPT-4.5 and DeepSeek-V3-R1—in answering frequently asked orthognathic patient questions, analyzing accuracy, completeness, readability, and quality across English (EN) and Turkish (TR). Twenty-five patient-centered questions categorized into five clinical domains yielded 200 AI-generated responses, independently evaluated by two oral and maxillofacial surgeons (OMFSs) using a multidimensional framework. Statistical analyses included non-parametric tests and inter-rater reliability assessments (Intraclass Correlation Coefficient (ICC), and Cohen's Kappa).

Results: Significant differences emerged across clinical categories in difficulty and accuracy scores ($p < 0.05$). Questions in the "Postoperative Complications & Rehabilitation" category were least difficult, while those in "Diagnosis & Indication" category were rated most difficult but achieved the highest accuracy and quality ratings. English (EN) responses significantly outperformed Turkish (TR) responses in readability, word count, and accuracy ($p < 0.05$), though completeness and quality did not differ significantly by language. No significant performance differences were found between the two chatbots. Inter-observer agreement was generally high, except for completeness ($p = 0.001$), where Observer-I assigned higher scores.

Conclusion: Both LLMs effectively generated clinically relevant responses, demonstrating substantial potential as supplemental tools for patient education, although the superior performance of EN responses emphasizes the need for further multilingual optimization.

Keywords: ChatGPT, DeepSeek, large language models, orthognathic surgery, patient education

Corresponding author: İpek Necla Güldiken **E-mail:** ipek.guldiken@istinye.edu.tr
Received: 18.08.2025 **Accepted:** 22.09.2025 **Published:** 26.10.2025

INTRODUCTION

The use of the internet for accessing health-related information has markedly increased in past three years (1). Orthognathic surgery, essential for managing dentofacial deformities, prompts many patients to seek information online due to limited access to specialists and the complexity of the treatment process. Patients typically investigate the surgical procedure, recovery timeline, associated risks, costs, and expected outcomes (2). Research shows that this information-seeking behavior intensifies during the decision-making period and preoperative phase, primarily via medical websites, blogs, and forums (3). Postoperative information needs also persist and are often addressed through professional communication and support networks (4). This trend highlights both the procedure's growing prevalence and the critical need for reliable, accessible information.

Findings over the past two years highlight AI's growing effectiveness and reliability in medical contexts, underscoring its expanding role in patient counselling and information dissemination. Recent studies have assessed how LLMs—including ChatGPT, DeepSeek—respond to patient queries in healthcare (5-9). ChatGPT-4.5 (developed by OpenAI, USA) and DeepSeek-V3-R1 (developed by DeepSeek-AI, China) are increasingly used in healthcare for responding to patient queries, owing to their advanced natural language processing capabilities. ChatGPT-4.5 is the latest model in OpenAI's GPT series, while DeepSeek-V3-R1 is the latest interactive model trained on large-scale datasets in DeepSeek-AI models. Both models stand out for their ability to generate fast and comprehensive responses to complex medical questions (10).

This study aimed to compare ChatGPT-4.5 and DeepSeek-V3-R1 in terms of response accuracy, comprehensiveness, readability, and overall quality for frequently asked patient questions on orthognathic surgery. It also evaluated the effects of language (EN vs. TR) and question difficulty on model performance. As the first study to provide a cross-linguistic comparison of LLM outputs in this context, it underscores the potential of these models to enhance access to trustworthy health information and reduce language-based communication barriers. Supporting standardized, multilingual communication in global healthcare was a key motivation for this research.

MATERIAL AND METHODS

Study design

Since this research did not involve human subjects or personal health data, formal ethical approval was not required. Nonetheless, all testing was conducted in a neutral setting to uphold the integrity of the study. To enhance transparency and ensure methodological consistency, the METRICS framework (Model, Evaluation, Timing, Range/Randomization, Individual factors, Count, and Specificity of prompts and language) was adopted (11). This structured approach also contributes to standardizing AI evaluations in healthcare and minimizing potential sources of bias. The methodological workflow of the study is summarized in Figure 1.

To simulate an average user experience and reduce potential bias, both models were accessed via a newly created Google account using the "Continue with Google" option. Prior to testing, all browsing history, cookies, and cache were deleted by selecting the "Clear Browsing Data" option with the time range set to "All Time".

Question development, categorization and sampling

The sample of patient-centered questions was derived using a purposive sampling strategy, specifically combining expert clinical input with a structured review of existing patient education materials. The authors created a preliminary question pool by adapting content from published literature (2,12-14) and reviewing patient guidelines issued by professional bodies, such as the American Association of Oral and Maxillofacial Surgeons (AAOMS) and the British Association of Oral and Maxillofacial Surgeons (BAOMS). Additional questions were identified through a targeted Google search using keywords such as "orthognathic surgery", "patient FAQs (frequently

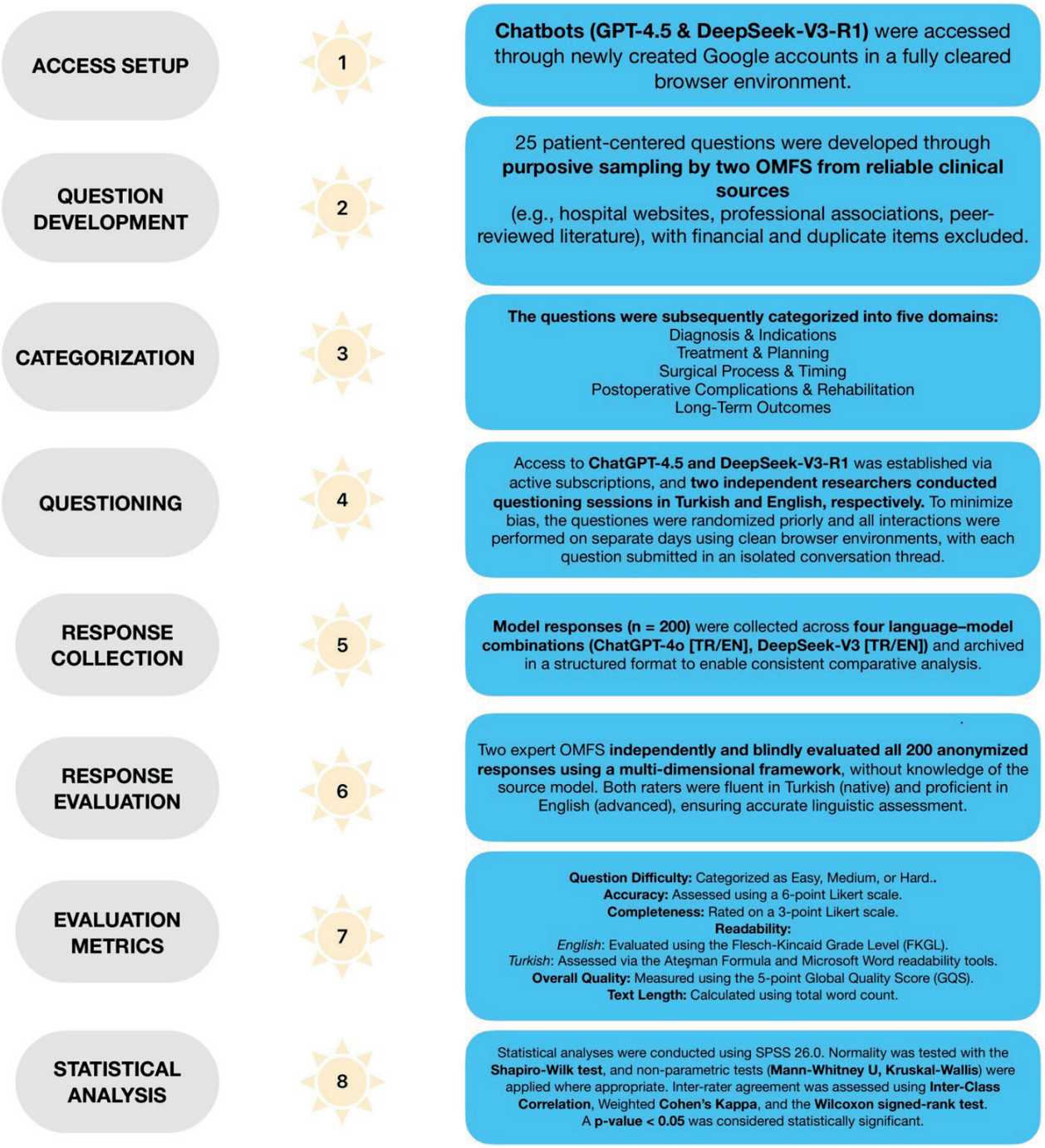


Figure 1. Flowchart of the study.

asked questions)", and "patient guide", focusing on the top-ranking results from reputable hospital and clinical websites, including large academic medical centers and national surgical institutes.

The emphasis was placed on thematic representativeness and content saturation, rather than statistical generalizability, aligning with the principles of purposive sampling as outlined by Etikan et al.

(15). This approach ensured the inclusion of clinically relevant, high-frequency, and informative questions for comparative evaluation across AI models.

The final set of 25 patient-centered questions was selected after excluding those with financial content and removing duplicates (Table 1). To enhance analytical consistency, the questions were ontologically categorized according to the medical process-based framework proposed by Chong et al. (16), which classifies patient inquiries into five distinct domains: (1) Diagnosis & Indication, (2) Treatment & Planning, (3) Surgical Process & Timing, (4) Postoperative Complications & Rehabilitation, and (5) Long-Term Outcomes. This classification facilitated a structured analysis of AI model performance across different stages of the orthognathic surgery care continuum (Figure 2).

Table 1. Patient questions and the categories

ID	Question	Category
1	What exactly is orthognathic surgery?	
2	Who needs orthognathic surgery?	
3	Which conditions can orthognathic surgery address?	
4	Can this surgery be done purely for cosmetic reasons, even if there's no functional issue?	Diagnosis & Indication
5	At what age is orthognathic surgery performed?	
6	How do I know if I need orthognathic surgery vs. just orthodontic treatment?	
7	How do I find or choose a qualified surgeon and orthodontist for my treatment?	
8	Do I need braces before orthognathic surgery?	Treatment & Planning
9	How should I prepare for jaw surgery and how?	
10	How is orthognathic surgery performed?	
11	How long does orthognathic surgery take?	Operation Process & Time
12	What type of anesthesia is used?	
13	Will I need to have my jaws wired shut, and for how long?	
14	What is the typical recovery time after orthognathic surgery?	
15	Is the procedure painful, and what can I expect in terms of post-operative discomfort?	Postoperative Complications & Rehabilitation
16	How do I manage swelling and other potential side effects after surgery?	
17	Is numbness or loss of sensation normal after surgery, and will it go away over time?	
18	What are the potential risks and complications of orthognathic surgery?	
19	How soon can I return to work or school after surgery?	
20	Does having this surgery improve facial appearance as well as jaw function?	
21	Will orthognathic surgery affect speech or eating in the long run?	
22	Are there any long-term lifestyle changes required after orthognathic surgery?	
23	Can I have symmetry disorder in my face after surgery?	Long-Term Results
24	Does this surgery correct breathing or snoring problems?	
25	After orthognathic surgery, may I need to have another operation in the future?	

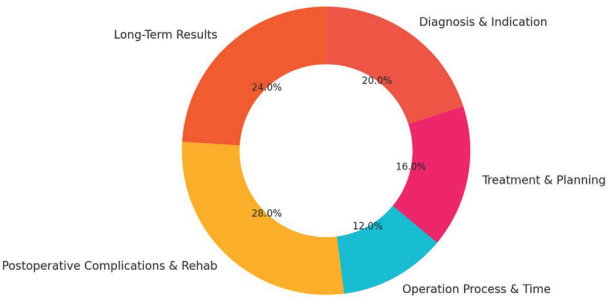


Figure 2. Question distrubution according to the categories.

Questioning

Access to ChatGPT (OpenAI) and DeepSeek AI was obtained through active subscriptions to ChatGPT-4.5 and DeepSeek-V3-R1, which offer improved processing power, priority usage, and increased accuracy through the latest updates. To systematically evaluate the models, two independent researchers conducted the questioning sessions. One researcher interacted with the models in Turkish (TR), while the other did so in English (EN).

All interactions were carried out using newly created user accounts in a clean browsing environment (with cleared cookies, cache, and history) to minimize contextual bias. Each session was conducted on a different day to reduce potential carry-over effects, and every question was submitted in a distinct conversation thread to ensure isolated contextual modeling using newly created accounts and a clear browsing environment.

Response collection

Model outputs were systematically collected across four distinct language–model pairings: ChatGPT (TR), ChatGPT (EN), DeepSeek (TR), and DeepSeek (EN). All responses were preserved in their entirety and archived in a structured format to ensure consistency and enable comparative evaluation.

Response evaluation

Model-generated answers were assessed using a multi-dimensional framework incorporating both quantitative and qualitative indicators (2,14). The evaluation was conducted by two board-certified OMFSs, who independently and blindly scored all 200 responses. The responses—randomized across both models—were anonymized such that evaluators were unaware of the originating model. Both raters possessed advanced proficiency in English and native-level fluency in Turkish, ensuring reliable linguistic judgment across both language sets.

Question difficulty was categorized as easy, medium, or hard based on Goodman et al. (17). Accuracy was rated on a 6-point Likert scale, and completeness on a 3-point scale, with intermediate scores reflecting partial correctness. Readability was assessed using language-specific tools: the Flesch-Kincaid Grade Level (FKGL) for English (7,18) and the Ateşman Formula for Turkish (19-21). Overall answer quality was evaluated using the 5-point Global Quality Score (GQS), reflecting scientific accuracy, clarity, and informativeness (22). Text length was calculated as total word count using Microsoft Word (Microsoft Corporation, Redmond, WA, USA) (see Table 2 for detailed scoring criteria).

The FKGL score is derived from average sentence length and syllables per word, with lower scores indicating simpler, more accessible language (18). For TR responses, readability was evaluated using the Ateşman Formula, a well-established metric adapted from the Flesch Reading Ease Index (FREI) for the TR language. The formula incorporates average word and sentence lengths to generate a numerical readability score, where higher values indicate easier comprehension (19,23). In addition, Microsoft Word's built-in readability analysis was used as a supplementary tool to validate the scoring consistency.

Statistics

All statistical analyses were performed using SPSS version 26.0 (IBM Corp., Armonk, NY, USA). Descriptive statistics were used to summarize participant and response characteristics. The normality of continuous variables was assessed using the Shapiro-Wilk test ($p < 0.001$). For variables that did not follow a normal distribution, non-parametric tests were employed, including the Mann-Whitney U test and Kruskal-Wallis test for group comparisons. The inter-rater agreement between the two evaluators was analyzed using the

Table 2. Evaluation framework for model responses

Evaluation Dimension	Metric/Tool	Scale/Range	Description
Question difficulty	Cathegorical	Easy Medium Hard	Classification based on clinical and linguistic complexity
Accuracy	6-Point Likert Scale	1–6	1 = Completely incorrect 2 = Mostly incorrect 3 = Contains some factual errors 4 = Partially correct 5 = Mostly accurate 6 = Completely accurate
Completeness	3-Point Likert Scale	1–3	1 = Inadequate 2 = Moderately complete 3 = Comprehensive
Readability (English)	Flesch-Kincaid Grade Level (FKGL)	Grade level (e.g., 8.0)	1–5 = Very easy to read 6–8 = Easy to read 9–12 = Fairly difficult / Standard 13–16 = Difficult to read 17+ = Very difficult / Academic or technical material
Readability (Turkish)	Ateşman Formula / Microsoft Word	Score (e.g., 0–100)	90–100 = Very easy to read 70–89 = Easy to read 50–69 = Fairly difficult / Standard 30–49 = Difficult to read 0–29 = Very difficult / Academic or technical material
Overall quality	Global Quality Score (GQS)	1–5	1 = Very poor 2 = Poor 3 = Moderate 4 = Good 5 = Excellent quality
Text length	Microsoft Word (Microsoft Corp., USA)	Numeric word count	Total number of words in the model response

Intraclass Correlation Coefficient (ICC) and Weighted Cohen's Kappa. Furthermore, the inter-observer consistency was appraised by means of the Wilcoxon signed-rank test. All results were interpreted within a 95% confidence interval, and a p-value <0.05 was considered statistically significant.

RESULTS

The majority of the questions were categorized under the headings "Postoperative Complications & Rehabilitation" and "Long-Term Outcomes," with a balanced distribution across language groups, chatbots, and observers. Descriptive statistics for question difficulty, word count, accuracy, completeness, readability, and overall quality are presented in Table 3.

Statistical comparisons revealed significant differences in difficulty scores based on question category ($p < 0.001$), with "Diagnosis & Indication" questions rated as significantly more difficult than those in "Surgical Process & Timing" and "Postoperative Complications & Rehabilitation". Word count did not vary significantly by category. Despite the higher difficulty scores, accuracy scores were also significantly higher in the "Diagnosis & Indication" compared to "Postoperative Complications & Rehabilitation" category ($p = 0.013$) (Table 4).

Table 3. Descriptive statistics about questionnaire

Variables	**Mean. ± S.D.**	**Median (Min.Max.)**
Difficulty	1.9 ± 0.7	2 (1-3)
Word count	289.8 ± 108.3	279 (81-596)
Accuracy	4.9 ± 0.9	5 (3-6)
Completeness	2.7 ± 0.5	3 (2-3)
Readability	3.9 ± 1.1	4 (2-6)
Quality	3.9 ± 0.9	4 (2-5)
Category	**n**	**%**
Diagnosis & Indication	40	20
Treatment & Planning	32	16
Operation process & Time	24	12
Postoperative Complications & Rehabilitation	56	28
Long term results	48	24
Language	**n**	**%**
TR	100	50
ENG	100	50
Chatbot	**n**	**%**
ChatGPT	100	50
DeepSeek	100	50
Observer	**n**	**%**
Observer-I	100	50
Observer-II	100	50

Language comparisons showed that EN responses had significantly higher word counts ($p<0.001$), higher accuracy ($p = 0.047$), and higher readability scores ($p <0.001$) compared to TR responses (Table 5). No statistically significant differences were observed between ChatGPT and DeepSeek models in terms of difficulty, word count, accuracy, completeness, readability, or quality (Table 4). Similarly, completeness and quality scores did not significantly differ by language (Table 5).

A comparison of the evaluators' scores revealed significant discrepancies in the completeness ratings ($p <0.001$), with Observer I assigned higher ratings than Observer II. In a similar vein, Observer I provided substantially higher quality scores. Despite these differences, the inter-rater reliability was found to be strong to perfect across all evaluation criteria, with ICC and weighted Cohen's Kappa values ranging from 0.374 to 1.000 ($p <0.001$ for all metrics). These findings suggest that while subjective interpretation may influence certain dimensions, particularly completeness and quality, the overall scoring framework demonstrated a high level of consistency and reliability between evaluators (Table 5).

DISCUSSION

In this study, responses generated by two large language models—ChatGPT-4.5 and DeepSeek-V3-R1—to patient questions on orthognathic surgery were evaluated using a multidimensional framework. Analysis of 200 outputs showed that both models produced responses with high accuracy and quality, and similar word count and scope. However, EN responses significantly outperformed TR ones in terms of accuracy and readability, likely due to the uneven distribution of training data across languages (24). Notably, the "Diagnosis & Indication" category, despite its higher difficulty level, received the highest accuracy scores—contrary to previous findings (6). This suggests that structured knowledge domains, such as diagnostic content, may enhance model performance even on complex queries.

The methodological framework of this study aligns with prior research in the field (6,17,25,26). In designing the evaluation protocol, several established health content quality assessment guidelines were reviewed, including METRICS, CLEAR (Communication, Language, Evaluation, and Review), and MI-CLEAR-LLM (Minimum Reporting Items for Clear Evaluation of Accuracy Reports of Large Language Models in Healthcare) (11,27,28). The most applicable and pragmatic elements from these frameworks were selectively integrated into the METRICS-based assessment applied in this study.

Table 4. Comparative metrics by question type

			Difficulty			Word count			Accuracy		
Variables		**N**	**Mean. ± S.D.**	**Median (Min.Max.)**	**p**	**Mean. ± S.D.**	**Median (Min.Max.)**	**p**	**Mean. ± S.D.**	**Median (Min.Max.)**	**p**
Category	Diagnosis& indication	40	2.3 ± 0.8	2.5 (1-3)	p<0.001	289.6 ± 110.5	281.5 (141-593)	0.117	5.2 ± 0.6	5 (4-6)	0.013*
	Treatment& planning	32	2.0 ± 0.7	2 (1-3)		330.0 ±116.5	324 (189-593)		4.8 ± 1.1	5 (3-6)	
	Surgical process& time	24	1.7 ± 0.8	1.5 (1-3)		294.1 ± 134.8	268.5 (149-593)		5.0 ± 0.8	5 (4-6)	
	Postoperative complication & rehabilitation	56	1.4 ± 0.5	1 (1-3)		262.0 ± 107.7	238 (81-489)		4.6 ± 0.9	5 (3-6)	
	Long-term results	48	2.1 ± 0.5	2 (1-3)		293.5 ± 77.6	296 (130-436)		5.0 ± 0.9	5 (3-6)	
Language	TR	100	1.8 ± 0.7	2 (1-3)	0.069	255.3 ± 101.9	230 (81-593)	p<0.001	4.8 ± 0.9	5 (3-6)	0.047*
	EN	100	2.0 ± 0.8	2 (1-3)		324.3 ± 103.8	333.5 (134-596)		5.0 ± 0.8	5 (3-6)	
Chatbot	ChatGPT	100	1.9 ± 0.7	2 (1-3)	0.727	278.1 ± 96.8	263.5 (134-593)	0.107	4.9 ± 0.8	5 (3-6)	0.276
	DeepSeek	100	1.8 ± 0.7	2 (1-3)		301.5 ± 117.9	292 (81-596)		5.0 ± 0.9	5 (3-6)	
Obserever	Observer-I	100	1.9 ± 0.7	2 (1-3)	1.000	289.8 ± 108.5	279 (81-596)	1.000	5.0 ± 1.0	5 (3-6)	0.068
	Observer-II	100	1.9 ± 0.7	2 (1-3)		289.8 ± 108.5	279		4.8 ± 0.8	5 (3-6)	

*p<0.05

Table 5. Comparisons of completeness, readability and quality on characteristic specialities of questionnaires

			Completeness			Readability			Quality		
Variables		**N**	**Mean. ± S.D.**	**Median (Min.Max.)**	**p**	**Mean. ± S.D.**	**Median (Min.Max.)**	**p**	**Mean. ± S.D.**	**Median (Min.Max.)**	**p**
Catergory	Diagnosis & Indication	40	2.7 ±0.5	3 (2-3)	0.159	4.1 ± 1.1	4 (2-6)	0.001*	4.2 ± 0.7	4 (2-5)	0.026*
	Treatment & Planning	32	2.7 ±0.5	3 (2-3)		3.8 ± 0.9	4 (2-5)		3.8 ±1.0	4 (2-5)	
	Surgical process & Time	24	2.5 ± 0.5	3 (2-3)		3.8 ± 0.9	3.5 (3-5)		4.2 ±0.7	4 (3-5)	
	Postoperative complication & Rehabilitation	56	2.8 ± 0.4	3 (2-3)		3.5 ±1.2	3 (2-6)		3.7 ±0.9	4 (2-5)	
	Long-term results	48	2.5 ± 0.5	3 (2-3)		4.3 ± 0.9	4 (3-6)		3.8 ±0.8	4 (2-5)	
Language	TR	100	2.6 ± 0.5	3 (2-3)	0.237	3.2 ± 0.7	3 (2-5)	p<0.001	3.8 ± 0.9	4 (2-5)	0.236
	ENG	100	2.7 ± 0.5	3 (2-3)		4.6 ± 0.9	5 (3-6)		4.0 ± 0.8	4 (2-5)	
Chatbot	ChatGPT	100	2.7 ± 0.5	3 (2-3)	1.000	3.8 ± 1.0	3.5 (2-6)	0.370	3.8 ± 0.9	4 (2-5)	0.105
	DeePSeeK	100	2.7 ± 0.5	3 (2-3)		3.9 ± 1.1	4 (2-6)		4.0 ± 0.8	4 (2-5)	
Observer	Observer-I	100	2.8 ± 0.4	3 (2-3)	0.001*	3.9 ± 1.1	4 (2-6)	1.000	4.0 ±0.9	4 (2-5)	0.170
	Observer-II	100	2.5 ± 0.5	3 (2-3)		3.9 ± 1.1	4 (2-6)		3.8 ±0.8	4 (2-5)	

*p<0.05

Previous studies have shown that LLMs generally provide satisfactory responses to medical questions intended for patient education. Goodman et al. and Beheshti et al. reported that models such as ChatGPT are capable of generating accurate and useful medical content (17,29). Comparative evaluations of various LLMs—such as ChatGPT, Gemini, Bard, Claude, Copilot, and DeepSeek—indicate that ChatGPT is the most frequently studied model, with English being the dominant language of analysis (5-9,22). While direct comparisons between ChatGPT and DeepSeek remain limited, it has been showed that both models performed similarly when responding to complex, multi-domain queries—an observation that aligns with our own findings (10,23).

In the present study, open-ended, prompt-free questions were used to simulate a realistic conversational setting. The impact of prompt usage and question format on model performance has also been addressed in the literature, particularly in healthcare, where the absence of prompts in open-ended queries has been linked to accuracy issues (e.g., hallucination effects) (6,7,14,30-32). In their systematic review, Beheshti et al. highlighted that many studies failed to evaluate the influence of prompt design (29). Nevertheless, further research is warranted to clarify the role of prompt engineering in medical applications.

This study offers one of the first comparative analyses of two LLMs across both EN and TR, addressing the gap in the literature where evaluations are often limited to a single language (2,17,26). The dominance of EN in model training and the lack of standardized cross-linguistic evaluation metrics contribute to this limitation (24).

Readability was assessed using language-specific tools: the FKGL for EN and the Ateşman Formula for Turkish (a well-established metric adapted from the FREI) (18,21). To our knowledge, this is the first study to directly compare the readability of ChatGPT and DeepSeek responses in TR versus EN. While the American Medical Association (AMA) and National Institutes of Health (NIH) recommend health materials be written at a sixth-grade reading level (4), our findings, consistent with previous work (14,26), show that most responses exceed this threshold. Notably, TR responses exhibited lower (i.e., better) readability scores, suggesting improved accessibility for native speakers.

Prompt usage has been shown to enhance factual accuracy but not necessarily readability (7). Overall, the higher accuracy and readability of EN responses likely reflect the greater volume of training data in EN (24). These findings underscore the need for multilingual training and evaluation strategies, especially for low-resource languages like TR. Moreover, differences in text length and linguistic structure may also influence readability outcomes. By addressing the bilingual evaluation gap, this study contributes to promote equitable and comprehensible health communication (2,14,26).

Previous research indicates that online sources, including social media, provide predominantly low-quality orthognathic surgery information, characterized by subjective patient experiences (33). Bavbek et al. particularly highlighted the poor quality of Turkish-language online resources on orthognathic surgery (3). However, recent advances in AI-based chatbots, such as ChatGPT and Google Gemini, have significantly enhanced content reliability by incorporating academic literature and professional guidelines (2,14). Despite these improvements, issues regarding readability and accuracy persist, warranting cautious integration into clinical practice (4). Notably, no prior studies have evaluated DeepSeek's performance in orthognathic patient education, emphasizing our study's novel contribution in comparing model performance across languages, particularly in TR.

This study offers several strengths, including the evaluation of four model-language combinations, the ontological categorization of patient questions, blinded expert-based scoring, and a multidimensional assessment framework encompassing accuracy, completeness, readability, and quality. Furthermore, although there is currently no consensus on standardized criteria for evaluating chatbot performance in health literacy contexts (11,14,27,28), this study addressed that gap by adapting elements from existing guidelines—particularly the METRICS

framework—to guide its evaluation process. However, certain limitations of this study should be acknowledged.

First, the limited sample size may not fully capture the range of responses these models can generate. Both question development and scoring were conducted by two OMFSs from the same institution, potentially introducing selection bias. Nonetheless, the inclusion of inter-rater reliability analysis partially mitigates this limitation.

The reproducibility of model outputs could not be assessed, as each question was asked only once and sessions were not repeated. Additionally, restricting the study to only two widely used LLMs limits the generalizability of the findings. The lack of transparency in source usage and inconsistent citation practices may also affect the perceived accuracy of content (14,29). Although both models exhibited citation behaviors, inconsistencies in source attribution prevented statistical analysis. Moreover, the assertive tone adopted by LLMs may create a false sense of confidence among users, which has been noted in previous studies (17,29).

Finally, while this study aimed to address the limitations of monolingual evaluations by including both EN and TR, structural differences between the two languages pose inherent challenges for direct comparison. Despite these limitations, this study remains one of the few bilingual and multidimensional evaluations of LLMs in the specialized field of orthognathic surgery, offering a foundation for future research.

CONCLUSIONS AND SUGGESTIONS

This study demonstrated that large language models, specifically ChatGPT-4.5 and DeepSeek-V3-R1, are capable of producing accurate and clinically relevant responses to patient-centered questions in orthognathic surgery. While both models performed similarly in overall quality, responses in EN showed higher readability and accuracy than those in TR. Differences across clinical categories and the presence of moderate inter-observer variability emphasize the need for standardized evaluation frameworks. With careful implementation and ongoing validation, LLMs may serve as valuable tools to support patient education and preoperative communication in oral and maxillofacial surgery settings.

Acknowledgments

The authors would like to thank the Department of Oral and Maxillofacial Surgery at İstinye University for supporting this research.

Ethical approval

The authors confirm that an ethical waiver was obtained due to the study design (retrospective evaluation of publicly accessible data without direct patient interaction), and that the study was conducted according to the Helsinki Declaration (2013).

Author contribution

Surgical and Medical Practices: ING, ED; Concept: ING, ED; Design: ING, ED; Data Collection or Processing: ING, ED; Analysis or Interpretation: ING, ED; Literature Search: ING, ED; Writing: ING. All authors reviewed the results and approved the final version of the article.

Source of funding

The authors declare the study received no funding.

Conflict of interest

The authors declare that there is no conflict of interest.

REFERENCES

1. Su H, Sun Y, Li R, et al. Large Language Models in Medical Diagnostics: Scoping Review With Bibliometric Analysis. J Med Internet Res. 2025; 27: e72062. **[Crossref]**
2. Aziz AAA, Abdelrahman HH, Hassan MG. The use of ChatGPT and Google Gemini in responding to orthognathic surgery-related questions: A comparative study. J World Fed Orthod. 2025; 14(1): 20-6. **[Crossref]**
3. Bavbek NC, Tuncer BB. Information on the Internet Regarding Orthognathic Surgery in Turkey: Is It an Adequate Guide for Potential Patients? Turk J Orthod. 2017; 30(3): 78-83. **[Crossref]**

4. Balel Y. Can ChatGPT be used in oral and maxillofacial surgery? J Stomatol Oral Maxillofac Surg. 2023; 124(5): 101471. [Crossref]
5. Dursun D, Bilici Geçer R. Can artificial intelligence models serve as patient information consultants in orthodontics? BMC Med Inform Decis Mak. 2024; 24(1): 211. [Crossref]
6. Gumilar KE, Indraprasta BR, Faridzi AS, et al. Assessment of Large Language Models (LLMs) in decision-making support for gynecologic oncology. Comput Struct Biotechnol J. 2024; 23: 4019-26. [Crossref]
7. Gordon EB, Towbin AJ, Wingrove P, et al. Enhancing Patient Communication With Chat-GPT in Radiology: Evaluating the Efficacy and Readability of Answers to Common Imaging-Related Questions. J Am Coll Radiol. 2024; 21(2): 353-9. [Crossref]
8. Metin U, Goymen M. Information from digital and human sources: A comparison of chatbot and clinician responses to orthodontic questions. Am J Orthod Dentofacial Orthop. 2025; 168(3): 348-57. [Crossref]
9. Daraqel B, Wafaie K, Mohammed H, et al. The performance of artificial intelligence models in generating responses to general orthodontic questions: ChatGPT vs Google Bard. Am J Orthod Dentofacial Orthop. 2024; 165(6): 652-62. [Crossref]
10. Zhou M, Pan Y, Zhang Y, Song X, Zhou Y. Evaluating AI-generated patient education materials for spinal surgeries: Comparative analysis of readability and DISCERN quality across ChatGPT and deepseek models. Int J Med Inform. 2025; 198: 105871. [Crossref]
11. Sallam M, Barakat M, Sallam M. A Preliminary Checklist (METRICS) to Standardize the Design and Reporting of Studies on Generative Artificial Intelligence-Based Models in Health Care Education and Practice: Development Study Involving a Literature Review. Interact J Med Res. 2024; 13: e54704. [Crossref]
12. Modig M, Andersson L, Wårdh I. Patients' perception of improvement after orthognathic surgery: pilot study. Br J Oral Maxillofac Surg. 2006; 44(1): 24-7. [Crossref]
13. Lee S, McGrath C, Samman N. Impact of orthognathic surgery on quality of life. J Oral Maxillofac Surg. 2008; 66(6): 1194-9. [Crossref]
14. Yurdakurban E, Topsakal KG, Duran GS. A comparative analysis of AI-based chatbots: Assessing data quality in orthognathic surgery related patient information. J Stomatol Oral Maxillofac Surg. 2024; 125(5): 101757. [Crossref]
15. Etikan I, Musa SA, Alkassim RS. Comparison of convenience sampling and purposive sampling. American Journal of Theoretical and Applied Statistics. 2015; 5(1): 1-4. [Crossref]
16. Chong Q, Marwadi A, Supekar K, Lee Y. Ontology based metadata management in medical domains, Journal of Research and Practice in Information Technology. 2003; 35(2): 139-54.
17. Goodman RS, Patrinely JR, Stone CA, et al. Accuracy and Reliability of Chatbot Responses to Physician Questions. JAMA Netw Open. 2023; 6(10): e2336483. [Crossref]
18. Kincaid JP, Fishburne RP Jr, Rogers RL, Chissom BS. Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel (Report No. ADA006655). 1975. [Crossref]
19. Ateşman E. Türkçede okunabilirliğin ölçülmesi. Dil Dergisi. 1997; 58: 71-4. Available at: https://web.archive.org/web/20201031111301/http://www.atesman.info/wp-content/uploads/2015/10/Atesman-okunabilirlik.pdf
20. Bezirci B, Yılmaz AE. A software library on measuring the readability of texts and a new readability criterion for Turkish. Dokuz Eylul University Faculty of Engineering Journal of Science and Engineering. 2010; 12(3): 49-62.
21. Er A, Ay IE, Horozoğlu Ceran T. Evaluating Turkish Readability and Quality of Strabismus-Related Websites. Cureus. 2024; 16(4): e58603. [Crossref]
22. Hancı V, Ergün B, Gül Ş, Uzun Ö, Erdemir İ, Hancı FB. Assessment of readability, reliability, and quality of ChatGPT®, BARD®, Gemini®, Copilot®, Perplexity® responses on palliative care. Medicine (Baltimore). 2024; 103(33): e39305. [Crossref]
23. Chan L, Xu X, Lv K. DeepSeek-R1 and GPT-4 are comparable in a complex diagnostic challenge: a historical control study. Int J Surg. 2025; 111(6): 4056-9. [Crossref]
24. Qin L, Chen Q, Zhou Y, et al. A survey of multilingual large language models. Patterns (N Y). 2025; 6(1): 101118. [Crossref]
25. Asfuroğlu ZM, Yağar H, Gümüşoğlu E. High accuracy but limited readability of large language model-generated responses to frequently asked questions about Kienböck's disease. BMC Musculoskelet Disord. 2024; 25(1): 879. [Crossref]
26. Hassona Y, Alqaisi D, Al-Haddad A, et al. How good is ChatGPT at answering patients' questions related to early detection of oral (mouth) cancer? Oral Surg Oral Med Oral Pathol Oral Radiol. 2024; 138(2): 269-78. [Crossref]
27. Park SH, Suh CH, Lee JH, Kahn CE, Moy L. Minimum Reporting Items for Clear Evaluation of Accuracy Reports of Large Language Models in Healthcare (MI-CLEAR-LLM). Korean J Radiol. 2024; 25(10): 865-8. [Crossref]
28. Kocak B, Baessler B, Bakas S, et al. CheckList for EvaluAtion of Radiomics research (CLEAR): a step-by-step reporting guideline for authors and reviewers endorsed by ESR and EuSoMII. Insights Imaging. 2023; 14(1): 75. [Crossref]
29. Beheshti M, Toubal IE, Alaboud K, et al. Evaluating the Reliability of ChatGPT for Health-Related Questions: A Systematic Review. Informatics. 2025; 12(1): 9. [Crossref]
30. Huang AS, Hirabayashi K, Barna L, Parikh D, Pasquale LR. Assessment of a Large Language Model's Responses to Questions and Cases About Glaucoma and Retina Management. JAMA Ophthalmol. 2024; 142(4): 371-5. [Crossref]

31. Pandey S, Sharma S. A comparative study of retrieval-based and generative-based chatbots using Deep Learning and Machine Learning. Healthcare Analytics. 2023; 3: 100198. **[Crossref]**

32. Lehman E, Hernandez E, Mahajan D, et al. Do We Still Need Clinical Language Models? arXiv 2302.08091[Preprint]. 2023 Feb 16. **[Crossref]**

33. Ngo M, Jensen E, Meade M. The quality of orthognathic surgery information on social media: A scoping review. Int Orthod. 2025; 23(1): 100959. **[Crossref]**

NORTHWESTERN MEDICAL JOURNAL

Northwestern Med J. 2025;5(4):222-233
DOI: 10.54307/NWMJ.2025.174

RESEARCH ARTICLE

Daedalus fragment from the CBX3 ubiquitous chromatin opening element as an optimum UCOE candidate

Ali Osman Arslan[1], **Ömer Faruk Anakök**[1]

[1]Department of Medical Biology, Faculty of Medicine, Abant Izzet Baysal University, Bolu, Türkiye

Cite as: Arslan AO, Anakök ÖF. Daedalus fragment from the CBX3 ubiquitous chromatin opening element as an optimum UCOE candidate. Northwestern Med J. 2025;5(4):222-233.

ABSTRACT

Aim: The A2UCOE, a well-characterized chromatin-modifying element originating from a constitutively expressed human gene locus, promotes consistent and durable gene activity, even when foreign DNA sequences are inserted into densely packed heterochromatin regions. The aim of this study is to examine the Daedalus fragment in A2UCOE to identify key sub-regions responsible for its prominent chromatin opening activity. This would enable the discovery of a minimal yet fully functional element, allowing for more efficient utilization of vector capacity to accommodate therapeutic transgenes when incorporated into lentiviral vectors (LVs).

Method: To assess the chromatin-opening activity of the Daedalus sub-region within the A2UCOE, lentiviral vectors carrying eGFP reporter constructs were generated and transduced into murine embryonal carcinoma cell lines (P19 and F9). Transgene expression stability was assessed across both pluripotent cells and those differentiated with retinoic acid. HEK293T cells were used for lentiviral packaging. Plasmid propagation was performed in *E. coli* DH5α, and standard molecular cloning techniques were applied. Flow cytometry was used to measure eGFP expression levels, and significance was assessed via Student's t-test ($p < 0.05$).

Results: Flow cytometry and RT-qPCR analyses revealed that the positive control vector (1.5A2UCOE-SEW) maintained stable eGFP expression in both differentiated states of P19 and F9 cells, whereas the UCOE vectors named Daedalus-F and Daedalus-R were unable to prevent transgene silencing. In undifferentiated cultures, eGFP expression from these vectors decreased by 40–50% within two weeks, with a similar decline observed in differentiated cells. The average vector copy numbers remained stable, indicating transcriptional silencing rather than vector loss.

Conclusion: The Daedalus sub-region of A2UCOE alone could not prevent transgene silencing in pluripotent or differentiated cells, indicating additional A2UCOE elements are needed for stable long-term expression.

Keywords: chromatin structure modulation, daedalus vector system, lentiviral gene delivery, neuroectodermal and endodermal differentiation, transcriptional silencing, ubiquitous chromatin opening element

Corresponding author: Ali Osman Arslan **E-mail:** aliosmanarslanist@hotmail.com
Received: 08.01.2025 **Accepted:** 15.04.2025 **Published:** 26.10.2025

INTRODUCTION

Gene therapy involves introducing therapeutic nucleic acids into living cells to correct genetic mutations or to alter cellular functions for medical treatment (1). These nucleic acids can be delivered using either viral or non-viral vectors. While non-viral systems, such as DNA-drug complexes like liposomes, are used, viral vectors remain the more efficient option due to their natural infection mechanisms (2,3). Advances in the field have led to the development of recombinant viral systems that are now widely applied in both research and clinical settings (4,5).

Lentiviral vectors (LVs), derived from HIV-1, are a prominent type of RNA-based viral vectors. Their capacity to effectively deliver genetic material into both proliferating and quiescent cells renders them ideal for targeting a range of tissues, including the brain and stem cells (6). Additionally, LVs have a relatively large carrying capacity (~8 kb) and can be modified to target specific cell types by changing the viral envelope glycoproteins (7). Despite their advantages, recombinant viral vectors pose certain biosafety concerns, such as the risk of insertional mutagenesis or activation of replication-competent viruses, especially in the context of HIV-1. To address this, self-inactivating (SIN) lentiviral vectors have been designed, reducing the likelihood of transcriptional interference and off-target gene activation (8). One major challenge in gene therapy is the epigenetic silencing of integrated vectors, often caused by methylation of CpG islands and chromatin condensation (9,10). Such silencing impedes stable transgene expression, particularly in retroviral vectors (11,12). To mitigate this, researchers have incorporated elements known as ubiquitous chromatin opening elements (UCOEs) into vectors. UCOEs help resist silencing by maintaining an open chromatin structure and, ensuring sustained gene expression (13,14).

A widely studied UCOE is derived from the HNRPA2B1-CBX3 locus, which includes a methylation-resistant CpG island with dual divergent promoters. This element, known as A2UCOE, has been shown to support stable gene expression even in heterochromatic regions (13,15). It functions via two main components: an unmethylated CpG island and bidirectional transcription from the native promoters. A2UCOE's performance in lentiviral contexts, including in hematopoietic stem cells, highlights its utility in gene therapy (16).

Currently, researchers are exploring whether shorter sub-regions of A2UCOE can preserve its anti-silencing capabilities while providing more space for therapeutic genes within lentiviral vectors. The aim was to discover compact and effective UCOE variants for future applications in stable gene therapy.

The primary goal of this study was to enhance the stability of transgene expression while minimizing gene silencing during lentiviral gene therapy. Lentiviral vectors are particularly advantageous because they can efficiently transduce both dividing and non-dividing cells, making them suitable for targeting complex tissues such as the liver and brain. To combat epigenetic silencing, elements like UCOEs—especially A2UCOE—are incorporated into these vectors to preserve an open chromatin configuration. Ongoing research is investigating novel CpG-rich segments and smaller A2UCOE fragments that may improve functionality while optimizing the vector capacity. These advancements have the potential to make gene delivery safer and more effective.

MATERIALS AND METHODS

This study aimed to determine whether a specific sub-region of the A2UCOE, known as the 'Daedalus fragment', retains its chromatin-opening functionality. This was based on emerging questions regarding whether UCOE efficacy depends on intrinsic promoter features. The exact genomic locations of the tested sub-fragments are illustrated in Figures 1 and 2.

To assess the activity of the vector constructs, embryonic carcinoma cell lines originating from mice, namely P19 and F9, were used. These cell lines are well-established models for evaluating the expression stability of viral vectors (15,16). To evaluate the reliability and strength of expression mediated by the various vector designs, eGFP production was tracked in cells maintained in both their pluripotent state and following differentiation.

Luria-bertani (LB) medium preparation

To prepare Luria-Bertani (LB) medium, 10 g of tryptone, 5 g of yeast extract, and 10 g of NaCl were dissolved in one liter of deionized water. The solution was autoclaved at 121°C (15 psi) for 20 minutes to ensure sterility. Once cooled to approximately 55°C, ampicillin was incorporated to a final concentration of 50 µg/mL. The medium was subsequently stored at 4°C until needed. For solid media, 20 g of agar was added per liter of LB prior to sterilization. Selective LB broth or agar was prepared by supplementing the medium with ampicillin at a concentration of 100 µg/mL.

Competent E. coli DH5α Transformation

Chemically competent DH5α E. coli cells, obtained from Life Technologies, were transformed following the standard procedure provided by the supplier. Plasmid-containing colonies were cultured on ampicillin plates (100 µg/mL) overnight at 37°C and then grown in LB medium with ampicillin (100 µg/mL) in small (5 mL) or large (200–500 mL) cultures. Plasmids were extracted using Qiagen kits and eluted in TE buffer.

Cell culture

HEK293T cells were cultured in DMEM supplemented with 10% fetal bovine serum, 1% L-glutamine (200 mM), and penicillin-streptomycin (10 µg/mL each) at 37°C in a humidified incubator containing 5% CO_2. Cell cultures were initiated by plating 2×10^7 cells into T162 flasks and maintained until they reached approximately 80–90% confluence prior to transfection. The culture supernatant was collected 48 hours after transfection, the medium was refreshed with DMEM, and a second harvest was performed at the 72-hour mark. Culture medium was collected 48 hours post-transfection, replaced with fresh DMEM, and harvested again at 72 hours. For viral titration, $1–2 \times 10^5$ cells were seeded in the wells of a 24-well culture plate, followed by exposure to gradually diluted viral suspensions, resulting in a multiplicity of infection (MOI) ranging from 1 to 10^{-5} per µL. Detached cells were treated with PBS and Tryple Red reagent, then neutralized with serum-based medium. For flow cytometry, cells were fixed with 4% formaldehyde-PBS, shaken, and kept in darkness at 4°C until they were ready for analysis.

Mouse embryonic carcinoma cells culture

P19 cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) enriched with 2 millimolar L-glutamine and supplemented with 1% non-essential amino acids. The culture conditions included the addition of 10 µg/mL penicillin-streptomycin, and incubation at 37°C within a 5% CO_2 atmosphere. Neuronal differentiation was initiated by forming embryoid bodies in non-adherent dishes containing DMEM supplemented with 5% FBS and retinoic acid at a concentration of 1 micromolar. The culture of F9 cells was carried out in flasks coated with gelatin, employing DMEM enriched with 10% FBS, alongside antibiotics penicillin (100 U/mL) and streptomycin (100 µg/mL), plus 2 mM L-glutamine. The formation of embryoid bodies in DMEM/F12 supplemented with 5% FBS and 50 nM RA triggered the cells to differentiate toward a parietal endoderm lineage.

Reporter gene analysis

The UCOE and A2UCOE eGFP reporter constructs within the LV system were assessed by transducing P19 and F9 cells. The stability of gene expression was then monitored both prior to and following differentiation into neuroectodermal and endodermal lineages.

At least 2×10^5 fixed cells were analyzed for GFP expression using flow cytometry (BD FACSymphony™). During analysis, only viable cells were included by gating based on cell size and granularity, determined respectively through forward scatter and side scatter measurements. GFP+ cells were identified by detecting fluorescence emission at 525 nm (FL1 channel) compared to 575 nm (FL2 channel). Untransduced cells served as negative controls to set the gating threshold for positive GFP signals. The resulting percentage of GFP-positive cells represents the fraction of viable cells within the sample that expressed GFP. Subsequently, the proportion of GFP-positive cells for each sample was analyzed in relation to the amount of virus administered. By scaling this relationship, the quantity of cells infected with a specified volume of viral preparation could be estimated, allowing for the calculation of the multiplicity of infection (MOI) based on mean fluorescence intensity. For this calculation, values ranging from 1% to 20% were selected. Samples

exhibiting more than 20% GFP-positive cells likely contained multiple integration events per cell, which could lead to an underestimation of the viral titer. Conversely, samples with less than 1% GFP expression were considered unreliable, as detectable signals at such low levels are prone to interference from background noise.

Explanation of the chemicals and devices used in the study: Thermo Fisher Scientific, Gibco, Qiagen, Sigmaaldrich (MERK), Bio-Rad, BD FACSymphony

Statistical analyses

Statistical evaluations were conducted using Prism version 7 software. To assess the eGFP-positive populations from flow cytometry, as well as mean fluorescence intensity (MFI) and vector copy number (VCN) measurements in transduced cell cultures, Student's t-test was employed. Results with p-values below 0.05 were considered statistically significant.

In our cell culture study, the cell line used was obtained commercially. Therefore, ethical approval was not required for our study.

Lentiviral vectors used in this study

The experimental approach aimed at testing the UCOE two-component model involved the use of 750bp length Daedalus fragments. To test their effectiveness in preventing suppression of gene activity, these sequences were inserted in both forward and reverse directions into the silencing-prone SFFV-eGFP vector. Approximate locations and plasmid schematics for the Daedalus fragments derived from A2UCOE sub-regions examined in this study are presented in Figures 1 and 2.

RESULTS

Evaluation of the functional properties of Daedalus-F and Daedalus-R lentiviral vectors containing candidate UCOEs in undifferentiated P19 and F9 cell lines

The generated lentiviral vectors (LVs) were utilized to transduce P19 and F9 cells at infection ratios (MOI) of 3 and 6, aiming for an initial eGFP+ cell proportion between 40 and 60% across all pools. Cells

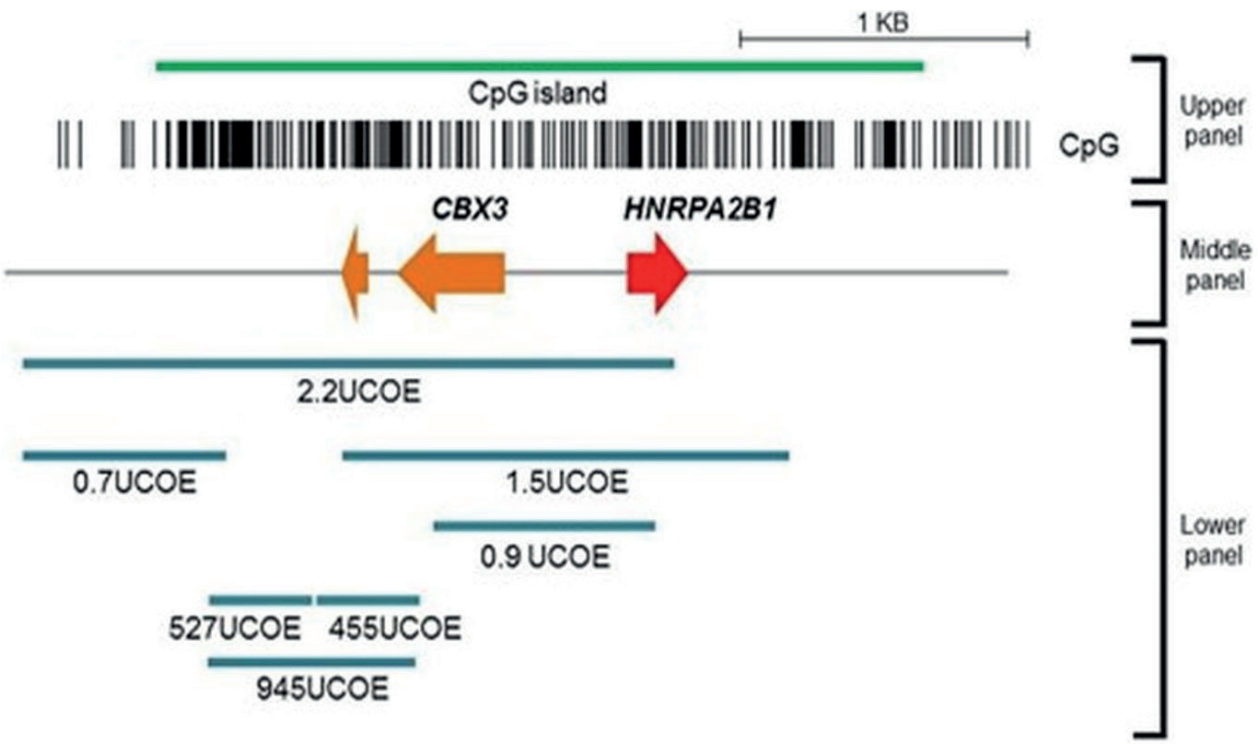


Figure 1. The positions of A2UCOE region sub-fragments analyzed for UCOE activity.

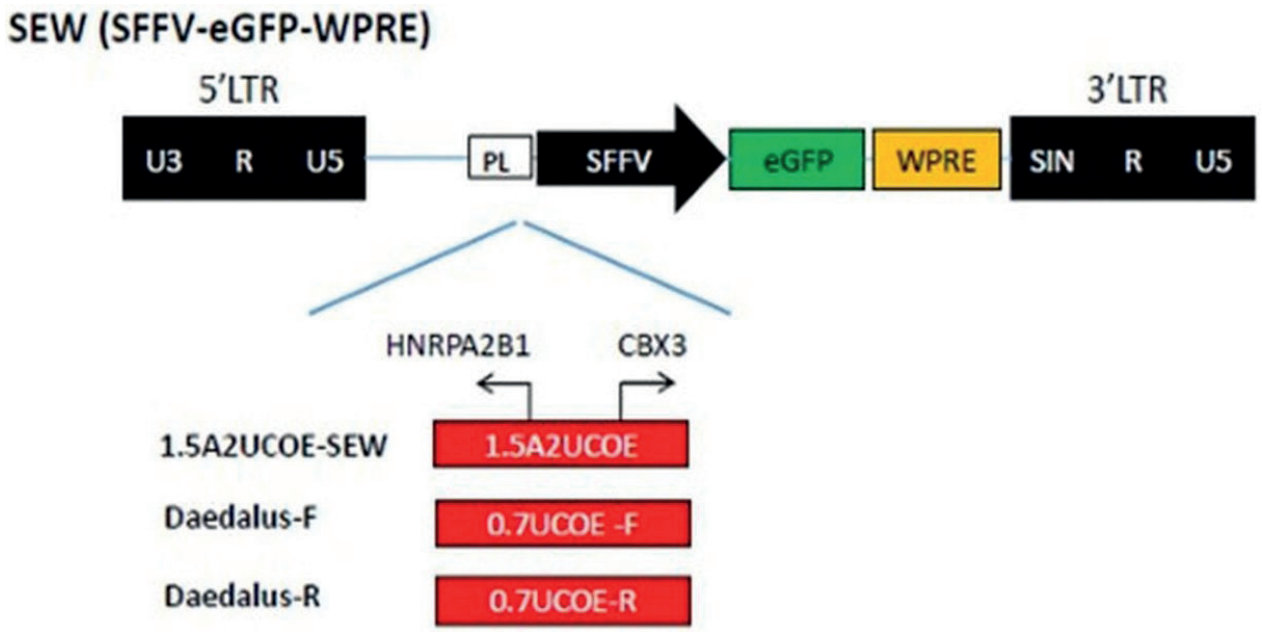


Figure 2. Illustration of novel candidate UCOE and control lentiviral vectors.

that underwent transduction, along with negative controls, were maintained in culture and periodically assessed using flow cytometric analysis to quantify the percentage of cells expressing eGFP, as well as their mean fluorescence intensity during ongoing culture following a single vector transduction. The first three days after transduction, cell samples were collected once a week up to 45 days (see Figure 3) to analyze eGFP expression via flow cytometry. At designated time points, DNA was isolated for RT-qPCR to quantify the average number of vector copies present in each cell. Flow cytometry analysis results in Figure 4A illustrate the proportion of eGFP+ cells. At the start, all vectors demonstrated comparable transduction efficiencies, ranging between 45% and 60% eGFP+ cells; however, the expression driven by SEW (SFFV-eGFPWPRE) sharply decreased from 46% to just 3% of positive cells over a period of 17 days. Conversely, the percentage of eGFP+ cells within the 1.5A2UCOE-SEW group showed consistent stability throughout the 45-day cultivation period. Therefore, while the proportion of eGFP+ cells was notably greater in 1.5A2UCOE-SEW transduced samples compared to alternative vectors, the reduction in expression observed in both Daedalus-F and Daedalus-R mirrored the pattern seen with SEW. SEW transduced cells saw an almost 80% reduction in eGFP expression after two weeks. Except for 1.5A2UCOE-SEW, the remaining vectors, including both Daedalus-F and Daedalus-R, demonstrated a 40–50% decline in eGFP+ cell populations within the corresponding time interval. The MFI across all vectors reflected the percentage of eGFP+ cells (Figure 4B). MFI remained stable in 1.5A2UCOE-SEW transduced cells but fluctuated in SEW and the new Daedalus-F and Daedalus-R constructs.

Figure 5A presents the flow cytometry analysis showing the proportion of eGFP-expressing cells measured at successive time points following transduction in undifferentiated F9 cells. These findings are consistent with those observed in P19 cells (Figure 4A). Initial transduction efficiencies were comparable across all vectors, ranging from 39% to 50% eGFP+ cells. However, expression driven by the SEW vector declined sharply, dropping from 47% to just 7% within 17 days. Conversely, cells genetically transduced with the 1.5A2UCOE-SEW vector sustained a stable eGFP-positive population of approximately 45% throughout the 38-day culture period. The candidate UCOE constructs, Daedalus-F and Daedalus-R, showed a reduction in gene activity comparable to the decrease observed with the SEW vector. Specifically, eGFP

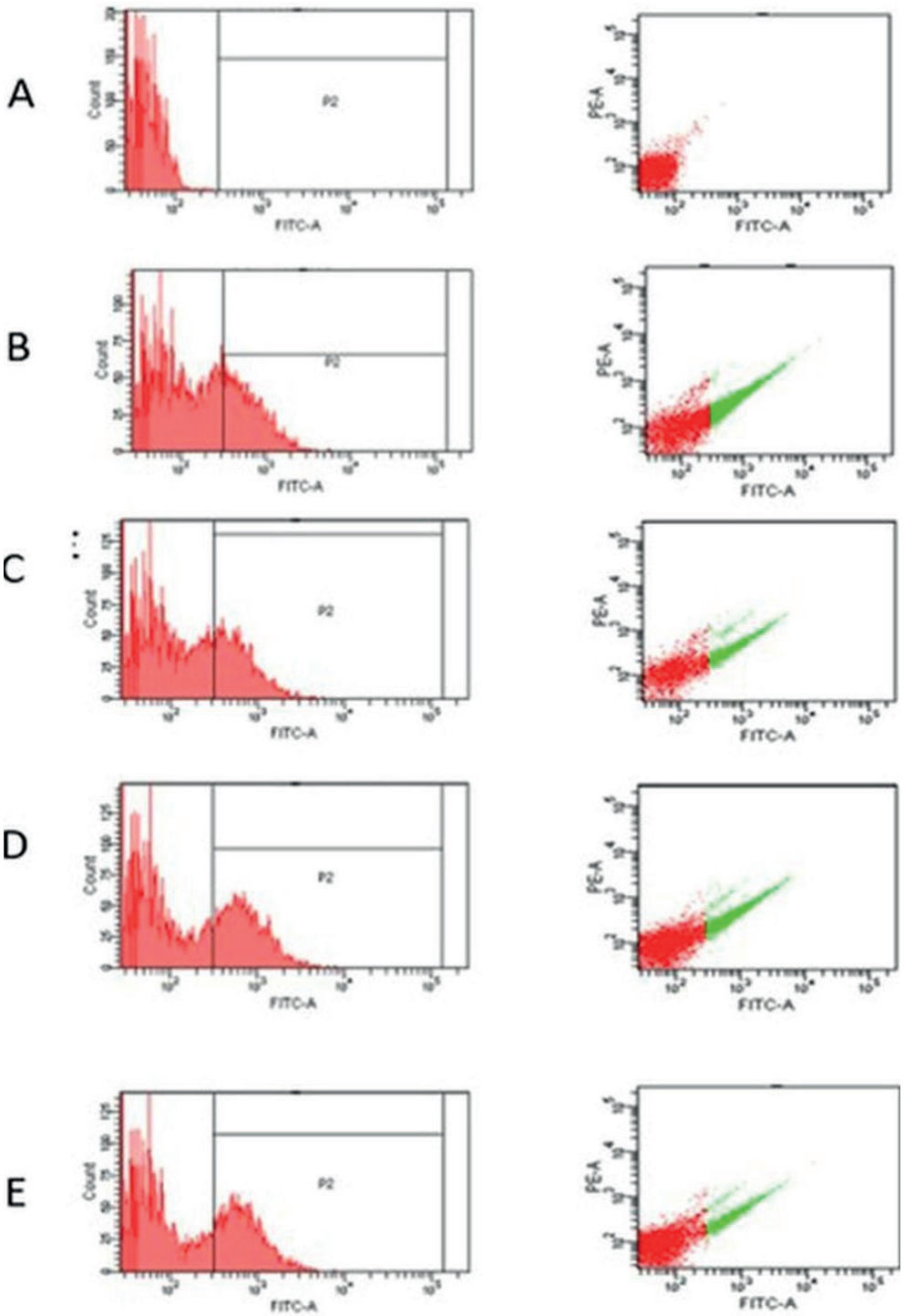


Figure 3. Flow cytometry plots of untransduced and transduced P19 and F9 cells prior to differentiation.*

* Flow cytometry analysis was performed 3 days post-transduction, showing GFP-negative cells in red and GFP-positive cells in green. (A) Negative control—untransduced cells. (B) SEW transduced cells. (C) 1.5A2UCOE-SEW transduced cells. (D) Daedalus-F transduced cells. (E) Daedalus-R transduced cells.

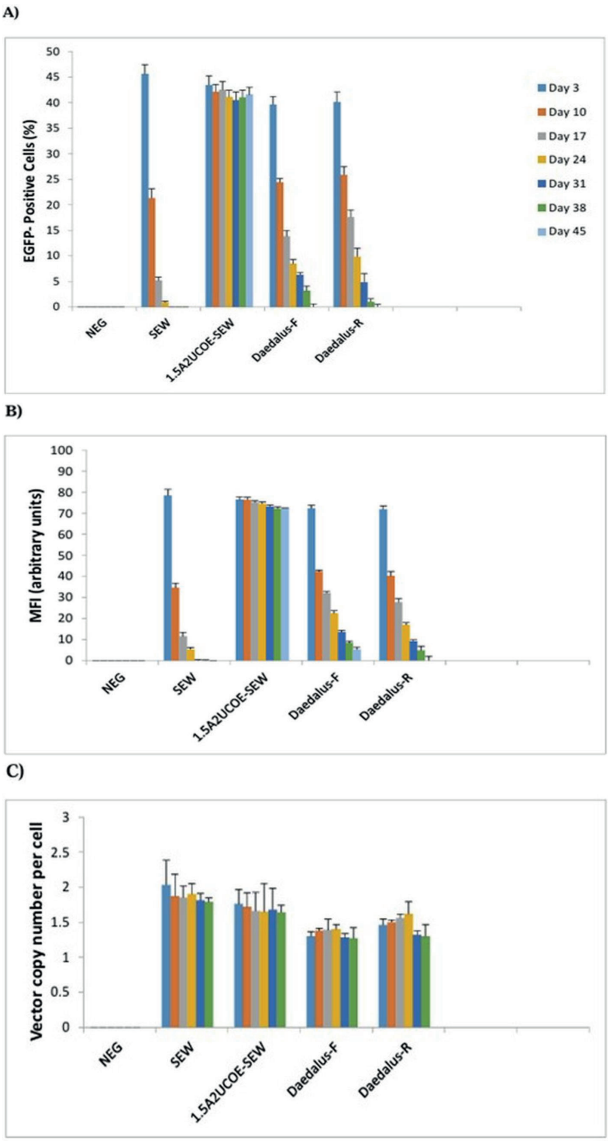


Figure 4. Novel candidate Daedalus-F and Daedalus-R UCOEs offer only partial protection against silencing in undifferentiated P19 cells. *

* Cells were analyzed by flow cytometry to detect the percentage of eGFP reporter gene expressing (eGFP+) cells, mean fluorescence intensity (MFI) and by RT-qPCR for average VCN per cell. The data show combined results from three independent transductions for each vector, plus negative control (NEG), over a period of 3 to 45 days post transduction. (A) Time course of percentage eGFP positive cells; (Mean + SEM, n=4; **p<0.01). (B) As in (A) but showing MFI; (Mean + SEM, n=4; **p<0.01). (C) As in (A)/(B) but average VCN/cell; (Mean + SEM, n=4; **p<0.01).

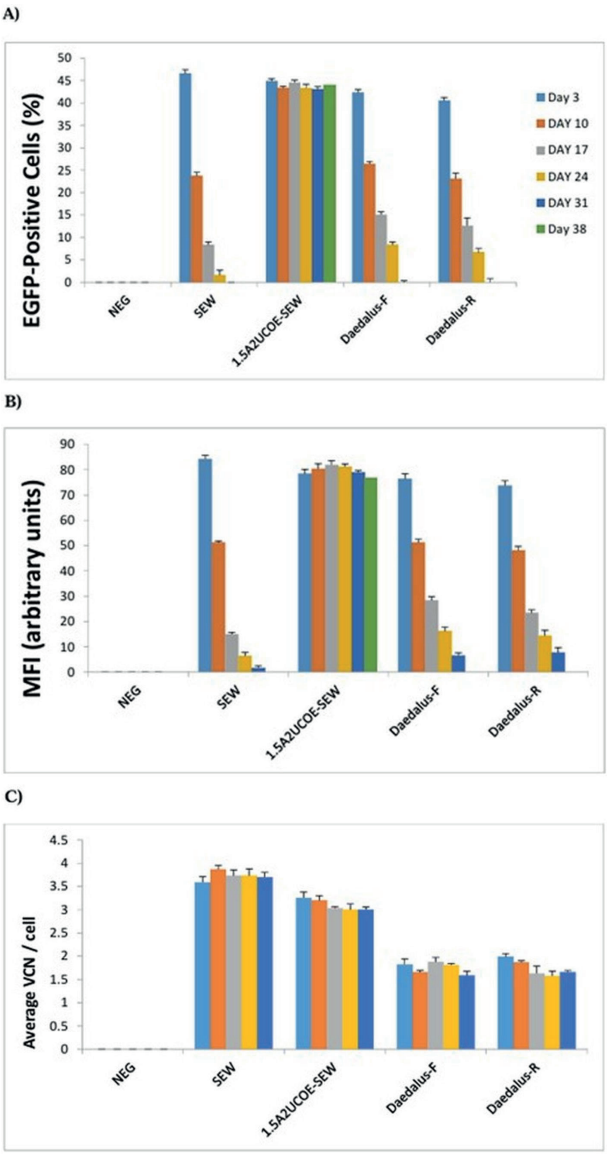


Figure 5. Novel candidate UCOE Daedalus-F and Daedalus-R offer only partial protection against silencing in undifferentiated F9 cells. *

* Cells were analyzed by flow cytometry to detect the percentage of eGFP reporter gene expressing (eGFP+) cells and MFI, and to quantify the average VCN per cell by real-time qPCR. The data show combined results from three independent transductions for each vector, plus negative control (NEG), over a period of 3 to 38 days post-transduction. (A) Time course of percentage eGFP positive cells; (Mean + SEM, n=4; **p<0.01). (B) As in (A) but showing MFI; (Mean + SEM, n=4; **p<0.01). (C) As in (A)/(B) but average VCN/cell; (Mean + SEM, n=4; **p<0.01).

levels in cells treated with SEW dropped by nearly 80% fourteen days after gene delivery. Apart from the 1.5A2UCOE-SEW group, the reduction in eGFP+ cells for Daedalus-F and Daedalus-R vectors ranged between 40% and 50% over the same timeframe. The MFI measurements across all vectors reflected the trends seen in eGFP expression (Figure 5B), showing stability in cells transduced with 1.5A2UCOE-SEW, instability in those with SEW alone, and partial stability in cells treated with the new UCOE vectors. Throughout the entire culture duration, the mean number of vector copies per individual cell in the P19 (Figure 4C) and F9 (Figure 5C) transduction experiments remained stable, suggesting that the reduction in eGFP-positive cells observed with the SEW, Daedalus-F, and Daedalus-R vectors was due to transcriptional silencing of the integrated sequences rather than a loss of vector presence.

Assessment of the functional performance of Daedalus-F and Daedalus-R candidate UCOEs in P19 and F9 cells following differentiation

The newly developed UCOE vectors, Daedalus-F and Daedalus-R, failed to sustain consistent gene expression in undifferentiated P19 (Figure 4) and F9 (Figure 5) cell lines. Subsequently, we evaluated their effectiveness in maintaining expression stability throughout the differentiation process of P19 and F9 cells into neuroectodermal and endodermal lineages, respectively. Cells transduced with the respective lentiviral vectors were cultured in differentiation medium supplemented with retinoic acid, on non-adherent plastic surfaces to promote the formation of embryoid bodies, a critical step for their subsequent differentiation into neuronal and endodermal lineages. The cells differentiated into a mixture of neuroectodermal and parietal endoderm cell types, with fibroblast-like cells being the most abundant during the early stages—2 to 3 days for P19 cells and 5 to 6 days for F9 cells. Neurons and endoderm cells were also detected in the cultures. To enrich the population for non-dividing cells, 10 μM cytosine arabinoside was added, which selectively eliminated dividing cells and allowed only neurons and endoderm cells to survive. The optimal duration for embryoid body formation and differentiation into neuroectodermal and endodermal lineages was 2 to 3 days for P19 cells and 5 to 6 days for F9 cells. Following the differentiation process, Figures 6 and 7 display the flow cytometric profiles of entire cell populations derived from triplicate cultures transduced with the SEW, 1.5A2UCOE-SEW, Daedalus-F, and Daedalus-R lentiviral constructs. The data indicate that the lentiviral vector incorporating the 1.5 kb A2UCOE sequence (1.5A2UCOE-SEW), used as a positive control, maintained consistent transgene expression following the differentiation of P19 and F9 cells into neuronal and endodermal lineages, respectively (Figures 6A and 7A). In contrast, the newly developed UCOE vectors, Daedalus-F and Daedalus-R, exhibited a rapid decline in eGFP expression. The suppression pattern mirrored that of the SEW vector used as a silencing control, wherein eGFP expression was rapidly diminished, as illustrated in Figures 6A and 7A. As seen in undifferentiated cells (Figures 4B and 5B), the MFI values measured during the differentiation process of both P19 and F9 cells displayed a pattern consistent with the eGFP+ cell data (Figures 6B and 7B). The mean VCN remained consistent over the entire experimental timeline (see Figures 6C and 7C). Comparable results were observed in both pluripotent and differentiated forms of F9 and P19 cells following transduction with the lentiviral constructs employed in this study. It is noteworthy that differentiating cells exhibited a greater loss of eGFP expression after being transduced with lentiviral particles carrying SEW and newly characterized UCOE elements, as seen in Figures 6A and 7A. This decline was more pronounced compared to undifferentiated cells, illustrated in Figures 4A and 5A. Additionally, because Daedalus-F and Daedalus-R vectors failed to sustain stable expression in differentiated P19 and F9 cells, immunofluorescence analysis for neuroectodermal and endodermal markers was not performed.

Given that neither of the Daedalus constructs exhibited stable expression, we decided against performing immunofluorescence staining on the differentiated cells.

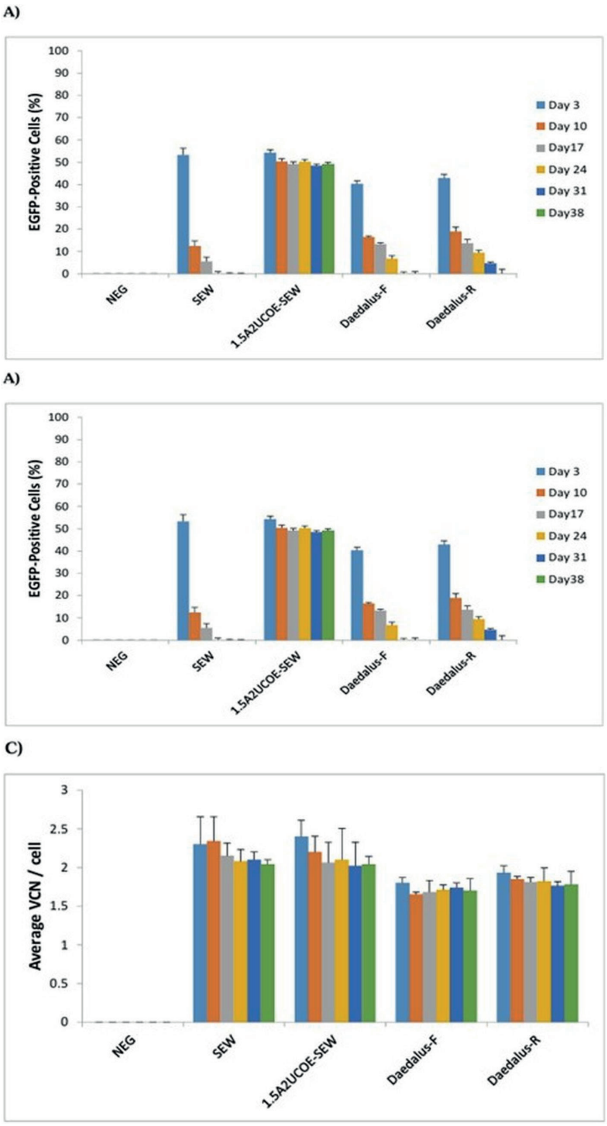


Figure 6. Novel candidate UCOEs Daedalus-F and Daedalus-R offer only partial protection against silencing in differentiated P19 cells. *

* Cells were analyzed by flow cytometry to detect the percentage of eGFP reporter gene expressing (eGFP+) cells and MFI, and to quantify the average VCN per cell by real-time qPCR. The data show combined results from three independent transductions for each vector, plus negative control (NEG), over a period of 3 to 38 days post-transduction. (A) Time course of percentage eGFP positive cells; (Mean + SEM, n=4; **p<0.01). (B) As in (A) but showing MFI; (Mean + SEM, n=4; **p<0.01). (C) As in (A)/(B) but average VCN/cell; (Mean + SEM, n=4; **p<0.01).

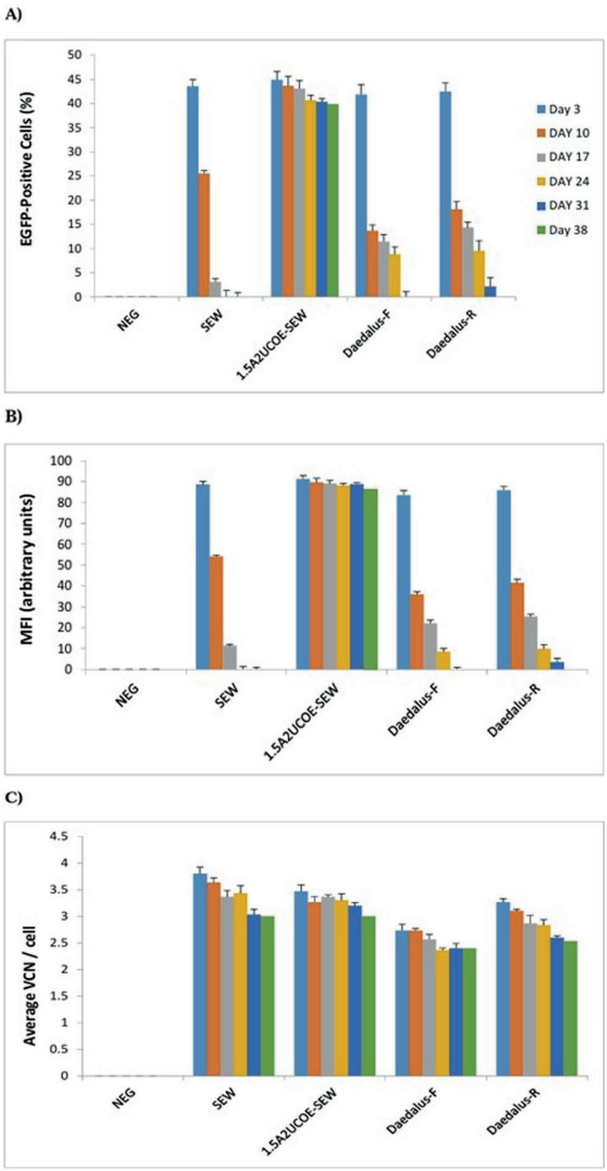


Figure 7. Novel candidate UCOEs Daedalus-F and Daedalus-R offer only partial protection against silencing in differentiated F9 cells. *

* Cells were analyzed by flow cytometry to detect the percentage of eGFP reporter gene expressing (eGFP+) cells and MFI, and to quantify the average VCN per cell by real-time qPCR. The data show combined results from three independent transductions for each vector, plus negative control (NEG), over a period of 3 to 38 days post-transduction. (A) Time course of percentage eGFP-positive cells; (Mean + SEM, n=4; **p<0.01). (B) As in (A) but showing MFI; (Mean + SEM, n=4; **p<0.01). (C) As in (A)/(B) but average VCN/cell; (Mean + SEM, n=4; **p<0.01).

DISCUSSION

The gammaretroviral and lentiviral vector classes, known for their integration capability, continue to be the most reliable tools for ensuring stable incorporation and expression of therapeutic genes. This is particularly evident in the context of stem cell subsets undergoing active proliferation. Over the course of the last fifteen years, clinical trials employing these vector systems in ex vivo applications targeting hematopoietic stem cells (HSCs) have yielded promising results. Genetic modification of patients' hematopoietic stem cells (HSCs) using gammaretroviral vectors has shown therapeutic success in various primary immunodeficiencies, including severe combined immunodeficiency subtypes such as SCID-X1 and SCID-ADA (17,18), as well as in disorders like Wiskott-Aldrich syndrome and chronic granulomatous disease (19,20). More recently, lentiviral vectors have demonstrated success in treating inherited demyelinating disorders, including X-ALD (21) and MLD (22), in addition to WAS (23). Despite these advancements, two critical issues continue to arise in the context of using integrating viral vectors and require ongoing attention. A major limitation is the potential for insertion-related genomic alterations; another is the suppression of therapeutic gene activity over time through epigenetic mechanisms (1). Insertional mutagenesis resulting from gammaretroviral integration, which inadvertently triggered the activation of proto-oncogenes within the host genome and contributed to malignant transformation, was detected in five patients out of a cohort of twenty undergoing treatment for SCID-X1 (17,18). Meanwhile, epigenetic silencing triggered by promoter methylation resulted in a loss of sustained therapeutic benefit in two CGD patients, despite an initially positive response (19,20).

The main goal of this study phase was to evaluate whether particular sub-regions within the A2UCOE, which originates from the human HNRPA2B1-CBX3 locus (see Figures 1 and 2), retain the characteristic UCOE activity. Earlier investigations have indicated that a 0.7 kb fragment within the first intron of CBX3 preserves UCOE functionality when fused to an exogenous CMV promoter (24), and that this segment can sustain appropriate DNA methylation patterns in the context of transgene expression (25). These results prompt reconsideration of the mechanism by which UCOEs operate through bidirectional divergent transcription (26). To directly evaluate whether a 0.7 kb segment derived from the first intron of CBX3 (see Figure 2) possesses UCOE functionality, lentiviral constructs named Daedalus-F and Daedalus-R (Figures 1 and 2) were engineered. Our findings from experiments involving P19 and F9 cells, both before and after their differentiation into neural ectoderm and endoderm cell types, distinctly demonstrates that this sequence fails to promote expression stability when associated with an SFFV promoter (Figures 4-7). The difference observed between our findings and earlier studies likely stems from the fact that previous investigations reporting UCOE activity of the 0.7 kb intronic region within CBX3 were conducted using transductions that yielded approximately 8 to 10 lentiviral vector copies per cell (24). Under these conditions, even a single active copy of the integrated vector within a cell can generate detectable GFP reporter gene expression. This may misleadingly suggest that the transgene remains functionally stable. However, a notable reduction of approximately 55% in MFI was observed at four weeks after transduction with the 0.7UCOE-GFP ("Daedalus") lentiviral vector. This decline highlights considerable vector silencing, despite the continued presence of GFP+ cells (see Figure 4) (24). Our findings provide strong evidence that the expression of the 0.7UCOE (Daedalus) vector undergoes silencing. This is supported by experiments specifically designed to generate cells containing a low average vector copy number per cell, between 1 and 4 (Figures 4C, 7C). Furthermore, the P19 and F9 embryonal carcinoma cell lines used in our study may serve as a more sensitive and reliable model for evaluating functional gene silencing compared to the CHO cells employed in previous research (24,15,16).

CONCLUSION

In summary, our findings demonstrate that the 0.7 kb Daedalus sub-region of A2UCOE is insufficient to maintain stable transgene expression in both pluripotent and differentiated P19 and F9 cells. Despite stable vector integration, significant transcriptional silencing occurred, particularly at low vector copy numbers. These results indicate that additional regulatory sequences within A2UCOE are necessary

to sustain long-term and robust transgene expression. Identifying a minimal yet fully functional UCOE element remains crucial for optimizing lentiviral vector design and enhancing the efficiency of therapeutic transgene delivery.

Ethical approval

In this study, ethical approval is not required.

Author contribution

Surgical and Medical Practices: ÖFA; Concept: ÖFA; Design: ÖMA, AOA; Data Collection or Processing: ÖFA; Analysis or Interpretation: AOA; Literature Search: ÖFA, AOA; Writing: ÖFA, AOA. All authors reviewed the results and approved the final version of the article.

Source of funding

The authors declare the study received no funding.

Conflict of interest

The authors declare that there is no conflict of interest.

REFERENCES

1. Antoniou MN, Skipper KA, Anakok O. Optimizing retroviral gene expression for effective therapies. Hum Gene Ther. 2013; 24(4): 363-74. [Crossref]
2. Griesenbach U, Alton EW. Moving forward: cystic fibrosis gene therapy. Hum Mol Genet. 2013; 22(R1): R52-8. [Crossref]
3. Neville JJ, Orlando J, Mann K, McCloskey B, Antoniou MN. Ubiquitous Chromatin-opening Elements (UCOEs): applications in biomanufacturing and gene therapy. Biotechnol Adv. 2017; 35(5): 557-64. [Crossref]
4. Sizer RE, White RJ. Use of ubiquitous chromatin opening elements (UCOE) as tools to maintain transgene expression in biotechnology. Comput Struct Biotechnol J. 2022; 21: 275-83. [Crossref]
5. Fu Y, Han Z, Cheng W, Niu S, Wang T, Wang X. Improvement strategies for transient gene expression in mammalian cells. Appl Microbiol Biotechnol. 2024; 108(1): 480. [Crossref]
6. Naldini L, Blömer U, Gallay P, et al. In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. Science. 1996; 272(5259): 263-7. [Crossref]
7. Bouard D, Alazard-Dany D, Cosset FL. Viral vectors: from virology to transgene expression. Br J Pharmacol. 2009; 157(2): 153-65. [Crossref]
8. Gurumoorthy N, Nordin F, Tye GJ, Wan Kamarul Zaman WS, Ng MH. Non-integrating lentiviral vectors in clinical applications: a glance through. Biomedicines. 2022; 10(1): 107. [Crossref]
9. Teich NM, Weiss RA, Martin GR, Lowy DR. Virus infection of murine teratocarcinoma stem cell lines. Cell. 1977; 12(4): 973-82. [Crossref]
10. Speers WC, Gautsch JW, Dixon FJ. Silent infection of murine embryonal carcinoma cells by Moloney murine leukemia virus. Virology. 1980; 105(1): 241-4. [Crossref]
11. Razin A. CpG methylation, chromatin structure and gene silencing-a three-way connection. EMBO J. 1998; 17(17): 4905-8. [Crossref]
12. Fuks F. DNA methylation and histone modifications: teaming up to silence genes. Curr Opin Genet Dev. 2005; 15(5): 490-5. [Crossref]
13. Benton T, Chen T, McEntee M, et al. The use of UCOE vectors in combination with a preadapted serum free, suspension cell line allows for rapid production of large quantities of protein. Cytotechnology. 2002; 38(1-3): 43-6. [Crossref]
14. Ellis J. Silencing and variegation of gammaretrovirus and lentivirus vectors. Hum Gene Ther. 2005; 16(11): 1241-6. [Crossref]
15. Zhang F, Thornhill SI, Howe SJ, et al. Lentiviral vectors containing an enhancer-less ubiquitously acting chromatin opening element (UCOE) provide highly reproducible and stable transgene expression in hematopoietic cells. Blood. 2007; 110(5): 1448-57. [Crossref]
16. Zhang F, Frost AR, Blundell MP, Bales O, Antoniou MN, Thrasher AJ. A ubiquitous chromatin opening element (UCOE) confers resistance to DNA methylation-mediated silencing of lentiviral vectors. Mol Ther. 2010; 18(9): 1640-9. [Crossref]
17. Cavazzana-Calvo M, Fischer A, Hacein-Bey-Abina S, Aiuti A. Gene therapy for primary immunodeficiencies: part 1. Curr Opin Immunol. 2012; 24(5): 580-4. [Crossref]
18. Cavazza A, Moiani A, Mavilio F. Mechanisms of retroviral integration and mutagenesis. Hum Gene Ther. 2013; 24(2): 119-31. [Crossref]
19. Ott MG, Seger R, Stein S, Siler U, Hoelzer D, Grez M. Advances in the treatment of chronic granulomatous disease by gene therapy. Curr Gene Ther. 2007; 7(3): 155-61. [Crossref]
20. Aiuti A, Bacchetta R, Seger R, Villa A, Cavazzana-Calvo M. Gene therapy for primary immunodeficiencies: part 2. Curr Opin Immunol. 2012; 24(5): 585-91. [Crossref]
21. Cartier N, Hacein-Bey-Abina S, Bartholomae CC, et al. Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy. Science. 2009; 326(5954): 818-23. [Crossref]
22. Biffi A, Montini E, Lorioli L, et al. Lentiviral hematopoietic stem cell gene therapy benefits metachromatic leukodystrophy. Science. 2013; 341(6148): 1233158. [Crossref]

23. Bosticardo M, Ferrua F, Cavazzana M, Aiuti A. Gene therapy for Wiskott-Aldrich Syndrome. Curr Gene Ther. 2014; 14(6): 413-21. **[Crossref]**

24. Bandaranayake AD, Correnti C, Ryu BY, Brault M, Strong RK, Rawlings DJ. Daedalus: a robust, turnkey platform for rapid production of decigram quantities of active recombinant proteins in human cell lines using novel lentiviral vectors. Nucleic Acids Res. 2011; 39(21): e143. **[Crossref]**

25. Lienert F, Wirbelauer C, Som I, Dean A, Mohn F, Schübeler D. Identification of genetic elements that autonomously determine DNA methylation states. Nat Genet. 2011; 43(11): 1091-7. **[Crossref]**

26. Lindahl Allen M, Antoniou M. Correlation of DNA methylation with histone modifications across the HNRPA2B1-CBX3 ubiquitously-acting chromatin open element (UCOE). Epigenetics. 2007; 2(4): 227-36. **[Crossref]**

NORTHWESTERN MEDICAL JOURNAL

Northwestern Med J. 2025;5(4):234-239
DOI: 10.54307/NWMJ.2025.185

RESEARCH ARTICLE

Self-reported COVID-19 prevalence among isotretinoin users vs. non-users: a cross-sectional survey

Rukiye Yasak Güner[1], Mehmet Kayhan[2], Mustafa Tosun[1], Ahmet Turan Ünsal[1], Ahmet Turan İnce[3], Volkan Erguvanlı[4]

[1]Department of Dermatology, Faculty of Medicine, Sivas Cumhuriyet University, Sivas, Türkiye
[2]Department of Internal Medical Sciences, Faculty of Medicine, Bolu Abant İzzet Baysal University, Bolu, Türkiye
[3]Department of Cardiovascular Surgery, Faculty of Medicine, Sivas Cumhuriyet University, Sivas, Türkiye
[4]Department of Medicine, Sivas Cumhuriyet University, Sivas, Türkiye

Cite as: Yasak Güner R, Kayhan M, Tosun M, Ünsal AT, İnce AT, Erguvanlı V. Self-reported COVID-19 prevalence among isotretinoin users vs. non-users: a cross-sectional survey. Northwestern Med J. 2025;5(4):234-239.

ABSTRACT

Aim: Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused a global pandemic, and effective drug treatments for COVID-19 remain unavailable. Isotretinoin (ISO), a retinoic acid derivative commonly used for severe acne, is a potent down-regulator of ACE2 receptors, which play a key role in the pathophysiology of COVID-19. This study aims to investigate the effect of ISO treatment on COVID-19 infection in patients with acne vulgaris.

Materials and Methods: Between January 2021 and April 2021, 520 ISO users with acne vulgaris and 400 controls were included in the study. Data were collected using an electronic questionnaire distributed via social media, e-mail and mobile phones.

Results: Among ISO users, 66 (12.7%) had COVID-19 compared to 89 (22.3%) in the control group, indicating a significantly lower infection rate in the ISO group ($p < 0.001$). The risk of COVID-19 infection was 1.76 times higher in the control group. COVID-19 rates decreased steadily during the first three months of ISO treatment, particularly at the 20 mg/day dose. However, loss of smell and taste was more common in the ISO group than in controls.

Conclusion: In this cross-sectional survey, isotretinoin use was associated with lower self-reported COVID-19 prevalence.

Keywords: ACE2 receptor, acne, COVID-19, isotretinoin, off-label use

Corresponding author: Rukiye Yasak Güner **E-mail:** rky.guner@gmail.com
Received: 14.03.2025 **Accepted:** 02.09.2025 **Published:** 26.10.2025

INTRODUCTION

The high mortality rate, global impact, and lack of effective treatments have made COVID-19 a major public health concern. Although the pandemic has moved into an endemic phase due to vaccines and viral mutations, as of August 2025, approximately 778,474,21 cases and 7 million deaths have been reported worldwide since the onset of the pandemic (1). Despite these advancements, an effective drug treatment remains unavailable.

Isotretinoin (13-cis-retinoic acid) (ISO), approved by the FDA in 1982 for the treatment of severe acne, has also been used off-label for various cancers, including cutaneous T-cell lymphoma, neuroblastoma, and squamous cell carcinoma prevention (2,3). Similarly, drugs approved for other indications are undergoing clinical trials for COVID-19 treatment.

Isotretinoin has demonstrated significant suppression of ACE2 receptors, key players in COVID-19 pathophysiology, and may inhibit the papain-like proteinase enzyme encoded by SARS-CoV-2 genes (4). Furthermore, as a metabolite of retinoic acid, isotretinoin has been reported to prevent disseminated intravascular coagulation (DIC), a severe COVID-19 complication, and to reduce lymphocyte counts (5). While these effects suggest potential benefits, isotretinoin's anti-inflammatory and immunomodulatory properties may position it as a candidate for reducing SARS-CoV-2-related mortality and morbidity.

This study explores whether isotretinoin use is associated with differences in self-reported COVID-19 infection rates in a real-world sample. Given isotretinoin's modulatory effect on ACE2 receptor, the analysis is hypothesis-generating and not designed to test a treatment effect.

MATERIAL AND METHOD

Although data were collected through a questionnaire, participants reported their history of COVID-19 infection rather than current disease status. For this reason, the study design was considered retrospective in nature, based on self-reported past medical events rather than real-time follow-up.

The questionnaire was carefully developed in line with the study objectives and distributed to a diverse group of participants. Due to pandemic restrictions, face-to-face interviews were not possible; therefore, data collection was conducted online between January and April 2021. A total of 520 isotretinoin users with acne vulgaris and 400 controls were included in the study.

The 29-item questionnaire covered demographic characteristics, comorbidities, the amount and duration of isotretinoin use, COVID-19 infection status, disease severity, hospitalization history, contact history, and COVID-19 cases within the family. Individuals with chronic diseases were excluded from the study.

Ethical approval was obtained from the Local Ethics Committee and the Ministry of Health (2021-01/41).

Statistics

The data were evaluated using the SPSS version 22.0 package program. Data were analyzed using logistic regression analysis and the chi-square test. P-values less than 0.05 were considered statistically significant.

RESULTS

The study included 520 patients with acne vulgaris using ISO and 400 healthy controls. Of the participants using ISO, 68 (13.07%) were male and 452 (86.9%) were female. In the control group, 165 (41.25%) were male and 235 (58.75%) were female.

The mean age was 22.79 ± 4.90 in the ISO group and 22.18 ± 4.72 in the control group. There was no statistical difference between the two groups in terms of age and gender.

A total of 148 (16.1%) of all participants were married and 772 (83.9%) were single. As for the education level, 878 (95.4%) of all participants were high school graduates or above and 42 (4.56%) were primary school graduates.

Of the participants, 66 (12.7%) were from the ISO-using group and 89 (22.3%) from the control group had COVID-19 infection. When the patient and control groups were compared, the probability of having COVID-19 was lower in the group using ISO (p <0.001). In the group using isotretinoin, 189 patients tested for COVID -19, and 66 patients tested positive for SARS-CoV-2 PCR (+) (Table 1).

The cumulative dose was 34.35 ± 10.44 mg for the use of ISO. The cumulative dose in the ISO group was 35 ± 9.65 mg in those with COVID-19 infection, while it was 34.25 ± 10.56 mg in the group without COVID-19 infection. There was no statistically significant difference between the groups. The isotretinoin use period was 5.18 ± 2.4 months.

When both groups were evaluated together, 49 (21%) of the individuals who had COVID-19 were male, and 106 (15.4%) were female. When all groups were evaluated, the rate of COVID-19 transmission in men was statistically more significant than in women (p < 0.05). While 8 (11.8%) of 66 participants who used ISO and had COVID-19 infection were men, 58 (12.8%) were women. There was no statistically significant difference between them. While 41 (24.8%) of the 89 patients in the control group who had COVID-19 infection were men, 48 (20.4%) were women. There was no statistically significant difference between them.

When the treatment durations were evaluated, there was a consistent decrease in the rate of COVID-19

Table 1. The risk of having COVID-19

		Have you had Covid-19?		Total	p	OR
		Yes	No			
Group	Patient	66	454	520	<0.001	1,76
		12,7%	87,3%	100,0%		
Control	Control	89	311	400		
		22,3%	77,8%	100,0%		
Total		155	765	920		
		16,8%	83,2%	100,0%		

Table 2. The risk of having COVID-19 associated with the duration and dose of systemic ISO treatment

Variable	Category	Have you had COVID-19?		p
		Yes	No	
Group	Patient	66 (%12,7)	454 (%87,3)	<0,001
	Control	89 (%22,3)	311 (%77,8)	
Systemic ISO use time	0	89 (%22,3)	311 (%77,8)	0,002
	0-1 month	10 (%15,4)	55 (%84,6)	
	1-3 months	12 (%9,4)	116 (%90,6)	
	3-6 months	27 (%14,6)	158 (%85,4)	
	Over 6 months	17 (%12)	125 (%88)	
Systemic ISO dosage per day	0	89 (%22,3)	311 (%77,8)	0,002
	20 mg	12 (%9,8)	111 (%90,2)	
	30 mg	19 (%13,2)	125 (%86,8)	
	40 mg	25 (%15,9)	132 (%84,1)	
	Over 40 mg	10 (%10,4)	86 (%89,6)	

ISO: isotretinoin

infection with the use of ISO for the first 3 months. After a slight increase in the rate of COVID-19 infection between 3 and 6 months of treatment, the rate of COVID-19 decreased at the 6th iteration. The rate of COVID-19 infection significantly decreased with the use of ISO at 20 mg/day. Although there was a slight increase in the incidence of COVID-19 infection with the use of ISO at 30–40 mg/day, the rate tends to decrease again at doses above 40 mg (Table 2).

Multiple complaints were not observed in receiving drug treatment. Loss of smell and taste stood out compared to other complaints. Headache, joint, and muscle pains, and multiple complaints were more common in those who were not on drug treatment.

Of the participants, 51 continued ISO use during the COVID-19 pandemic, whereas 13 discontinued it. In total, 10 participants from the control group were hospitalized, but there were no hospitalizations among participants in the ISO group.

The risk of having COVID-19 infection in those who received drug treatment was calculated to be 1.76 times higher than the risk of having it in those who did not receive drug treatment (Table 1).

These findings suggest a statistically significant reduction in COVID-19 incidence among isotretinoin users. This aligns with previous studies on isotretinoin's role in ACE2 receptor modulation and immunomodulatory effects, emphasizing its potential as a preventive measure.

DISCUSSION

Our findings support the hypothesis that the downregulation of ACE2 receptors by isotretinoin may reduce susceptibility to SARS-CoV-2 by inhibiting viral entry, a critical step in COVID-19 pathophysiology (5,6). Consistent with previous studies, isotretinoin also has immunomodulatory properties that may attenuate inflammatory responses and potentially reduce the risk of severe outcomes. The absence of hospitalizations in the isotretinoin group in our cohort further supports a possible protective effect. By suppressing ACE2 receptors, isotretinoin may interfere with viral entry (7,8), while additional protective effects may arise from its regulation of immune responses and suppression of the mTORC1 pathway, which promotes inflammation (9- 13). Such mechanisms suggest that isotretinoin merits further investigation as a potential preventive agent against COVID-19.

Male patients typically experience higher COVID-19 mortality due to androgen-driven TMPRSS2 expression (14,15). The ability of isotretinoin to reduce DHT levels may therefore provide additional protection in males, consistent with our observation of no gender differences within the isotretinoin group. However, isotretinoin-related side effects such as nasal mucosa dryness could theoretically disrupt the nasal barrier and increase viral susceptibility (16,17), which may explain the transient mid-treatment increase in COVID-19 rates.

In line with this, Karadag et al. reported a suppression of Th1, Th2, and Th17 functions with isotretinoin treatment via reductions in TNF-α, IL-4, IL-17, and IFN-γ after three months (18). This immunosuppressive effect could explain the decline in COVID-19 rates during the first three months of isotretinoin use observed in our data. Loss of smell and taste was more frequent in the isotretinoin group, possibly related to disruption of the nasal barrier and ACE2/TMPRSS2 receptor expression in olfactory epithelial cells (19). Upregulation of genes such as LCN2, KRT23, and SERPINA3 has also been described in association with isotretinoin and in COVID-19 studies, which may contribute to inflammatory responses and could account for the transient increase in cases between the 3rd and 6th months of treatment (20,21).

Our current study further reveals that isotretinoin use does not increase the risk of contracting COVID-19 and may exert a protective effect. Acar et al. reported no increase in COVID-19 incidence or severity and no lung involvement in 186 patients on systemic retinoids (isotretinoin and acitretin) (22). Kuş et al. similarly found comparable infection rates between isotretinoin users and controls but with milder symptoms in the isotretinoin group (23). Demirel et al. also demonstrated the safe continuation of isotretinoin therapy during the pandemic (24). Havet et al. confirmed through larger epidemiological data that isotretinoin did not increase COVID-19 risk, despite reporting decreased drug access

during lockdowns (25). Our results are also supported by an interventional clinical trial by Shirvani et al., which demonstrated that isotretinoin use in COVID-19 patients accelerated clinical recovery and improved respiratory symptoms compared to controls (26). This study provides further evidence that isotretinoin may not only be safe during the pandemic but could also exert therapeutic benefits in infected patients.

Overall, these findings support the safety of isotretinoin under pandemic conditions and align with our observations.

Limitations of this study include its retrospective, questionnaire-based design, reliance on self-reported COVID-19 histories, and a relatively modest sample size. Nonetheless, in combination with previous evidence, our results provide further real-world support for the safe use of isotretinoin during the COVID-19 pandemic.

CONCLUSION

In light of these findings, isotretinoin use was not associated with increased COVID-19 risk. Its ability to modulate inflammatory responses warrants larger, controlled randomized trials to evaluate whether retinoid pathways merit clinical testing in COVID-19.

Ethical approval

This study has been approved by the Sivas Cumhuriyet University Non-Interventional Clinical Research Ethics Committee (approval date 13.01.2021, number 2021-01/41). Written informed consent was obtained from the participants.

Author contribution

Design: RYG; Data Collection: ATÜ, ATİ, VE; Analysis: MK, MT; Literature Search: RYG; Writing: RYG, MK. All authors reviewed the results and approved the final version of the article.

Source of funding

The authors declare the study received no funding.

Conflict of interest

The authors declare that there is no conflict of interest.

REFERENCES

1. World Health Organization (WHO). Coronavirus (COVID-19) dashboard. Available at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (Accessed on August 22, 2025).
2. Chu S, Michelle L, Ekelem C, Sung CT, Rojek N, Mesinkovska NA. Oral isotretinoin for the treatment of dermatologic conditions other than acne: a systematic review and discussion of future directions. Arch Dermatol Res. 2021; 313(6): 391-430. [Crossref]
3. Alkhaldy HY, Assiri AM, Fatima S, Owaidah T. Isotretinoin is active in the initial management of acute pro-myelocytic leukemia. Leuk Res Rep. 2020; 14: 100220. [Crossref]
4. Abbas AM, Ebrahim IM, Ramadan RG, Ali AR, Isotretinoin: a potential treatment for COVID-19. World Journal of Engineering Research and Technology. 2020; 6(8): 01-03.
5. Lillicrap D. Disseminated intravascular coagulation in patients with 2019-nCoV pneumonia. J Thromb Haemost. 2020; 18(4): 786-7. [Crossref]
6. Gundogdu M, Dere G. Is systemic isotretinoin use a risk factor for coronavirus disease 2019 (COVID-19)? J Cosmet Dermatol. 2021; 20(6): 1568-70. [Crossref]
7. Cousin VL, Giraud R, Bendjelid K. Pathophysiology of COVID-19: Everywhere You Look You Will See ACE2! Front Med (Lausanne). 2021; 8: 694029. [Crossref]
8. Sinha S, Cheng K, Aldape K, Schiff E, Ruppin E. Systematic cell line-based identification of drugs modifying ACE2 expression. Preprints. 2020: 2020030446. [Crossref]
9. Mashayekhi-Sardoo H, Hosseinjani H. A new application of mTOR inhibitor drugs as potential therapeutic agents for COVID-19. J Basic Clin Physiol Pharmacol. 2021; 33(1): 17-25. [Crossref]
10. Maiese K. The Mechanistic Target of Rapamycin (mTOR): Novel Considerations as an Antiviral Treatment. Curr Neurovasc Res. 2020; 17(3): 332-7. [Crossref]
11. Maione F, Casillo GM, Raucci F, et al. Interleukin-17A (IL-17A): A silent amplifier of COVID-19. Biomed Pharmacother. 2021; 142: 111980. [Crossref]
12. Sardana K, Verma G. Propionibacterium acnes and the Th1/Th17 Axis, Implications in Acne Pathogenesis and Treatment. Indian J Dermatol. 2017; 62(4): 392-4. [Crossref]
13. Melnik B. Dietary intervention in acne: Attenuation of increased mTORC1 signaling promoted by Western diet. Dermatoendocrinol. 2012; 4(1): 20-32. [Crossref]
14. Ciarambino T, Para O, Giordano M. Immune system and COVID-19 by sex differences and age. Womens Health (Lond). 2021; 17: 17455065211022262. [Crossref]

15. Feily A, Taheri T, Meier-Schiesser B, et al. The effect of low-dose isotretinoin therapy on serum androgen levels in women with acne vulgaris. Int J Womens Dermatol. 2019; 6(2): 102-4. **[Crossref]**

16. Gorpelioglu C, Ozol D, Sarifakioglu E. Influence of isotretinoin on nasal mucociliary clearance and lung function in patients with acne vulgaris. Int J Dermatol. 2010; 49(1): 87-90. **[Crossref]**

17. Takci Z, Simsek GG, Karabulut H, Buran Y, Karadag AS. Effect of systemic isotretinoin therapy on mucociliary clearance and nasal surface mucosa in acne patients. J Drugs Dermatol. 2013; 12(8): e124-8.

18. Karadag AS, Ertugrul DT, Bilgili SG, Takci Z, Akin KO, Calka O. Immunoregulatory effects of isotretinoin in patients with acne. Br J Dermatol. 2012; 167(2): 433-5. **[Crossref]**

19. Meng X, Pan Y. COVID-19 and anosmia: The story so far. Ear Nose Throat J. 2024; 103(5): NP312-20. **[Crossref]**

20. Jha D, Kand KS, Gautam HK. Regulation of Gene Expression by Quantitative Real Time PCR in Low Dose Isotretinoin Treated Acne Patients. Gene Technology. 2015; S1: 001. **[Crossref]**

21. Karakurt HU, Pir P. Integration of transcriptomic profile of SARS-CoV-2 infected normal human bronchial epithelial cells with metabolic and protein-protein interaction networks. Turk J Biol. 2020; 44(3): 168-77. **[Crossref]**

22. Acar EM, Özyurt K, Akyol R. Evaluation of COVID-19 risk in patients on systemic retinoid therapy. Turkderm - Turk Arch Dermatol Venereol. 2022; 56(3): 109-12. **[Crossref]**

23. Kuş MM, Öztürk P, Bulut T, et al. How are acne vulgaris patients using oral isotretinoin affected by COVID-19? Dermatol Pract Concept. 2022; 12(3): e2022146. **[Crossref]**

24. Demirel Öğüt N, Kutlu Ö, Erbağcı E. Oral isotretinoin treatment in patients with acne vulgaris during the COVID-19 pandemic: A retrospective cohort study in a tertiary care hospital. J Cosmet Dermatol. 2021; 20(7): 1969-74. **[Crossref]**

25. Havet A, Moskal A, Payet C, et al. Prescribing and dispensing of oral isotretinoin and noncompliance with recommendations in France before and during the COVID-19 pandemic. Dermatol Ther (Heidelb). 2025; 15(6): 1533-41. **[Crossref]**

26. Shirvani M, Vaziri S, Akrami MR, Sarmasti A, Hassanpour K, Akrami A. Evaluating the therapeutic effects of isotretinoin on patients with coronavirus disease 2019 (COVID-19): A controlled open-label clinical trial. Heliyon. 2024; 10(5): e26685. **[Crossref]**

NORTHWESTERN MEDICAL JOURNAL

Northwestern Med J. 2025;5(4):240-247
DOI: 10.54307/NWMJ.2025.131

RESEARCH ARTICLE

Age- and gender-related changes in cerebral ventricle parameters during the pediatric period

Abdullah Ray[1], **İbrahim Kürtül**[1]

[1]Department of Anatomy, Faculty of Medicine, Bolu Abant İzzet Baysal University, Bolu, Türkiye

Cite as: Ray A, Kürtül İ. Age- and gender-related changes in cerebral ventricle parameters during the pediatric period. Northwestern Med J. 2025;5(4):240-247.

ABSTRACT

Aim: This study focused on investigating age- and gender-related changes in the brain ventricles of healthy pediatric individuals.

Methods: Brain MR images of 200 healthy children aged 0-18 years were included in the study. The variables measured were as follows: axis of the third ventricle (ATV), anterior width of the frontal horn (ACF), posterior width of the frontal horn (PCF), width of the frontal horn (WCF), oblique diameter of the frontal horn (OCF), maximum transverse diameter of the skull (MTDS), vertical diameter of the skull (VDS), anteroposterior width of the right temporal horn (ARCT), anteroposterior width of the left temporal horn (ALCT), anteroposterior width of the fourth ventricle (AFV), transverse width of the fourth ventricle (TFV). In addition, the Evans index (EI) has been calculated.

Results: Statistically significant results were found between the individuals of the first, second and fourth groups for the ACF and PCF variables; the fourth group for the WCF, VDS, and TFV variables; the first and fourth groups for ARCT; the first group for ALCT; and the third and fourth groups for MTDS. In the pediatric period, while there was no significant difference between the genders until a certain age, it was observed that the difference between the genders increased especially after a certain age (between 7-18 years).

Conclusion: It is thought that the study will provide basic data for clinical sciences in the stages of diagnosis and treatment planning.

Keywords: Pediatric period, brain ventricles, MRI brain, brain, central nervous system

INTRODUCTION

The nervous system begins to develop rapidly at a very early embryonic stage and continues to build up at a slower rate after birth. It regulates the control mechanisms of the body, including vital functions such as reflexes, in relation to the endocrine system. As the central nervous system (CNS) develops, the central canal expands in several regions to form the brain's ventricles. The ventricular system of the brain, which is

Corresponding author: Abdullah Ray **E-mail:** abdullahray10@gmail.com
Received: 09.07.2024 **Accepted:** 30.04.2025 **Published:** 26.10.2025

hollow and filled with cerebrospinal fluid (CSF), consists of four components: the lateral right and left ventricles, and the third and fourth ventricles. The production and circulation of CSF take place in these structures (1-3). CSF acts both as a protector of the CNS and plays an important role in brain metabolism. Various pathologies in these structures are observed in several neurodegenerative conditions (4,5), resulting in brain atrophy. Ventricular enlargement is the most essential indicator in determining the prognosis of Alzheimer's disease. Therefore, a detailed understanding of the anatomical associations between the ventricular system and the neighboring brain tissue is vital in defining intraventricular pathology and the potential safety risks associated with surgical treatment (6,7). The morphologic development of these structures in the prenatal and postnatal periods is valuable for neurosurgeons, neurologists, and radiologists (8-10). With this in mind, this study conducted in Bolu Abant Izzet Baysal University (BAIBU), Training and Research Hospital (TRH) has focused on examining age- and gender-related changes in the parameters of the brain ventricles during the pediatric period. The results of the study will contribute to the basic data in the relevant literature and clinical sciences.

MATERIAL AND METHOD

This study was conducted using brain MRI images of 200 children (100 girls, 100 boys) aged 0-18 years who were admitted to BAIBU TRH, with no pathological conditions in the brain or brain ventricles. The ethics committee permission, obtained from BAIBU Clinical Research Ethics Committee with decision number 2022/99, was strictly followed at every stage of the study.

Brain MR images of the subjects who fulfilled the criteria for admission were selected from the PACS of BAIBU TRH in T2 acquisition protocol. MR images were obtained with a 1.5 T Signa Explorer MRI Scanner (GE Medical Systems, Milwaukee, Wisconsin, USA). Images in the PACS system were downloaded in DICOM format. They then were imported into the Radiant DICOM Viewer (RDV) program, which was used as a personal workstation. The parameters of the measurements performed on axial T2 prop MR images were determined as slice thickness 3.5 mm, repetition time 7143 ms, and echo time 115.136 ms.

In the study, the following measurements were acquired; axis of the third ventricle (ATV), anterior width of the frontal horn (ACF), posterior width of the frontal horn (PCF), width of the frontal horn (WCF), oblique diameter of the frontal horn (OCF), maximum transverse diameter of the skull (MTDS), vertical diameter of the skull (VDS), anteroposterior width of the right temporal horn (ARCT), anteroposterior width of the left temporal horn (ALCT), anteroposterior width of the fourth ventricle (AFV), transverse width of the fourth ventricle (TFV), the Evans index (EI). Demonstrations of the measured variables are shown in Figure 1.

To reduce inter-observer error, three measurements were taken by an expert and their averages were recorded. Images of the brain ventricles were examined by dividing the subjects into four groups of 50 individuals (25 boys, 25 girls) in each group according

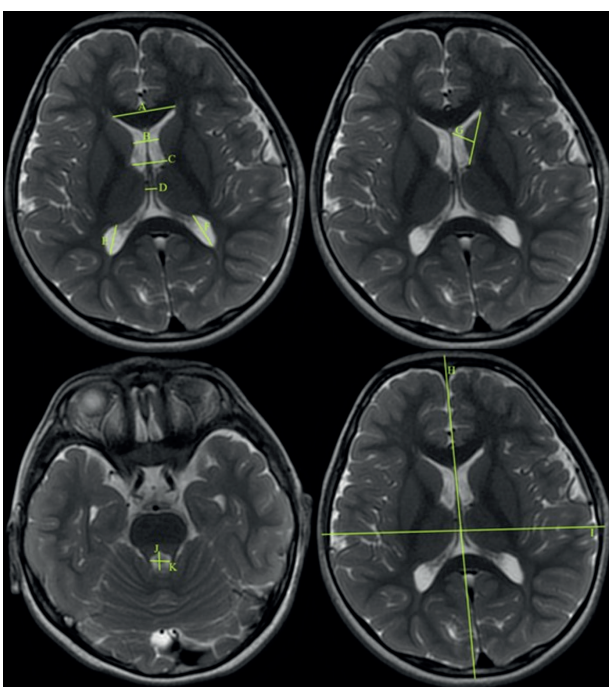


Figure 1. Variable demonstrations

A: ACF [anterior width of the frontal horn], B: WCF [width of the frontal horn], C: PCF [posterior width of the frontal horn], D: ATV [axis of the third ventricle], E: ALCT [anteroposterior width of the left temporal].

to age. The groups were as follows: group 1 – healthy children aged 0-2 years; group 2 – healthy children aged 3-6 years; group 3 – healthy children aged 7-11 years; group 4 – healthy children aged 12-18 years.

Statistical analysis

Statistical analyses were performed with the Minitab® 21.2 program. The appropriateness of the residuals to normal distribution was tested with the Anderson-Darling test. Logarithmic and square root transformations were applied to nonparametric variables. As a result of the transformation, it was determined that the transformations had no effect. The difference between the genders forming each group was tested with independent t test for parametric variables and Mann-Whitney U test for nonparametric variables. For parametric variables, two-factor analysis of variance was applied, and for nonparametric variables, the Kruskal-Wallis test was used. Tukey analysis was performed as a post-hoc test for variables for which the difference between groups was found to be significant as a result of variance analysis. The inter-group test of the variables that were significant as a result of the Kruskal-Wallis test was performed using the Mann-Whitney U test. BoxPlot graphs showing the change of variables according to age and gender were drawn.

RESULTS

The mean and standard deviation (SD) values of the parametric variables, and the min, max, and median values of the nonparametric variables were calculated. The p values from the two-way ANOVA for age and group interactions of the parametric variables, and the p values from the Kruskal-Wallis test for the nonparametric distributed variables were acquired. Table 1 presents the results of the descriptive statistics for the variables.

Table 1. Descriptive statistics results of the variables

Variables	Gender	n	Group 1	Group 2	Group 3	Group 4	p value
Age	M	25	1.0 (0.0-2.0)^{a}	5.0 (3.0-6.0)^{b}	9.0 (7.0-11.0)^{c}	15.0 (12,0-18.0)^{d}	<0.001^{k}
	F	25	1.0 (0.0-2.0)^{a}	6.0 (3.0-6.0)^{b}	9.0 (7.0-11.0)^{c}	13.0 (12.0-18.0)^{d}	<0.001^{k}
	p value		0.561^{x}	0.559^{x}	0.985^{x}	0.042^{x}	
ATV	M	25	4.0 (3.3-4.5)^{b}	3.5 (3.0-4.0)^{a}	3.8 (3.0-4.8)^{b}	3.8 (3.3-5.1)^{b}	0.002^{k}
	F	25	3.9 (2.5-4.5)	3.4 (3.0-4.3)	3.5 (3.0-4.8)	3.6 (3.0-5.1)	0.065^{k}
	p value		0.816^{x}	0.376^{x}	0.103^{x}	0.048^{x}	
ACF	M	25	29.3±3.2^{b}	32.0±3.2^{a}	31.8±2.5^{a}	33.2±3.0^{a}	<0.001^{m}
	F	25	26.8±3.8^{b}	29.9±2.9^{a}	31.4±2.5^{a}	31.4±2.9^{a}	<0.001^{m}
	p value		0.018^{y}	0.017^{y}	0.412^{y}	0.042^{y}	0.380^{m}
PCF	M	25	15.7±2.4^{a}	15.0±2.8^{ab}	13.8±2.0^{ab}	14.8±2.0^{b}	0.039^{m}
	F	25	13.9±2.4	13.1±2.5	13.6±2.5	12.8±2.0	0.410^{m}
	p value		0.013^{y}	0.012^{y}	0.587^{y}	0.001^{y}	0.180^{m}
WCF	M	25	12.3 (8.4-17.8)	10.7 (6.3-18.6)	11.2 (6.6-17.1)	11.5 (8.4- 15.9)	0.065^{k}
	F	25	11.7 (5.6-17.1)	10.2 (6.6-15.3)	9.6 (6.6-17.1)	9.5 (6.3-17.3)	0.466^{k}
	p value		0.130^{x}	0.101^{x}	0.132^{x}	0.003^{x}	
OCF	M	25	11.1 (7.2-13.1)	10.8 (7.2-18.3)	10.6 (7.6-16.3)	11.1 (7.7-15.3)	0.723^{k}
	F	25	9.4 (5.8-16.0)	10.1 (7.8-13.9)	10.9 (8.7-16.0)	10.5 (7.3-13.5)	0.131^{k}
	p value		0.165^{x}	0.258^{x}	0.578^{x}	0.180^{x}	
ARCT	M	25	13.6±2.1	14.3±2.7	14.7±2.7	15.1±2.6	0.215^{m}
	F	25	12.2±2.0	13.1±2.5	13.9±2.2	13.1±2.7	0.095^{m}
	p value		0.019^{y}	0.103^{y}	0.288^{y}	0.010^{y}	0.610^{m}

k: The p value of the Kruskal-Wallis test result, m: The p value of the two-way ANOVA result, x: The p value of the Mann-Whitney U test result, y: The p value of the two-sample t test result.

Table 1. Continued

Variables	Gender	n	Group 1	Group 2	Group 3	Group 4	p value
VDS	M	25	150.5±14.4^{c}	167.8±8.8^{b}	170.2±10.0^{ab}	176.2±6.5^{a}	<0.001^{m}
	F	25	144.2±15.3^{c}	163.9±8.9^{b}	166.5±6.2^{ab}	172.1±6.5^{a}	<0.001^{m}
	p value		0.139^{y}	0.132^{y}	0.101^{y}	0.034^{y}	0.909^{m}
MTDS	M	25	125.1±11.3^{c}	139.2±7,2^{b}	142.7±4,9^{b}	153.2±8.8^{a}	<0.001^{m}
	F	25	118.6±14.5^{c}	136.2±5.0^{b}	139.1±7.0^{b}	144.9±9.3^{a}	<0.001^{m}
	p value		0.085^{y}	0.089^{y}	0.033^{y}	0.002^{y}	0.402^{m}
AFV	M	25	7.8±1.8^{b}	8.7±1.2^{ab}	9.6±1.5^{ab}	9.5±1.5^{a}	0.001^{m}
	F	25	8.0±1.7^{b}	8.6±1.7^{ab}	8.8±1.2^{a}	9.6±1.7^{a}	0.009^{m}
	p value		0.670^{y}	0.710^{y}	0.041^{y}	0.878^{y}	0.374^{m}
TFV	M	25	10.2±1.8^{b}	11.3±1.3^{b}	12.6±2.0^{a}	12.9±1.5^{a}	0.001^{m}
	F	25	10.7±1.7	11.4±1.8	11.7±1.3	11.9±1.6	0.078^{m}
	p value		0.341^{y}	0.990^{y}	0.070^{y}	0.034^{y}	0.072^{m}
EI	M	25	0.2±0.02^{a}	0.2±0.02^{ab}	0.2±0.02^{ab}	0.2±0.02^{b}	0.037^{m}
	F	25	0.2±0.02	0.2±0.01	0.2±0.01	0.2±0.02	0.321^{m}
	p value		0.263^{y}	0.067^{y}	0.776^{y}	0.988^{y}	0.389^{m}

k: The p value of the Kruskal-Wallis test result, m: The p value of the two-way ANOVA result, x: The p value of the Mann-Whitney U test result, y: The p value of the two-sample t test result.

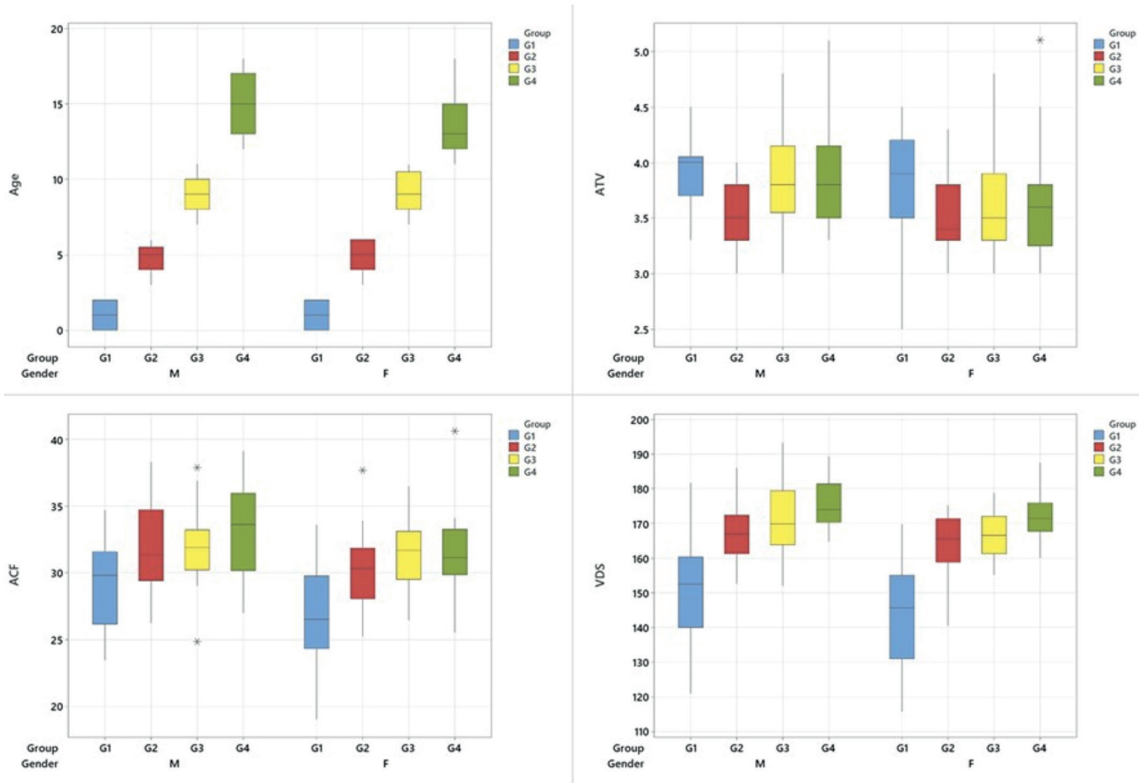


Figure 2. Boxplot for the Age, ATV (axis of the third ventricle), ACF (anterior width of the frontal horn), VDS (vertical diameter of the skull).

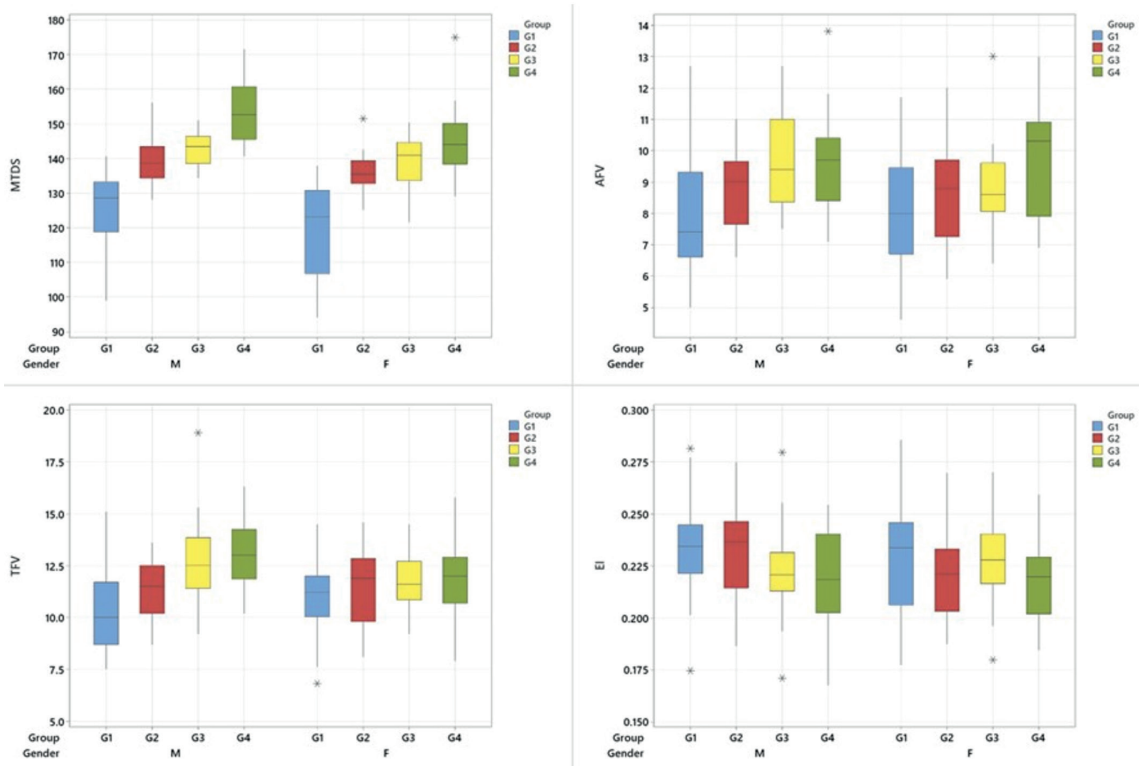


Figure 3. Boxplot for the MTDS (maximum transverse diameter of the skull), AFV (anteroposterior width of the fourth ventricle), TFV (transverse width of the fourth ventricle), EI (Evans index).

BoxPlots for Age, ATV, ACF, VDS, MTDS, AFV, TFV, EI variables are presented in Figure 2 and Figure 3.

DISCUSSION

This research has focused on revealing the changes in the ventricular system of the brain in relation to age and gender in the pediatric period, using brain MR images of 200 children aged 0-18 years, admitted to BAIBU TRH without any pathological condition affecting the brain or brain ventricles. The images of the brain ventricles have been evaluated in four groups based on age. The following parameters have been measured, and the following results have been interpreted: ATV, ACF, PCF, WCF, OCF, MTDS, VDS, ARCT, ALCT, AFV, TFV, and EI. The results have indicated that after a certain age in the pediatric period, the morphometry of the brain ventricles and other variables of the skull differ significantly in relation with gender. The development of brain ventricles in the pediatric period is critical for the early diagnosis of many diseases. It can be seen from the accumulating literature that the number of studies conducted on the ventricular system of the brain and the number of individuals examined in the pediatric period are limited (11-13). Yet, to our knowledge, there is no study using MR images of healthy individuals in this period to evaluate the development of the ventricular system of the brain in particular. Hence, in this study, we have focused on the age- and gender-related changes of the brain ventricles in individuals aged 0-18 years.

It has been shown in the literature that neurodegenerative diseases play an aggregative role in the disruption of the morphometric structure of the brain ventricles (11). Hydrocephalus, an example of a neurodegenerative disease, is a disorder characterized by dilatation of the ventricular system and obstruction of CSF flow. In a study investigating the differences among the brain ventricles of individuals diagnosed with hydrocephalus in relation to gender, a total of 50 patients (27 boys, 23 girls) diagnosed with either

myelomeningocele-related hydrocephalus (MHC) or non-myelomeningocele-related hydrocephalus (HC) in the pediatric period have been put under research. ATV and EI variables have been analyzed on CT brain and MR brain images of the individuals. While no statistically significant result has been seen between the genders, the results have differed significantly with age. Statistically significant changes have been found between HC and MHC groups in terms of EI and ATV. While our research has focused on the cerebral ventricle development in healthy individuals, the study indicated above has examined individuals diagnosed with hydrocephalus, expressing greater differences in the median values. This is surely due to hydrocephalus having a very high impact on the morphometry of the brain and brain ventricles. One study has investigated the importance of EI and ventricular indexes in revealing the prognosis of hydrocephalus in 137 patients from birth to six years of age out of the Nigerian population (14). This research has documented that the ventricular index and the EI variable reflect the prognosis of the disease. Another research has examined the structure of the brain ventricles of individuals with multiple sclerosis (MS), another example of a neurodegenerative disease, and the structures of the brain ventricles of healthy people (15). The study has focused on a group of patients with two stages of MS (RR-Relapsing Remitting and SPMS-Secondary Progressive). MR images of 40 MS patients (20 RR and 20 SPMS), and 10 control subjects were used to analyze ATV, AFV, CFAG, ARCT, ALCT, and VDS variables. As a result of the analysis, the ATV variable has been found to be lower in the control group, and the differences among all groups have been found statistically significant. The EI variable has not demonstrated a level of statistical significance, while the AFV, ACF, ARCT, and ALCT variables have differed significantly among the groups. The individuals with MS have shown higher values compared to the control group. Upon analyzing the results, it is well understood that neurodegenerative diseases affect the size and volume of brain ventricles. Therefore, it is vital to know the morphometric dimensions of brain ventricles in healthy individuals, as this can facilitate both the prognosis and the diagnosis of the disease.

Similarly, a study has measured ACF, ATV, TFV, AFV, MTDS, and EI variables in adults, and has found that ACF, ATV, and MTDS variables have shown a statistically significant difference between genders, while no difference has been seen in AFV, TFV, and EI variables (16). Likewise, in a study conducted on MR images of adolescents (17), ATV, OCF and ACF variables have been analyzed, reporting no statistically significant difference in ATV and OCF variables, while a statistically significant difference in ACF in terms of gender was reported. ACF, EI, and MTDS variables have been evaluated in another study conducted on 100 individuals from the Indian population with an average age between 5-90 years, including seven pediatric patients, and the differences among the genders and groups have not been statistically significant in the EI variable (12). Moreover, the differences in the variables ACF and MTDS have been significant, in terms of gender, while no difference has been found between the groups. In a study in which the same variables have been examined in the adult Nigerian population, a statistically significant result has been found for ACF and MTDS variables in terms of gender, and no statistically significant result has been detected in the EI variable (18). In the current study, the EI variable among these variables has not been found to be statistically different in terms of gender, whereas it has been found to be significantly different between individuals in the first and fourth groups. The EI of the individuals in the first group has been found to be lower than that of individuals in the other group. This result shows that the EI variable progresses with age but remains constant after a certain age, ultimately indicating no difference between the genders. As far as the ACF and MTDS variables are concerned, no statistically significant difference has been observed between the second and third groups, while a statistically significant difference has been observed between the first group and the fourth group. It has been determined that these variables have not differed in the period from the age of three to the age of 11. Statistically significant results have been found between the individuals in the first, second, and fourth groups for the ACF variable, and between the individuals in the third and fourth groups for the MTDS variable. After a certain age (seven years), the difference between the genders has increased in the transverse diameter of the skull, indicating that pediatric individuals start to differentiate between genders after the age of seven.

A study conducted on the temporal lobe development of pediatric individuals has examined MR images of individuals aged from three weeks to 14 years (13). The study found that the temporal lobe grows sharply until the age of two, but the growth rate slows down right after this age. The difference in growth rates between the right and left temporal lobes has been found to be statistically significant. Moreover, while temporal lobe volume has been found to be inversely proportional to age, the right temporal lobe has been found to be larger than the left temporal lobe, except for the individuals between the ages of one and two. Similar results have also been acquired in our research. While the ARCT variable has had a higher value in all groups compared to the ALCT variable, no statistically significant difference has been detected between the groups. However, a statistically significant difference has been found among individuals in the first and fourth groups for the ARCT variable, and among individuals in the first group for the ALCT variable, in terms of gender. Having evaluated these results, it is obvious that the growth in the temporal lobe, brain, and skull affects the growth of the left and right temporal horn at a similar rate.

Consequently, this study was conducted to examine changes in the brain ventricles of pediatric individuals in relation to age and gender, has determined that the ACF, PCF, ARCT, and ALCT variables for the first group; the ACF and PCF variables for the second group; the VDS, MTDS, and AFV variables for the third group; and the ATV, ACF, PCF, WCF, ARCT, VDS, MTDS, and TFV variables for the fourth group have differed significantly with respect to gender. Analysis has revealed a significant difference among the age groups for all variables except WCF, OCF, ARCT, and ALCT. It has been observed that in the pediatric period, the differences in the brain ventricles increased between the ages of 12 and 18, and both the vertical and transversal dimensions of the skull notably changed between the ages of seven and 18.

Ethical approval

This study has been approved by the Bolu Abant İzzet Baysal University Clinical Research Ethics Committee (approval date 10.05.2022, number 2022/99).

Author contribution

Concept: AR, İK; Design: AR, İK; Data Collection or Processing: AR; Analysis or Interpretation: AR; Literature Search: AR; Writing: AR, İK. All authors reviewed the results and approved the final version of the article.

Source of funding

The authors declare the study received no funding.

Conflict of interest

The authors declare that there is no conflict of interest.

REFERENCES

1. Kerscher SR, Schweizer LL, Nägele T, Weichselbaum A, Haas-Lude K, Schuhmann MU. Changes of third ventricle diameter (TVD) mirror changes of the entire ventricular system after initial therapy and during follow-up in pediatric hydrocephalus. Eur J Paediatr Neurol. 2019; 23(4): 571-80. [Crossref]
2. Khan AF, Drozd JJ, Moreland RK, et al. A novel MRI-compatible brain ventricle phantom for validation of segmentation and volumetry methods. J Magn Reson Imaging. 2012; 36(2): 476-82. [Crossref]
3. Fame RM, Cortés-Campos C, Sive HL. Brain Ventricular System and Cerebrospinal Fluid Development and Function: Light at the End of the Tube: A Primer with Latest Insights. Bioessays. 2020; 42(3): e1900186. [Crossref]
4. Stratchko L, Filatova I, Agarwal A, Kanekar S. The Ventricular System of the Brain: Anatomy and Normal Variations. Semin Ultrasound CT MR. 2016; 37(2): 72-83. [Crossref]
5. Karaca O, Buyukmert A, Tepe N, Ozcan E, Kus I. Volume estimation of brain ventricles using Cavalieri's principle and Atlas-based methods in Alzheimer disease: Consistency between methods. J Clin Neurosci. 2020; 78: 333-8. [Crossref]
6. Korzh V. Development of brain ventricular system. Cell Mol Life Sci. 2018; 75(3): 375-83. [Crossref]
7. Kremen WS, Panizzon MS, Neale MC, et al. Heritability of brain ventricle volume: converging evidence from inconsistent results. Neurobiol Aging. 2012; 33(1): 1-8. [Crossref]
8. Ma W, Li XJ, Li W, Xiao L, Ji XJ, Xu Y. MRI findings of central nervous system involvement in children with haemophagocytic lymphohistiocytosis: correlation with clinical biochemical tests. Clin Radiol. 2021; 76(2): 159.e9-159.e17. [Crossref]

9. Shao M, Han S, Carass A, et al. Brain ventricle parcellation using a deep neural network: Application to patients with ventriculomegaly. Neuroimage Clin. 2019; 23: 101871. [Crossref]

10. Zhuravlova I, Kornieieva M, Rodrigues E. Anatomic Variability of the Morphometric Parameters of the Fourth Ventricle of the Brain. J Neurol Surg B Skull Base. 2018; 79(2): 200-4. [Crossref]

11. Idowu OE, Balogun MM. Visual function in infants with congenital hydrocephalus with and without myelomeningocoele. Childs Nerv Syst. 2014; 30(2): 327-30. [Crossref]

12. Arun Kumar S, Meena Kumari S, Vijai Anand M, Saraswathy R, Rajeshwari M. Evaluation of Evan's Index in South Indian Population using Computed Tomography. International Journal of Anatomy, Radiology and Surgery. 2017; 6(3): RO28-RO31. [Crossref]

13. Utsunomiya H, Takano K, Okazaki M, Mitsudome A. Development of the temporal lobe in infants and children: analysis by MR-based volumetry. AJNR Am J Neuroradiol. 1999; 20(4): 717-23.

14. Idowu O, Olumide A. Etiology and cranial CT scan profile of nontumoral hydrocephalus in a tertiary black African hospital. J Neurosurg Pediatr. 2011; 7(4): 397-400. [Crossref]

15. Turner B, Ramli N, Blumhardt LD, Jaspan T. Ventricular enlargement in multiple sclerosis: a comparison of three-dimensional and linear MRI estimates. Neuroradiology. 2001; 43(8): 608-14. [Crossref]

16. Polat S, Öksüzler FY, Öksüzler M, Kabakci AG, Yücel AH. Morphometric MRI study of the brain ventricles in healthy Turkish subjects. International Journal of Morphology. 2019; 37(2): 554-60. [Crossref]

17. Aukland SM, Odberg MD, Gunny R, Chong WKK, Eide GE, Rosendahl K. Assessing ventricular size: is subjective evaluation accurate enough? New MRI-based normative standards for 19-year-olds. Neuroradiology. 2008; 50(12): 1005-11. [Crossref]

18. Hamidu AU, Olarinoye-Akorede SA, Ekott DS, Danborno B, Mahmud MR, Balogun MS. Computerized tomographic study of normal Evans index in adult Nigerians. J Neurosci Rural Pract. 2015; 6(1): 55-8. [Crossref]

NORTHWESTERN MEDICAL JOURNAL

Northwestern Med J. 2025;5(4):248-254
DOI: 10.54307/NWMJ.2025.197

RESEARCH ARTICLE

Quality and reliability of Youtube videos on percutaneous ablation of thyroid nodules

Samet Genez[1], **Yunus Yılmazsoy**[1], **Hamza Özer**[1], **Soner Türkoğlu**[1], **Zehra Nur Baloğlu**[1]

[1]Department of Radiology, Faculty of Medicine, Bolu Abant İzzet Baysal University, Bolu, Türkiye

Cite as: Genez S, Yılmazsoy Y, Özer H, Türkoğlu S, Baloğlu ZN. Quality and reliability of Youtube videos on percutaneous ablation of thyroid nodules. Northwestern Med J. 2025;5(4):248-254.

ABSTRACT

Aim: We aimed to assess the quality and reliability of YouTube videos on percutaneous ablation of thyroid nodules (PATN) to evaluate their utility as educational tools for patients and healthcare professionals.

Materials and Methods: This cross-sectional study analyzed 76 YouTube videos identified through keyword searches ("percutaneous ablation of thyroid nodules", "thyroid radiofrequency ablation", and "thyroid microwave ablation") on November 10, 2024. Videos were categorized by type (informative, technical, personal experience, and news), duration (≤4 minutes, >4 minutes), upload source (professional healthcare providers, non-professionals, or independent users), and target audience (patients or healthcare professionals).

Quality and reliability were assessed using the Journal of the American Medical Association (JAMA) criteria, the Global Quality Score (GQS), and the modified DISCERN (mDISCERN). Statistical analyses, including the Kruskal-Wallis and Spearman correlation tests, were conducted.

Results: A total of 76 videos were evaluated. Informative videos scored significantly higher on quality metrics (GQS, 2.85 ± 0.15, p= 0.002; mDISCERN, 2.13 ± 0.16, p= 0.008). Videos >4 minutes demonstrated higher quality scores (p= 0.001). No statistically significant differences in popularity metrics (likes, comments, view rates) were observed across groups (p> 0.05). No significant correlations were found between quality scores (JAMA, GQS, mDISCERN) and popularity metrics (R^2 = -0.019 to 0.147).

Conclusion: While informative and longer videos exhibited higher quality, popularity metrics were not reliable indicators of video quality. These findings highlight the need for healthcare professionals to produce engaging and accurate content for platforms like YouTube to improve public education about PATN.

Keywords: YouTube, percutaneous ablation, thyroid nodule, video quality, patient education

Corresponding author: Samet Genez **E-mail:** sametgenez@hotmail.com
Received: 28.04.2025 **Accepted:** 26.06.2025 **Published:** 26.10.2025

INTRODUCTION

Percutaneous ablation of thyroid nodules (PATN) has recently received significant attention as a minimally invasive option for the treatment of benign thyroid nodules. Compared to traditional surgical approaches, ablation procedures offer numerous advantages, including preservation of thyroid function, shorter recovery times, and fewer complications. These benefits have increased popularity among clinicians and patients seeking less invasive management options (1).

With the rise of digital technology, the Internet has become a primary source of health-related information for patients. Video-sharing platforms such as YouTube play an important role in this shift by providing easily accessible and engaging visual content about medical procedures. Studies have shown that patients increasingly use YouTube to learn about new treatment options and understand procedures, outcomes, and potential risks (2).

However, the quality and reliability of health-related content on YouTube remain a significant concern. Unlike peer-reviewed medical literature, content uploaded to video-sharing platforms is not subject to quality-control, allowing the spread of inaccurate, biased, or incomplete information that can mislead patients and influence treatment decisions (3). As a result, patients face challenges distinguishing between reliable guidance and misinformation.

Given the increasing interest in thyroid RFA and the widespread use of YouTube as a patient education tool, it is essential to assess the quality and reliability of videos on this procedure. In this study, we aimed to evaluate the reliability and quality of YouTube videos on PATN to provide accurate information to patients and healthcare professionals.

MATERIALS AND METHODS

This cross-sectional study evaluated the quality and reliability of YouTube videos related to PATN on November 10, 2024; the keywords "percutaneous ablation of thyroid nodules", "thyroid radiofrequency ablation", and "thyroid microwave ablation" were searched on YouTube. Searches were conducted in incognito mode to avoid algorithmic biases, and the default relevance-based sorting filter was applied. Videos were included if they were in English, provided visual or verbal explanations of thyroid ablation techniques, and had a resolution of at least 360p. Videos that were advertisements, unrelated to the topic, or exceeded 30 minutes in length were excluded. After sorting the videos by relevance, 76 of them met the inclusion criteria and were included in the analysis.

Two physicians independently reviewed the videos, recording the upload date, duration (minutes), total views, number of likes and comments. The viewing rate was calculated as the number of views per day since the video was uploaded. Videos were categorized by the source of upload as those uploaded by professional healthcare providers, non-professional individuals, and independent users. The target audience was also classified as healthcare professionals/doctors or patients and their relatives (4).

The quality and reliability of the videos were evaluated using three validated tools (Figure 1). The modified DISCERN (mDISCERN) score assessed the reliability of the information on a scale of 0 to 5, with higher scores indicating better reliability (5). The Global Quality Scale (GQS) was used to evaluate the videos' overall educational value and organization on a 5-point scale (6). The Journal of the American Medical Association (JAMA) benchmark criteria were applied to assess authorship, attribution, disclosure, and currency, with a maximum score of 4 indicating high reliability (7).

The Shapiro-Wilk test was applied to assess the normality of the data distribution. Non-parametric tests were used as the data did not follow a normal distribution. The Kruskal-Wallis test was applied to analyze differences in scores across groups, while the Spearman correlation test was used to evaluate associations between continuous variables. Interobserver agreement was assessed using the linear weighted Kappa statistic. A p-value of less than 0.05 was considered statistically significant. All statistical analyses were conducted using SPSS software, version 24.

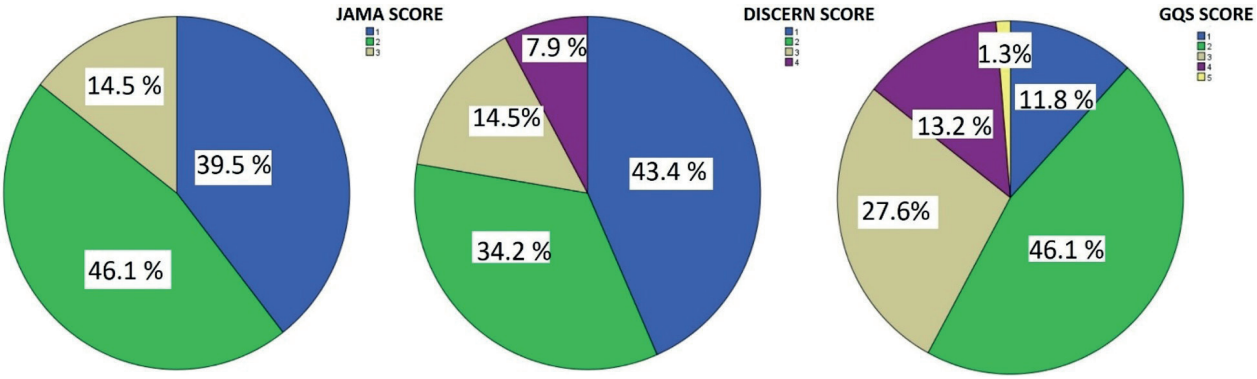


Figure 1. Distribution of JAMA, mDISCERN, and GQS scores across all analyzed YouTube videos.

RESULTS

A total of 76 YouTube videos were analyzed and categorized based on their characteristics. The mean number of views was 6,730.86, with a mean of 70.92 likes, 12.08 comments, and a viewing rate of 7.08 views per day. Among the analyzed videos, 71.1% (n=54) were uploaded by professional healthcare providers, 15.8% (n=12) by non-professional individuals, and 13.2% (n=10) by independent users. Regarding the

Modified DISCERN	
1. Is the aim clear, concise, understandable?	
2. Are sources of information reliable? (Cited publication, speaker is specialist in thyroid ablation procedure)	
3. Is the information presented balanced and unbiased? (Any reference to other treatment choices)	
4. Are additional sources of information listed?	
5. Does the video address areas of uncertainty?	
JAMA	
Authorship	Authors, partnerships and contributors, their affiliations and relevant credentials should be provided
Attribution	References and resources used for the content and a copyright statement should be listed clearly
Currency	Sponsorship, advertising, commercial financing, potential conflicts of interest should be prominently and fully disclosed
Disclosure	Dates when content was posted and updated should be indicated
GQS	
1	Poor quality, poor flow, most information missing, not useful for education
2	Generally poor quality and flow, of limited use to patients because only some information is present but many important topics missing.
3	Moderate quality, suboptimal flow, somewhat useful for patients as some important information is adequately discussed but others poorly discussed.
4	Good quality, generally good flow, useful to patients because most relevant information is covered but some topics not covered.
5	Excellent quality and flow, highly useful to patients.

Figure 2. Modified DISCERN, JAMA and GQS score.

JAMA: Journal of American Medical Association; GQS: Global Quality Score.

Table 1. Descriptive characteristics of the YouTube videos included in the analysis

Characteristic Features of Videos	Number (Percentages)
Video Type	
Informative	39 (51.3 %)
Technique	12 (15.8 %)
Personal Experience	24 (31.6 %)
News Update	1 (1.3 %)
Duration	
0-4 minutes	49 (64.5 %)
>4 minutes	27 (35.5 %)
Upload time	
0-6 months	8 (10.5 %)
6-12 months	6 (7.9 %)
>12 months	62 (81.6 %)
Upload Source	
Independent Users	10 (13.2 %)
Professional Health Care Providers	54 (71.1 %)
Non-professional	12 (15.8)
Target Population	
Patients and Relatives	62 (81.6 %)
Health Care Professionals	14 (18.4 %)
Popularity	**Mean (Range)**
Total number of views	6730.86 (53- 81218)
Total number of likes	70.92 (0-1000)
Total number of comments	12.08 (0-346)
View Rate	7.08 (0.13-59.05)

target audience, 81.6% (n=62) of the videos were aimed at patients and their relatives, while 18.4% (n=14) targeted healthcare professionals (Table 1).

The videos were evaluated for quality and reliability using the JAMA benchmark criteria, the GQS, and the mDISCERN score (Figure 2). Two interventional radiologists performed the assessments independently, and interobserver agreement was found to be strong, with a linear weighted Kappa value of 0.920 ± 0.056.

Informative videos demonstrated significantly higher mean GQS and mDISCERN scores compared to other video types, with GQS scores of 2.85 ± 0.145 (range 1–5, median 3) and mDISCERN scores of 2.13 ± 0.161 (range 1–5, median 2). No statistically significant differences were observed in the scores of other video types (Table 2).

When videos were categorized by target audience, those aimed at patients and relatives had significantly higher GQS and mDISCERN scores than those targeting healthcare professionals. The GQS scores were 2.60 ± 0.116 (range 1–5, median 2) for patients and relatives versus 1.86 ± 0.177 (range 1–3, median 2) for healthcare professionals. Similarly, the mDISCERN scores were 1.98 ± 0.121 (range 1–4, median 2) for patients and relatives versus 1.36 ± 0.199 (range 1–3, median 1) for healthcare professionals.

Shorter videos (< 4 min) had significantly lower scores on all three assessment tools compared to longer videos (> 4 min).

No statistically significant correlation was found between the total number of views and the scoring systems. The correlation coefficients for the mDISCERN, JAMA, and GQS scoring systems were R^2 = -0.019, R^2 = 0.147, and R^2 = 0.147, respectively.

There was no statistically significant difference between the groups and popularity metrics such as likes, comments, and view rates (p> 0.05)

DISCUSSION

Numerous studies have investigated the safety and efficacy of percutaneous ablation techniques in the treatment of benign thyroid nodules. These minimally invasive procedures offer shorter recovery times and fewer complications compared to surgical alternatives. With the increasing number of treatment options for thyroid nodules, patients need to have access to accurate educational resources to make informed decisions about their care.

Table 2. Comparison of video quality scores by video characteristics using three validated scoring systems

	JAMA		GQS		DISCERN	
	Mean (Min-Max/Median)	***P* value**	**Mean (Min-Max/Median)**	***P* value**	**Mean (Min-Max/Median)**	***P* value**
Video type						
Informative	1.79 (1-3/2)	0.382	2.85 (1-5/3)	**0.002**	2.13 (1-4/2)	**0.008**
Technique	1.50 (1-2/1.5)		1.83 (1-2/2)		1.17 (1-2/1)	
Personal Experience	1.83 (1-3/2)		2.17 (1-4/2)		1.83 (1-3/2)	
News Update	1 (1-1)		2 (2-2/2)		3 (3-3/3)	
Duration						
0-4 minutes	1.61 (1-3/2)	**0.043**	2.20 (1-4/2)	**0.001**	1.63 (1-4/1)	**0.009**
>4 minutes	2.00 (1-3/2)		2.93 (1-5/3)		2.30 (1-4/2)	
Upload time						
0-6 months	1.88 (1-3/2)	0.816	2.75 (1-5/2.5)	0.806	2.25 (1-4/2)	0.191
6-12 months	1.67 (1-2/2)		2.33 (2-3/2)		1.33 (1-2/1)	
>12 months	1.74 (1-3/2)		2.44 (1-4/2)		1.87 (1-4/2)	
Upload Source						
Independent Users	2.00 (1-3/2)	0.177	2.60 (2-4/2)	0.389	1.70 (1-3/2)	0.940
Professional Health Care Providers	1.76 (1-3/2)		2.50 (1-5/2)		1.91 (1-4/2)	

Social media platforms, especially YouTube, have become primary sources of health information, providing patients with accessible, visual, and engaging content. However, the uncontrolled nature of these platforms raises serious concerns about the quality and reliability of such videos as a source of health information (8).

Several studies analyzing YouTube as a source of health information have highlighted significant variability in video quality and reliability (9,10). Factors such as the number of views, likes, and comments often reflect popularity rather than educational value, making it difficult for patients to identify accurate content.

Studies evaluating the quality and reliability of YouTube videos on various medical topics have shown that videos uploaded by healthcare professionals generally receive higher DISCERN and GQS scores compared to non-professional sources (4,11-14). Specific studies, such as studies on neonatal sepsis and sports mouthguards, further confirm these findings, showing that videos uploaded by professional sources are generally more reliable but also inadequate as stand-alone educational tools (15,16). A study on uterine fibroid embolization found that although some content achieved higher DISCERN scores, the overall quality were average or poor, with no correlation between video popularity and reliability (17).

Analyses of musculoskeletal ultrasound, intragastric balloon procedures, and orthodontic aligners also found that professional videos were higher quality but represented only a small portion of the content. Furthermore, non-professional videos were often less reliable but more visually engaging, leading to higher popularity metrics such as views and likes (14,18).

Contrary to other studies, although the majority of videos in our study were uploaded by professional healthcare providers (71.1%), no significant difference was observed in all scoring metrics compared to non-professional users. Instead, our study highlights the critical role of content type and video duration in

determining quality and reliability, as informative and longer videos scored higher on all metrics. Therefore, healthcare professionals should create longer, more detailed, and engaging content to address gaps in patient education. However, the highly specialized and emerging nature of PATN videos results in a smaller pool of content, emphasizing an even greater disparity between professional and non-professional contributions. As with other studies, there is a disconnect between popularity metrics and content quality, highlighting the challenge of promoting accurate, high-quality information in niche areas. These findings reinforce the need for healthcare professionals to create detailed, accessible, and patient-centered content to address gaps in public education, particularly for new medical procedures such as PATN.

A notable observation in this study is the association between video duration and quality scores. Longer videos provided more comprehensive information, addressing key aspects such as procedure indications, benefits, risks, and alternatives. This pattern aligns with findings from studies on clear aligners and intragastric balloons, where detailed content correlated with higher DISCERN and GQS scores (13,14). However, the general audience's preference for shorter, more engaging content may limit the accessibility and visibility of longer, higher-quality videos.

Additionally, popularity metrics, such as views and likes, were not consistently correlated with video quality. This disconnect raises concerns about the influence of non-credible but visually appealing content on patient decision-making. As seen in evaluations of videos on skin cancer screening and the Zika virus pandemic, misleading or low-quality videos often garner more attention, which can propagate misinformation (4,19).

To address these challenges, healthcare professionals and organizations must create and promote high-quality educational content on platforms like YouTube. Incorporating elements such as engaging visuals, clear messaging, and comprehensive explanations can help improve the quality and reach of such content. Furthermore, collaboration with platforms to integrate quality indicators or certifications for health-related videos could guide users toward more reliable sources.

This study has several limitations. First, only English-language videos were included, which may introduce language bias and limit the generalizability of the findings to non-English-speaking populations. Second, the cross-sectional nature of the study, based on a single-day search, may not capture temporal changes in video content, popularity metrics, or search engine algorithms. As YouTube content is dynamic and constantly evolving, the results may not reflect long-term trends or newly uploaded videos. Third, while validated tools such as the mDISCERN, GQS, and JAMA scores were used to assess video quality and reliability, the study did not evaluate the factual accuracy or clinical correctness of the information presented in the videos. Therefore, a video could receive a high-quality score based on structure and presentation while still containing misleading or incorrect medical information. Future research should incorporate content validation by subject-matter experts as well as multilingual and longitudinal analysis to enhance the robustness of findings.

CONCLUSION

While YouTube offers significant opportunities for patients to learn about PATN, the variability in quality and reliability highlights the need for greater professional engagement and platform oversight. By adding verification systems, professional endorsements, and quality-focused algorithms, YouTube can serve as a more credible source of patient education. These efforts are essential to support informed patient decision-making and to promote a better understanding of emerging procedures such as PATN.

Ethical approval

This study has been approved by the Ethics Committee of Bolu Abant Izzet Baysal University (approval date 19.11.2024, number 2024/291). Written informed consent was obtained from the participants.

Author contribution

Concept: SG; Design: SG; Data Collection or Processing: ST, ZNB; Analysis or Interpretation: YY, HÖ; Literature

Search: SG, HÖ; Writing: SG, YY. All authors reviewed the results and approved the final version of the article.

Source of funding

The authors declare the study received no funding.

Conflict of interest

The authors declare that there is no conflict of interest.

REFERENCES

1. Che Y, Jin S, Shi C, et al. Treatment of Benign Thyroid Nodules: Comparison of Surgery with Radiofrequency Ablation. AJNR Am J Neuroradiol. 2015; 36(7): 1321-5. [Crossref]
2. Madathil KC, Rivera-Rodriguez AJ, Greenstein JS, Gramopadhye AK. Healthcare information on YouTube: A systematic review. Health Informatics J. 2015; 21(3): 173-94. [Crossref]
3. Syed-Abdul S, Fernandez-Luque L, Jian WS, et al. Misleading health-related information promoted through video-based social media: anorexia on YouTube. J Med Internet Res. 2013; 15(2): e30. [Crossref]
4. Reinhardt L, Steeb T, Mifka A, Berking C, Meier F, German Skin Cancer Council. Quality, Understandability and Reliability of YouTube Videos on Skin Cancer Screening. J Cancer Educ. 2023; 38(5): 1667-74. [Crossref]
5. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health. 1999; 53(2): 105-11. [Crossref]
6. Légaré F, Ratté S, Stacey D, et al. Interventions for improving the adoption of shared decision making by healthcare professionals. Cochrane Database Syst Rev. 2010; (5): CD006732. [Crossref]
7. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the Internet: Caveant lector et viewor-Let the reader and viewer beware. JAMA. 1997; 277(15): 1244-5.
8. Osman W, Mohamed F, Elhassan M, Shoufan A. Is YouTube a reliable source of health-related information? A systematic review. BMC Med Educ. 2022; 22(1): 382. [Crossref]
9. Drozd B, Couvillon E, Suarez A. Medical YouTube Videos and Methods of Evaluation: Literature Review. JMIR Med Educ. 2018; 4(1): e3. [Crossref]
10. Basch CH, Hillyer GC, MacDonald ZL, Reeves R, Basch CE. Characteristics of YouTube™ Videos Related to Mammography. J Cancer Educ. 2015; 30(4): 699-703. [Crossref]
11. Çapar SH, Şenkaya D, Ertürk P, Çetin Kara H. Quality and Reliability of YouTube Videos in the Dix-Hallpike Test. European Archives of Medical Research. 2023; 39(3): 153-8. [Crossref]
12. Kara M, Ozduran E, Mercan Kara M, Hanci V, Erkin Y. Assessing the quality and reliability of YouTube videos as a source of information on inflammatory back pain. PeerJ. 2024; 12: e17215. [Crossref]
13. Calisir A, Ece I. Assessment of the Quality and Reliability of Intragastric Balloon Videos on YouTube. Obes Surg. 2022; 32(4): 1157-63. [Crossref]
14. Cesur E, Tuncer K, Sevgi D, Balaban BC, Arslan C. Evaluation of the Quality and Reliability of YouTube™ Videos Created by Orthodontists as an Information Source for Clear Aligners. Turk J Orthod. 2024; 37(1): 44-9. [Crossref]
15. Hakyemez Toptan H, Kizildemir A. Quality and Reliability Analysis of YouTube Videos Related to Neonatal Sepsis. Cureus. 2023; 15(5): e38422. [Crossref]
16. Gezer I, Saygili S, Gunver MG, Kasimoglu Y, Tuna-Ince EB. Quality and Reliability of YouTube Video Contents About Sports Mouthguards: A Cross-Sectional Study. Dent Traumatol. 2025; 41(1): 118-27. [Crossref]
17. Gad B, Shanmugasundaram S, Kumar A, Shukla P. Quality and Reliability of YouTube Videos on Uterine Fibroid Embolization. J Am Coll Radiol. 2022; 19(7): 905-12. [Crossref]
18. Cüzdan N, Türk İ. Evaluation of quality and reliability of musculoskeletal ultrasound videos on YouTube. Mod Rheumatol. 2022; 32(5): 999-1005. [Crossref]
19. Bora K, Das D, Barman B, Borah P. Are internet videos useful sources of information during global public health emergencies? A case study of YouTube videos during the 2015-16 Zika virus pandemic. Pathog Glob Health. 2018; 112(6): 320-8. [Crossref]

NORTHWESTERN MEDICAL JOURNAL

Northwestern Med J. 2025;5(4):255-258
DOI: 10.54307/NWMJ.2025.187

RESEARCH ARTICLE

Azygos lobe retrospective evaluation

Suat Konuk[1], **Emine Özsarı**[1]

[1]Department of Chest Diseases, Faculty of Medicine, Abant İzzet Baysal University, Bolu, Türkiye

Cite as: Konuk S, Özsarı E. Azygos lobe retrospective evaluation. Northwestern Med J. 2025;5(4):255-258.

ABSTRACT

Aim: The azygos lobe is a well-known anatomical variation of the lung, formed due to the aberrant course of the azygos vein during fetal development. The incidence of the azygos lobe has been reported to be between 0.4% and 1%. This study aims to identify the prevalence of the azygos lobe in a specific patient population, classify its variations, and highlight its importance in differential diagnosis and associated pathologies.

Methods: A retrospective analysis was conducted on 4,549 patients who visited our pulmonology clinic between April 2014 and November 2017. The presence of an azygos lobe was identified through radiological screening.

Results: Among the 4,549 patients screened, 14 cases (0.31%) of azygos lobe were detected, comprising 10 males and four females. The azygos lobe was classified based on the course of the azygos fissure: Type A (lateral to the lung apex) in seven cases (50%), Type B (midline and more vertical) in three cases (21%), and Type C (medial, extending into the mediastinum) in four cases (29%). Additionally, 12 cases (85.7%) were located on the right side, while two cases (14.3%) were on the left.

Conclusions: The azygos lobe is most commonly detected incidentally during imaging for unrelated conditions. Our study emphasizes the importance of recognizing the azygos lobe as a differential diagnosis and investigating potential coexisting pathologies. The findings contribute to existing statistical data on the prevalence and anatomical variations of the azygos lobe.

Keywords: azygos lobe, anatomical variation, differential diagnosis, pulmonary imaging

Corresponding author: Suat Konuk **E-mail:** suatkonukk@windowslive.com
Received: 20.03.2025 **Accepted:** 25.04.2025 **Published:** 26.10.2025

INTRODUCTION

In normal human anatomy; The right lung is divided into three lobes—lobus superior, medius and inferior—by two fissures, and the left lung is divided into two lobes—lobus superior and inferior—by a single fissure (1). The lungs are divided into lobes by the combination of segments and the oblique and horizontal fissures (2). The most well-known accessory fissure is the azygos fissure. It occurs as a result of the azygos vein passing in front of the lung during the intrauterine period, and the apical or posterior segment of the right upper lobe remaining behind this vein. The azygos vein develops from the posterior cardinal vein. As a result of a part of the upper lobe remaining below and medial to this vein, an accessory lobe resembling a separate lobe is formed due to the pressure of the vein, which is called the azygos lobe. It has been reported to occur at a rate of 0.4-1% during routine examinations (3-5). It is usually asymptomatic and does not require treatment. The aim of this study is to contribute to the literature on the azygos lobe and to draw attention to the diseases that coexist with it.

MATERIALS AND METHODS

Among the 4,549 people who came to the Düzce private pulmonology clinic between April 2014 and November 2017, those with an azygos lobe were detected through retrospective screening. Statistical evaluation of the study was performed using the SPSS 16.0 package and p values <0.05 were considered significant.

Retrospective radiological evaluation was performed with posteroanterior (PA) radiographs in patients older than 18 years. Individuals over the age of 18 who presented to the chest diseases clinic for any reason were included in the study. Those without a chest X-ray were excluded from the study.

RESULTS

Among the 4,549 people who came for chest diseases examination between April 2014 and November 2017, those with an azygos lobe were detected through retrospective screening. Of these, 2,409 were male and 2,140 were female. An azygos lobe was detected in 10 of the 2,409 male patients and four of the 2,140 female patients.

The appearance of the azygos lobe is classified into three types, depending on the course of the azygos fissure in relation to the lung apex. The trigone is classified as Type A if it is lateral to the apex of the lung, Type B if it is located towards the middle and the fissure is more vertical, and Type C if it is medial and the fissure extends into the mediastinum. In our case series, there were seven patients (50%) with Type A, three patients (21%) with Type B, and four patients (29%) with Type C.

In our study, the azygos lobe was settled on the right side in 12 patients (85.7%) and on the left side in two patients (14.3%). Figure 1 shows the azygos lobe and fissure on the PA chest radiograph, and Figure 2

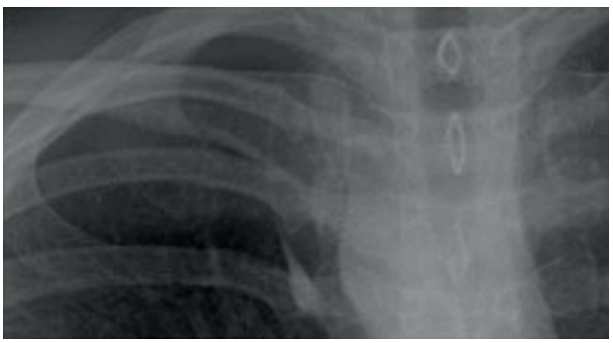


Figure 1. Azygos lobe and fissure on the PA chest radiograph.

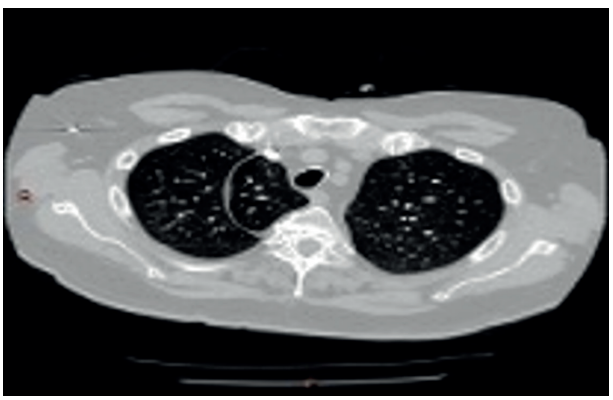


Figure 2. Azygos lobe and fissure on the thorax CT.

shows them on the thorax computed tomography (CT). Sequestration was observed on one patient's chest X-ray, and one patient had coexisting lung cancer, which was treated. The average age of the individuals included in the study was 48.6±3.5 years. It is important to identify diseases concomitant with the azygos lobe, as these conditions should not be overlooked.

DISCUSSION

The azygos lobe is a rare anatomical variation that can cause significant morphological changes in the upper mediastinum (6). It occurs as a result of inadequate migration of the azygos vein from the thoracic wall to its normal position, the tracheobronchial corner. The azygos lobe may be located behind the superior vena cava and trachea, and may contact the medial wall of the descending aorta and esophagus. It is generally seen twice as often in men (7,8). In our study, 10 out of 14 patients (71%) were male. The azygos lobe is often located on the right (9,10). In our study, 12 cases (85.7%) were on the right and two cases (14.3%) were on the left.

The azygos lobe is seen in 1% of cadaver samples, 0.4% in chest radiography, and 1.2% in high-resolution computed tomography (3-5,11). Diagnosis is usually made by chest radiography (11). Computed tomography can also be used in selected cases for further examination and differential diagnosis (12). As in our case, on the chest X-ray there is a convex line connected to the azygos fissure, a triangular-shaped area (trigone) connected to the extrapleural tissue on the fissure, and a teardrop appearance connected to the azygos vein at the bottom (13).

The appearance of the azygos lobe depends on the course of the azygos fissure in relation to the lung apex. It is classified in three ways: Type A, when the trigone is lateral to the lung apex; Type B, when it is located toward the middle and the fissure is more vertical; and Type C, when it is medial and the fissure extends into the mediastinum (14).

Although it does not represent a pathological condition on its own in practice, the azygos lobe requires attention due to possible accompanying pathological events. Although rare, the azygos lobe may be accompanied by pathologies such as fissures, tumors (small cell lung cancer), extrapulmonary sequestration, pneumothorax, bullous changes, vascular anomalies, and situs inversus totalis (15). In our study, one patient had extralobar sequestration, and another had lung cancer and was treated accordingly. Monaco et al. reported surgical treatment of a case in which they detected an azygos lobe with pneumothorax (16). The most important thing to consider in surgical interventions for lesions located in the azygos lobe is the risk of damaging the azygos vein while dissecting the azygos lobe (17). Gentle caudal traction of the azygos lobe is important in preventing this complication. The bronchus of the lobe usually branches from the upper lobe apical segment bronchus, but it has also been observed to arise from the main bronchus. Therefore, these anatomical variations should be taken into consideration during bronchial division (18). The azygos lobe is usually identified incidentally during imaging performed for other reasons. Accurate identification is important as the azygos lobe can mimic other lung pathologies and may cause technical challenges during surgery (19). In our study, no such findings were observed.

CONCLUSION

In most cases, the azygos lobe is detected incidentally during examinations for other reasons, as in our study. Our findings reveal that the azygos lobe should not be overlooked in differential diagnoses and emphasize the importance of investigating additional pathologies.

Ethical approval

This study has been approved by the Sakarya University Non-Interventional Research Ethics Committee (approval date 09.02.2017, number 71522473/050.10.04/33). Written informed consent was obtained from the participants.

Author contribution

Concept: SK; Design: SK; Data Collection or Processing: SK; Analysis or Interpretation: SK, EÖ; Literature Search: SK; Writing: SK, EÖ. All authors reviewed the results and approved the final version of the article.

Source of funding

The authors declare the study received no funding.

Conflict of interest

The authors declare that there is no conflict of interest.

REFERENCES

1. Kılıç C, Kocabıyık N, Yalçın B, Kırıcı Y, Yazar F, Ozan H. Accessory fissures of the lungs. Med J SDU. 2009; 13(3): 12-6.
2. Meenakshi S, Manjunath KY, Balasubramanyam V. Morphological variations of the lung fissures and lobes. Indian J Chest Dis Allied Sci. 2004; 46(3): 179-82.
3. BOYDEN EA. The distribution of bronchi in gross anomalies of the right upper lobe particularly lobes subdivided by the azygos vein and those containing pre-eparterial bronchi. Radiology. 1952; 58(6): 797-807. **[Crossref]**
4. Ariyürek OM, Gülsün M, Demirkazik FB. Accessory fissures of the lung: evaluation by high-resolution computed tomography. Eur Radiol. 2001; 11(12): 2449-53. **[Crossref]**
5. Aziz A, Ashizawa K, Nagaoki K, Hayashi K. High resolution CT anatomy of the pulmonary fissures. J Thorac Imaging. 2004; 19(3): 186-91. **[Crossref]**
6. Etter LE. Variations in the position of the azygos septum and its incidence in 50,000 roentgen examinations. Am J Roentgenol Radium Ther. 1947; 58(6): 726-9.
7. Postmus PE, Kerstjens JM, Breed A, vd Jagt E. A family with lobus venae azygos. Chest. 1986; 90(2): 298-9. **[Crossref]**
8. Fisher MS. Re: Adam's lobe. Radiology. 1985; 154(2): 547. **[Crossref]**
9. Takasugi JE, Godwin JD. Left azygos lobe. Radiology. 1989; 171(1): 133-4. **[Crossref]**
10. Kobayashi T, Satoh K, Kawase Y, et al. Bronchial and vascular supply in azygos lobe of inflated fixed lung: a case study. Radiat Med. 1995; 13(1): 31-3.
11. Mata J, Cáceres J, Alegret X, Coscojuela P, De Marcos JA. Imaging of the azygos lobe: normal anatomy and variations. AJR Am J Roentgenol. 1991; 156(5): 931-7. **[Crossref]**
12. Kauffman P, Wolosker N, de Campos JRM, Yazbek G, Jatene FB. Azygos lobe: a difficulty in video-assisted thoracic sympathectomy. Ann Thorac Surg. 2010; 89(6): e57-9. **[Crossref]**
13. Karre PR, Cooper GB. The azygos lobe and vein: interesting and typical clinical image. BMJ Case Rep. 2011; 2011: bcr0520114266. **[Crossref]**
14. Stibbe EP. The Accessory Pulmonary Lobe of the Vena Azygos. J Anat. 1919; 53(Pt 4): 305-14.
15. Gürkök S, Gözübüyük A, Yücel O, Çaylak H, Dakak M. Bullous azygos lobe: case report. Turk Gogus Kalp Damar Cerrahisi Derg. 2007; 15(2): 168-9.
16. Monaco M, Barone M, Barresi P, Carditello A, Pavia R, Mondello B. Azygos lobe and spontaneous pneumothorax. G Chir. 2000; 21(11-12): 457-8.
17. Fukuhara S, Montgomery M, Reyes A. Robot-assisted azygos lobectomy for adenocarcinoma arising in an azygos lobe. Interact Cardiovasc Thorac Surg. 2013; 16(5): 715-7. **[Crossref]**
18. Sadikot RT, Cowen ME, Arnold AG. Spontaneous pneumothorax in a patient with an azygos lobe. Thorax. 1997; 52(6): 579-80; discussion 575-6. **[Crossref]**
19. Caceres J, Mata JM, Andreu J. The azygos lobe: normal variants that may simulate disease. Eur J Radiol. 1998; 27(1): 15-20. **[Crossref]**

NORTHWESTERN MEDICAL JOURNAL

Northwestern Med J. 2025;5(4):259-267
DOI: 10.54307/NWMJ.2025.165

REVIEW

The impacts of artificial intelligence in healthcare

Elif Başer[1], **Nazlı Gülsoy Kırnap**[2]

[1]Nesibe Aydın Yıldızlar Science High School, Ankara, Türkiye
[2]Department of Internal Medicine, Arel University, İstanbul, Türkiye

Cite as: Başer E, Gülsoy Kırnap N. The impacts of artificial intelligence in healthcare. Northwestern Med J. 2025;5(4):259-267.

ABSTRACT

Artificial intelligence (AI) is a system that aims to bring human thinking ability to machines and thus solve complex tasks more easily and can continuously improve itself with the data it collects. AI technologies are used in many fields, such as education, media, banking, and the defense industry. In recent years, AI has begun to manifest itself in various fields of health services, from diagnosis to treatment and patient follow-up.

AI algorithms can facilitate patients' access to healthcare services, provide remote patient monitoring, shorten the diagnosis process of diseases, provide patient-specific treatment recommendations, or allow physicians to improve their practices. Thus, it can improve patient care and increase patient satisfaction, reduce costs, and speed up healthcare services. Increasing telemedicine applications during the pandemic contributed to the acceleration of the utilization of AI in healthcare, and AI-based algorithms for diagnosing and treating diseases began to be developed rapidly. With the increasing prevalence of wearable technologies and the introduction of electronic health records, there has been a tremendous explosion in individual health data. AI has contributed to health services in collecting and processing this rapidly increasing data.

Although AI has a promising future in health, it also brings many ethical problems. AI systems make decisions based on the data they are trained on. If there is not enough data diversity or if biased data is used, AI systems may give inaccurate predictions or learn and reproduce these preconceptions.

This manuscript is a literature review examining the impacts of AI in the healthcare sector and discusses the history of AI, AI studies applied for the diagnosis and treatment of diseases, telemedicine and preventive medicine applications, and the disadvantages of AI.

Keywords: artificial intelligence, diseases, healthcare, telemedicine

Corresponding author: Nazlı Gülsoy Kırnap **E-mail:** kirnapnazli@hotmail.com
Received: 04.12.2024 **Accepted:** 02.06.2025 **Published:** 26.10.2025

INTRODUCTION

In recent years, technological developments such as big data systems, the Internet of Things (IoT), health monitoring devices, telemedicine applications, and robotic surgeries have been used in various fields of the healthcare sector and have started to provide convenience to patients and physicians in protecting health and accelerating diagnosis and treatment processes. Thanks to big data systems, more data can be accessed by using health records, laboratory and imaging results, and information such as medical records, test results, imaging findings, and prescriptions of patients can be stored within the health cloud. IoT enables medical devices and sensors to be connected to the internet in order to collect and share data in real-time. Through IoT, communication between wearable technologies, sensors, medical devices, and other resources can be established, and information sharing can be provided (1). Mobile devices such as smartphones and tablets can be beneficial to the user by enabling patients to monitor their health status, use medication reminders, or meet with their doctors via telemedicine.

Artificial intelligence (AI) is the system used to solve tasks such as problem-solving, data analysis, and learning through computers by giving human-like thinking abilities to machines. AI includes many tools, such as deep learning, neural networks, and machine learning. An algorithm is a step-by-step approach designed to achieve a specific goal. AI advancements in the health industry aim to produce solutions to problems by using a diverse range of algorithms, such as deep learning, natural language processing, support vector machines, ensemble learning, and bioinformatics algorithms.

AI is used to speed up processes, increase efficiency, and reduce costs in numerous areas such as diagnosis, treatment, and follow-up of diseases, vaccine and pharmaceutical industry, making or reminding doctor appointments, and telemedicine with machine learning or deep learning technologies. This review discusses the areas of use of AI technologies in healthcare.

History of artificial intelligence

The emergence of AI dates back to the "Turing Test" in the 1950s. For the first time, with Alan Turing's question, "Can a man-made device move and make decisions in a way that is indistinguishable from that of humans?", AI was conceptualized, and with the Dartmouth Conference in 1956, AI became a topic of discussion (2,3). ELIZA, developed by Joseph Weizenbaum at the Massachusetts Institute of Technology (MIT) AI laboratory between 1964 and 1967, was one of the first chatbots and the first program to test the Turing Test. ELIZA could organize text responses that simulated a dialogue with a human therapist (4). Shakey, developed between 1966 and 1972, was the first robot capable of interpreting human instructions and performing actions based on these instructions. It was considered one of the greatest achievements of that period.

In the 1970s, AI started to be used in medical applications, and the first artificial medical advisor, INTERNIST-1, emerged (5). In this system, search algorithms were used to find a diagnosis from patients' symptoms. During these years, MYCIN was developed by Stanford University to identify bacteria giving rise to severe infections and to recommend appropriate doses of antibiotics to patients (6). In the 1980s, DXplain was designed at the Massachusetts General Hospital computer science laboratory to make it easier for doctors to diagnose (7). It was similar to INTERNIST-1 but had the ability to derive more diagnoses from the symptoms entered.

In the 2000s, AI applications in health gained momentum. IBM Watson was developed to respond to questions on a quiz show. It beat the best contestants to take first place. Watson was built as a question-answering computing system using natural language processing, information retrieval, knowledge representation, automatic reasoning, and machine learning technologies (8). In 2017, the Watson system identified RNA-binding proteins associated with amyotrophic lateral sclerosis (9).

Telemedicine

Telemedicine refers to delivering healthcare services across distances using information and communication technologies (10). Telemedicine applications can be used in "real time," involving active communication between doctors and patients via telephone or video calls, and through "store and forward" methods that enable clinicians to store and transmit data for later evaluation and interpretation. It facilitates access to healthcare, especially where access to clinicians is limited. Remote Patient Monitoring (RPM) is a subset of telemedicine, enabling healthcare professionals to monitor, investigate, and report on patient conditions remotely.

There has been an increase in the use of telemedicine during the COVID-19 outbreak, and it has provided significant benefits in reducing the spread of COVID-19, monitoring isolated patients, and identifying patients in need of hospital care (11). AI-equipped smartphones have enabled remote assessment of a patient's likelihood of being infected by analyzing the sound of speaking or coughing, and AI-enabled chatbots have contributed to reducing the increased triage burden by classifying patients according to their symptoms, helping to refer patients to home quarantine, the emergency department, or a COVID-19 clinic (12).

Integrating AI into telemedicine has contributed to the clinician's decision-making process by providing alternatives for interpreting medical images. AI methods in smartphones or wearable devices have been used to increase patients' treatment compliance, providing opportunities such as reminding patients of the time of medication intake and monitoring heart rhythm with smartwatches (13).

ChatGPT, an AI language model, launched by OpenAI, functions as an AI-powered chatbot that is able to comprehend natural language conversations and answer user queries. A ChatGPT-powered chatbot provides information about medical diseases and answers patients' questions about clinical features, prescription drugs, and therapeutic procedures. For medical staff, it summarizes patients' medical information and can help them perform RPM to maintain patient health. ChatGPT can offer solutions as a computer program that aids patients in managing their treatment, akin to a virtual assistant alerting them to adhere to their medical prescriptions and providing information about their current health condition. Digital platforms, including websites, voice assistants, and mobile applications, can be used to reach these virtual assistants. Nevertheless, ChatGPT in healthcare has shortcomings regarding medical ethics, privacy, security, consent, and liability.

Diagnosis of diseases

By processing biochemical and clinical test data with big data analysis methods, AI has become available in the diagnosis and treatment process of many diseases.

With the onset of digitalization, the interpretation of images by computers has been the subject of intense interest; thus, computer-aided detection (CADe) and computer-aided diagnosis (CADx) algorithms have emerged. Although there were initially high expectations from these systems, they were not very effective in the clinical field due to high false positive rates. AI has been increasingly used by radiologists for the early diagnosis of various diseases and the reduction of diagnostic errors. It has been reported that AI-supported systems perform at least as well as radiologists, show lower false diagnosis rates than radiologists, and reduce the workload of radiologists (14). Similarly, AI is a tool that cardiologists can use to analyze electrocardiography and echocardiography graphics. The Ultromics platform, founded at a hospital in Oxford, has used AI to analyze echocardiography scans to detect heart rate patterns and ischemic heart disease (15).

The first application of AI in the surgical field was by Gunn in 1976 when he tried to diagnose acute abdominal pain using computer analysis, and AI made an accurate diagnosis (16).

An AI-assisted diagnostic system was developed to classify malignant and benign lung nodules based on computed tomography (CT) data, and it was shown that this system can be applied to differentiate lung nodules accurately (17). In this study, Gong J et al. included a total of 243 pulmonary nodules, of which 76 were benign and 167 were malignant (81 were stage

1 and 86 were stage 3) according to biopsy results, and as a result of the study, they reported that a CADx scheme was feasible to discriminate between benign and malignant lung nodules accurately, and that there was a positive correlation between CADx performance and cancer stages (17).

Gudigar et al. declared that various medical imaging tools such as X-ray, CT, and ultrasound (US) using AI techniques have contributed substantially to the fight against the pandemic by aiding in the early diagnosis of COVID-19 (18). AI algorithms applied to lung CT scans of COVID-19 patients showed that pneumonia developed most frequently in the right lobe of the lung (19).

A study by Wang et al. proposed a new hybrid chest CT-based method to automatically detect COVID-19, using a diagnostic technique based on wavelet Renyi entropy (WRE) and a three-segment biogeography-grounded optimization (3SBBO) algorithm, which outperforms many AI algorithms (20).

AI has also started to be used in breast cancer detection. In related studies, it has been shown that the performance of AI methods developed in breast cancer detection is similar to human performance, and it has been revealed that AI may play an essential role in the field of breast cancer screening in the future (21). In one study, the interpretations made by 101 radiologists (28,296 independent interpretations) on 2,652 digital mammography exams (653 malignant) collected from seven countries were compared with the interpretations of the AI system, and it was shown that the performance of the AI system was not statistically inferior to the performance of radiologists. The AI system evaluated in this study was reported to achieve cancer detection accuracy similar to that of an average breast radiologist (14).

A study has shown that an AI-based method can classify thyroid nodules as benign or malignant. In this study, image features were extracted from thyroid US images in the spatial domain using deep learning, and in the frequency domain using Fast Fourier Transform (FFT). A stepwise classification scheme was created using these features, and it was found that thyroid nodules could be successfully classified (22).

Magnetic Resonance Imaging (MRI) is able to detect inflammation in the joints of patients with rheumatoid arthritis (RA) before it manifests clinically and can play a crucial role in early diagnosis. By collecting MRI data from healthy controls and patients with suspected progression to RA, specific patterns predicting the development of RA can be deduced. However, manually classifying a great deal of MRI data is impractical. Consequently, research has been conducted to detect early RA from MRI data using AI techniques, and an AI algorithm based on MRI data has been shown to increase the rate of early RA detection compared to specialized physicians (23).

In ophthalmology, AI is applied to fundus photographs, optical coherence tomography, and visual fields to detect diabetic retinopathy, macular edema, and macular degeneration. It is suggested that deep learning in ocular imaging can be combined with telemedicine to screen, diagnose, and monitor important eye diseases in primary care (24).

AI methods have also significantly progressed in interpreting small bowel capsule endoscopy images. In a study in which 7,556 small bowel capsule endoscopy images were analyzed with an AI neural network algorithm, lesions and their localization were detected with higher sensitivity and accuracy than the conventional model (25).

AI tools can analyze speech patterns to predict psychotic events and to recognize and screen for features of neurological diseases such as Parkinson's disease (26). Furthermore, AI algorithms may be helpful in the diagnosis of autism spectrum disorder by analyzing transcripts of spontaneous speech and assessing behavioral features or in the diagnosis of acute appendicitis in children by analyzing blood tests (27,28).

The use of AI in medical pathology can help in lesion detection, classification, and prognosis prediction. Kosaraju et al. developed a new deep-learning AI model called Deep-Hipo for pathological image analysis and reported that this model performed more efficiently and accurately than other AI methods in analyzing gastric cancer pathological images. Although there are several convolutional neural

network-based techniques to precisely analyze histopathological images for cancer detection, cancer subtype classification, and risk prediction, most of them perform patch-based examinations because of the large size of histopathological images. Still, these patches do not consist of adequate information or patterns. Deep-Hipo, on the other hand, was reported to extract two patches of the same size at both high and low magnification levels and to capture complex morphological patterns in both large and small receptive fields of the whole slide image, thus outperforming the latest deep learning methods (29).

Treatment of diseases

IBM Watson, a well-known AI system in the world, is designed to help physicians examine electronic health records of their patients, access academic publications related to diseases, and examine documents related to the disease with features including machine learning, data mining, and natural language processing. A study conducted in India found that the treatment recommendations provided by the Watson oncology system to 638 patients with breast cancer had 90% similarity to the treatments recommended by the tumor detection board working in this field, and saved a significant amount of time. The researchers used both the manual method and the Watson system to collect and analyze the data and then compared the time taken to provide treatment recommendations. They found that the manual approach took an average of 20 minutes, which decreased to 12 minutes as physicians became more familiar with it, while Watson took only 40 seconds to analyze the data and provide a treatment recommendation (30).

Today, robots are used in various surgical procedures to assist surgeons. The Da Vinci robotic surgery system is one of the systems often preferred by doctors in this context. The Da Vinci robotic surgery AI system has made surgical treatment more minimally invasive with the advantages of clearer image, more precise and comfortable operation, and even remote operation. With these advancements, it has provided high surgical success and low complication rates (31).

Three-Dimensional Printing (3DP) is a rapid prototyping technology using AI technology. Based on digital model files from CT or MRI data, it creates objects by printing layer by layer with powdered metal or other adhesive biomaterials with AI technology. Clinical imaging data are transferred to intelligent software. After artificially choosing regions of interest, the software outputs three-dimensional reconstructions. This method makes it possible to obtain an exact model of the injured part reconstructed from real CT scan data, make detailed plans before surgery, and perform a pre-simulated surgery on the model. After 3DP printing and sterilization, implants can be surgically implanted in place of human tissue injured for various reasons (32). Feng ZH et al. reported that pedicle screw placement with 3DP template guidance in spine surgery is both safe and easier than conventional methods, and the risks of neurovascular damage and radiation exposure are reduced (33).

Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR) technologies are novel types of digital holographic imaging technologies akin to 3DP. VR is a virtual digital image created by an intelligent computer algorithm that allows surgeons to practice and improve their surgical skills on a virtual system without making severe surgical errors. It cannot be applied in actual surgery due to the paucity of real-world experience. AR technology can provide a virtual reconstruction of the critical region, and by adding the virtual image to the actual image, it can recognize complex anatomical structures and guide the surgeon before or during surgery (34). MR combines virtuality and reality with real-time interaction and exact matching. With a wearable MR device or a device such as Hololens, the surgeon can immerse himself in a more complex surgical world, create a better treatment program, and develop robust doctor-patient communication (35).

AI also has original applications in rehabilitation. Physically and socially supportive robots can be utilized in the rehabilitation of individuals. Smart mobile and wearable devices can be used to gather information to assess the achievement of personalized rehabilitation goals (36). Sensors in wearable technologies can be used to explore whether individuals perform their exercise regimes appropriately (37). Although its long-

term effectiveness has not yet been proven, some studies indicate that patients with musculoskeletal dysfunction can be treated with simple mobilization using skilled or sensitive robotic hands (38).

Preventive medicine

Especially in many low- and middle-income countries, obstacles to routine vaccination activities are encountered. Moreover, quarantines imposed during the COVID-19 pandemic further disrupted routine vaccination services, affected people's demand for vaccines, and increased the risk of outbreaks of vaccine-preventable diseases (39). Studies show that vaccination records may contain errors. Some critical information, such as batch numbers on vaccine vials, is missing. Additionally, many vaccination-related data are entered manually by healthcare personnel (40). In recent years, developments in mobile technologies have effectively eliminated gaps in vaccination programs. Scanning vaccine vials with barcoding tools can help reduce data drift and transcription errors by uploading information directly to Internet Information Services (IIS). By scanning patient identification barcodes, it may be possible to provide data entries for demographic information and previous immunization records from electronic medical record software. AI can also help increase vaccination rates through features such as vaccination reminders and promotional campaigns (40). AI programs have effectively monitored COVID-19 cases and vaccine distributions during the pandemic, ensuring fair vaccine distribution (41).

As in the rest of the world, the population over 65 years of age is increasing gradually in Türkiye, and it is estimated to increase to 12.9% of the entire population in 2030 and 22.6% of the entire population in 2060 (42). As life expectancy increases over time, the number of patients with various chronic diseases such as Alzheimer's and dementia, or those who live alone at home and need external care, is rising. In recent years, the use of robots in elderly care has reduced caregivers' burden, and the quality of life of elderly individuals has increased with mobile applications and smart homes. Thanks to social robots, it has been observed that the daily life activities of individuals with dementia increase and their stress decreases (43). AI can be helpful in the follow-up of chronic diseases like diabetes, hypertension, sleep apnea, and chronic bronchial asthma in elderly individuals by using wearable, non-invasive sensors (44).

Disadvantages of AI in healthcare

Although AI has many advantages in the health sector, making decisions instead of humans in diagnosis and treatment and the implementation of these decisions bring some risks. Moreover, the digitalization of diagnosis and treatment processes may cause an increase in job anxiety among people working in this field. Being less physically and cognitively active due to technological dependence may lead to the emergence of psychological or musculoskeletal problems or a decrease in self-esteem because of the social isolation that technological dependence may result in. Feelings of dissatisfaction due to being examined and treated in a digital environment may also occur in patients.

Large datasets are required for AI systems to categorize or predict various tasks appropriately, and this data may be personal or sensitive. Since patient records are confidential in the health sector, there may be problems in terms of data accessibility, confidentiality, and security. The decision-making process of AI relies on the data that is provided. Consequently, AI systems that lack variety or contain flawed data may eventually lead to improper results being deduced and reproduced. Therefore, it might not be appropriate to accept the decisions of AI systems completely. When AI works on insufficient or erroneous data, it may provide incorrect diagnosis or treatment recommendations. In this case, it is uncertain whether this responsibility lies with the technology company that designed or implemented the erroneous algorithm or with the clinician who gave the final decision.

Furthermore, technological innovations may be more expensive and lack the empathy and creativity that humans have; they may produce different results across races and genders (45,46). A study by Obermeyer et al. found racial bias in a commonly

used commercial algorithm, highlighting that Black patients were clinically more ill compared to their White counterparts for a given projected risk score. This prejudice was due to the algorithm's prediction of healthcare costs rather than disease, implying that less budget was spent on Black patients than on White patients (47).

Moreover, there is a significant difference between the accuracy, precision, and abilities of AI systems. To illustrate, the progress from GPT-3 to GPT-4 is one of the best examples of groundbreaking improvements in AI's capability to comprehend and process complex data. To clarify this point, GPT-4 passed the bar exam and was in the top 10% of all examinees, whereas GPT-3.5 was in the bottom 10%. The differences in results are due to GPT-4's capability to manage an extensive range of data types, including photographs and written content, as well as the advantage of training on 45 gigabytes of data in contrast to 17 gigabytes of data for GPT-3. This improvement in performance makes GPT-4 more costly to implement than GPT-3 (48). Hence, when AI is utilized in underserved communities, it will be crucial not to be tempted to use cheaper and less effective forms of AI that may exacerbate health inequalities in these communities.

In conclusion, AI technologies are used in numerous fields such as medical imaging, diagnosis, treatment, pandemic response, patient care, preventive medicine, and telemedicine. Nevertheless, as AI advances in healthcare, it brings with it various technical and ethical challenges. AI cannot replace human connections and cooperation, even though it operates reasonably efficiently. Human functions like teamwork and leadership are not likely to be achieved, as machines are not able to form connections with humans. Machine learning's ability to transform data into insights could impact the medical field by taking over much of the work of radiologists and pathologists. However, clinical medicine requires evaluating vast amounts of data—from physical exams to laboratory results, imaging studies to genetic data—and the ability to manage this complexity puts doctors ahead of machines. In the future, it is necessary to prioritize ethical frameworks, and the technical and social aspects of AI to ensure its effective integration into healthcare. The goal is for AI technologies to comply with patient-centered care and ethical procedures with the contribution of clinicians, researchers, managers, and ethicists. Research aimed at improving the fairness, transparency, and reliability of AI is essential to both reduce bias and provide equitable healthcare.

Author contribution

Literature review: EB, Preparation and Writing: EB, Analysis or Interpretation: NGK. All authors reviewed the results and approved the final version of the article.

Source of funding

The authors declare the study received no funding.

Conflict of interest

The authors declare that there is no conflict of interest.

REFERENCES

1. Aceto G, Persico V, Pescapé A. Industry 4.0 and Health: Internet of Things, Big Data, and Cloud Computing for Healthcare 4.0. Journal of Industrial Information Integration. 2020; 18: 100129. [Crossref]
2. Shieber SM. The Turing Test as Interactive Proof. Nous. 2007; 41(4): 686-713. [Crossref]
3. Cordeschi R. AI's half century: on the thresholds of the Dartmouth Conference. Retrospettiva. 2006; 3(1-2): 109-14.
4. Weizenbaum J. ELIZA-A computer program for the study of natural language communication between man and machine. Communication of the ACM. 1983; 26(1): 23-8. [Crossref]
5. Miller RA, Pople HE, Myers JD. Internist-1, an experimental computer-based diagnostic consultant for general internal medicine. N Engl J Med. 1982; 307(8): 468-76. [Crossref]
6. Shortliffe EH. Mycin: A Knowledge-Based Computer Program Applied to Infectious Diseases. Proc Annu Symp Comput Appl Med Care. 1977: 66-9.
7. DXplain-The Laboratory of Computer Science. Available at: http://www.mghlcs.org/projects/dxplain
8. Ferrucci D, Levas A, Bagchi S, Gondek D, Mueller ET. Watson: Beyond Jeopardy! Artificial Intelligence. 2013; 199-200: 93-105. [Crossref]
9. Bakkar N, Kovalik T, Lorenzini I, et al. Artificial intelligence in neurodegenerative disease research: use of IBM Watson to identify additional RNA-binding proteins altered in amyotrophic lateral sclerosis. Acta Neuropathol. 2018; 135(2): 227-47. [Crossref]

10. Ryu S. Telemedicine: Opportunities and Developments in Member States: Report on the Second Global Survey on eHealth 2009 (Global Observatory for eHealth Series, Volume 2). Healthcare Informatics Research. 2012: 18(2): 153-5. [Crossref]

11. Mann DM, Chen J, Chunara R, Testa PA, Nov O. COVID-19 transforms health care through telemedicine: Evidence from the field. J Am Med Inform Assoc. 2020; 27(7): 1132-5. [Crossref]

12. Chamola V, Hassija V, Gupta V, Guizani M. A Comprehensive Review of the COVID-19 Pandemic and the Role of IoT, Drones, AI, Blockchain, and 5G in Managing its Impact. IEEE Access. 2020; 8: 90225-65. [Crossref]

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

14. Rodriguez-Ruiz A, Lång K, Gubern-Merida A, et al. Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists. J Natl Cancer Inst. 2019; 111(9): 916-22. [Crossref]

15. Ghosh P. AI Early Diagnosis Could Save Heart and Cancer Patients. Science Correspondent, BBC News. 2018. Available at: https://www.bbc.com/news/health-42357257

16. Gunn AA. The diagnosis of acute abdominal pain with computer analysis. J R Coll Surg Edinb. 1976; 21(3): 170-2.

17. Gong J, Liu JY, Sun XW, Zheng B, Nie SD. Computer-aided diagnosis of lung cancer: the effect of training data sets on classification accuracy of lung nodules. Phys Med Biol. 2018; 63(3): 035036. [Crossref]

18. Gudigar A, Raghavendra U, Nayak S, et al. Role of Artificial Intelligence in COVID-19 Detection. Sensors (Basel). 2021; 21(23): 8045. [Crossref]

19. Zhang HT, Zhang JS, Zhang HH, et al. Automated detection and quantification of COVID-19 pneumonia: CT imaging analysis by a deep learning-based software. Eur J Nucl Med Mol Imaging. 2020; 47(11): 2525-32. [Crossref]

20. Wang SH, Wu X. Zhang YD, Tang C, Zhang X. Diagnosis of COVID-19 by Wavelet Renyi Entropy and Three-Segment Biogeography-Based Optimization. Int J Comput Intell Syst. 2020; 13: 1332-44. [Crossref]

21. Sechopoulos I, Teuwen J, Mann R. Artificial intelligence for breast cancer detection in mammography and digital breast tomosynthesis: State of the art. Semin Cancer Biol. 2021; 72: 214-25. [Crossref]

22. Nguyen DT, Pham TD, Batchuluun G, Yoon HS, Park KR. Artificial Intelligence-Based Thyroid Nodule Classification Using Information from Spatial and Frequency Domains. J Clin Med. 2019; 8(11): 1976. [Crossref]

23. Stoel BC. Artificial intelligence in detecting early RA. Semin Arthritis Rheum. 2019; 49(3S): S25-8. [Crossref]

24. Ting DSW, Pasquale LR, Peng L, et al. Artificial intelligence and deep learning in ophthalmology. Br J Ophthalmol. 2019; 103(2): 167-75. [Crossref]

25. Hwang Y, Lee HH, Park C, et al. Improved classification and localization approach to small bowel capsule endoscopy using convolutional neural network. Dig Endosc. 2021; 33(4): 598-607. [Crossref]

26. Bedi G, Carrillo F, Cecchi GA, et al. Automated analysis of free speech predicts psychosis onset in high-risk youths. NPJ Schizophr. 2015; 1: 15030. [Crossref]

27. Küpper C, Stroth S, Wolff N, et al. Identifying predictive features of autism spectrum disorders in a clinical sample of adolescents and adults using machine learning. Sci Rep. 2020; 10(1): 4805. [Crossref]

28. Aydin E, Türkmen İU, Namli G, et al. A novel and simple machine learning algorithm for preoperative diagnosis of acute appendicitis in children. Pediatr Surg Int. 2020; 36(6): 735-42. [Crossref]

29. Kosaraju SC, Hao J, Koh HM, Kang M. Deep-Hipo: Multi-scale receptive field deep learning for histopathological image analysis. Methods. 2020; 179: 3-13. [Crossref]

30. Somashekhar SP, Kumarc R, Rauthan A, Arun KR, Patil P, Ramya YE. Double blinded validation study to assess performance of IBM artificial intelligence platform, Watson for oncology in comparison with Manipal multidisciplinary tumour board - First study of 638 breast cancer cases. Cancer Research. 2017; 77(Suppl 4): S6-07. [Crossref]

31. Tae K. Transoral robotic thyroidectomy using the da Vinci single-port surgical system. Gland Surg. 2020; 9(3): 614-6. [Crossref]

32. Tejo-Otero A, Buj-Corral I, Fenollosa-Artés F. 3D Printing in Medicine for Preoperative Surgical Planning: A Review. Ann Biomed Eng. 2020; 48(2): 536-55. [Crossref]

33. Feng ZH, Li XB, Phan K, et al. Design of a 3D navigation template to guide the screw trajectory in spine: a step-by-step approach using Mimics and 3-Matic software. J Spine Surg. 2018; 4(3): 645-53. [Crossref]

34. Creighton FX, Unberath M, Song T, Zhao Z, Armand M, Carey J. Early Feasibility Studies of Augmented Reality Navigation for Lateral Skull Base Surgery. Otol Neurotol. 2020; 41(7): 883-8. [Crossref]

35. Wu X, Liu R, Yu J, et al. Mixed Reality Technology Launches in Orthopedic Surgery for Comprehensive Preoperative Management of Complicated Cervical Fractures. Surg Innov. 2018; 25(4): 421-2. [Crossref]

36. Luxton DD, Riek LD. Artificial intelligence and robotics in rehabilitation. In: Brenner LA, Reid-Arndt SA, Elliott TR, Frank RG, Caplan B, editors. Handbook of Rehabilitation Psychology. Washington, DC, USA: American Psychological Association; 2019: 507-20. [Crossref]

37. Goldzweig CL, Orshansky G, Paige NM, et al. Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. Ann Intern Med. 2013; 159(10): 677-87. [Crossref]

38. Choi C, Schwarting W, DelPreto J, Rus D. Learning Object Grasping for Soft Robot Hands. IEEE Robotics and Automation Letters. 2018; 3(3): 2370-7. [Crossref]

39. Chandir S, Siddiqi DA, Setayesh H, Khan AJ. Impact of COVID-19 lockdown on routine immunisation in Karachi, Pakistan. Lancet Glob Health. 2020; 8(9): e1118-20. **[Crossref]**

40. Atkinson KM, Mithani SS, Bell C, Rubens-Augustson T, Wilson K. The digital immunization system of the future: imagining a patient-centric, interoperable immunization information system. Ther Adv Vaccines Immunother. 2020; 8: 2515135520967203. **[Crossref]**

41. White BM, Shaban-Nejad A. Utilization of Digital Health Dashboards in Improving COVID-19 Vaccination Uptake, Accounting for Health Inequities. Studies in Health Technology and Informatics. 2022; 295: 499-502. **[Crossref]**

42. Turkish Statistical Institute (TURKSTAT). İstatistiklerle Yaşlılar, 2022. March 17, 2023. Available at: https://data.tuik.gov.tr/Bulten/Index?p=Istatistiklerle-Yaslilar-2022-49667

43. Kachouie R, Sedighadeli S, Khosla R, Chu MT. Socially Assistive Robots in Elderly Care: A Mixed-Method Systematic Literature Review. International Journal of Human–Computer Interaction. 2014; 30(5): 369-93. **[Crossref]**

44. Kim J, Campbell AS, Wang J. Wearable non-invasive epidermal glucose sensors: A review. Talanta. 2018; 177: 163-70. **[Crossref]**

45. Braun M, Hummel P, Beck S, Dabrock P. Primer on an ethics of AI-based decision support systems in the clinic. J Med Ethics. 2020; 47(12): medethics-2019-105860. **[Crossref]**

46. Grote T, Berens P. On the ethics of algorithmic decision-making in healthcare. J Med Ethics. 2020; 46(3): 205-11. **[Crossref]**

47. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019; 366(6464): 447-53. **[Crossref]**

48. OpenAI; Achiam J, Adler S, Agarwal S, et al. GPT-4 Technical Report 2024. arXiv 2024, arXiv:2303.08774. **[Crossref]**

NORTHWESTERN MEDICAL JOURNAL

Northwestern Med J. 2025;5(4):268-273
DOI: 10.54307/NWMJ.2025.154

CASE REPORT

A case of placenta previa and placenta increta undergoing uterine sparing surgery

Elif Betül Esmer[1], **Ayşe Kübra Topal Aköz**[2], **Ebru Kaynar Varlık**[2], **Mustafa Ayhan Ekici**[2]

[1]Department of Obstetrics and Gynecology, İzzet Baysal State Hospital, Bolu, Türkiye
[2]Department of Obstetrics and Gynecology, Faculty of Medicine, Abant İzzet Baysal University, Bolu, Türkiye

Cite as: Esmer EB, Topal Aköz AK, Kaynar Varlık E, Ekici MA. A case of placenta previa and placenta increta undergoing uterine sparing surgery. Northwestern Med J. 2025;5(4):268-273.

ABSTRACT

Obstetric risks are increased in pregnancies following previous cesarean section deliveries. Placenta increta and placenta previa pregnancies are two important clinical manifestations of abnormal placentation entities. Placenta accreta occurs in 5-10% of pregnancies complicated by placenta previa. The most important risk factor for placenta accreta is a history of uterine surgery. A thin, defectively formed or absent decidua basalis layer cannot show resistance to deep penetration of trophoblasts and placenta invades this pathological region of uterus that has previously been traumatised by uterine surgery. Placenta accreta is diagnosed with characteristic signs during prenatal ultrasound examination. The first clinical sign of placenta accreta is excessive and life-threatening bleeding that occurs during manual placental separation. Prenatal diagnosis is important to avoid this complication, which may lead to maternal mortality. In this case, in which an invasion anomaly was detected in preoperative ultrasonography, the treatment of the patient accompanied by placenta increta and placenta praevia with uterus-preserving surgery was discussed.

Keywords: placenta increta, placenta previa, uterine sparing surgery, clinical approach

INTRODUCTION

Placenta previa refers to the presence of placental tissue extending above the internal cervical os.

Several clinical features increase the risk of placenta previa. Most importantly, women with one or more previous cesarean deliveries are at greater risk for subsequent placental disorders such as placenta previa, detachment, or placental invasion (1).

When placenta previa is diagnosed, the possibility of placenta accreta spectrum (PAS) should be considered. This is particularly important if the placenta is located above the previous hysterotomy site and is therefore most common with anterior placenta previa.

Two primary mechanisms are proposed to explain the development of placental invasion anomalies. The first involves damage to the decidua basalis, which may occur after uterine procedures such as curettage,

Corresponding author: Elif Betül Esmer **E-mail:** betuloneren@gmail.com
Received: 07.10.2024 **Accepted:** 03.10.2025 **Published:** 26.10.2025

cesarean delivery, or myomectomy. When this layer loses its protective barrier function, trophoblastic cells may directly invade the myometrium. The second mechanism has been associated with several molecular and genetic alterations, including increased CD44 receptor expression, dysregulation of angiogenic, and trophoblastic growth factors, changes in adrenomedullin gene expression at the cytokine level, and mitochondrial DNA mutations (2).

Placental invasion anomalies accompanying placenta previa are generally categorized into three subtypes based on the depth of trophoblastic penetration: accreta, increta, and percreta. In the accreta type, chorionic villi attach to the myometrium without invading it; in increta, they extend into the myometrial layer; and in percreta, villi penetrate through the myometrium and may reach the uterine serosa or adjacent organs, such as the bladder. Collectively, these entities are described under the term placenta accreta spectrum (PAS) or morbidly adherent placenta (3,4).

Reported risk factors for both placenta previa and PAS include advanced maternal age, multiparity, low socioeconomic background, previous infertility treatment, prior cesarean delivery, history of myomectomy or curettage, and maternal smoking (4,5). These anomalies are known to increase maternal morbidity through complications such as severe antepartum or intrapartum hemorrhage, need for cesarean hysterectomy, massive transfusion, or intensive care admission, and they may also contribute to preterm birth. In recent decades, while uterine atony used to be the leading indication for peripartum hysterectomy, PAS has now become the most frequent cause (3,4,6)

In placental invasion anomalies, the mode of delivery is a planned cesarean section in the early week, and the need for peripartum bleeding and hysterectomy increases due to increased myometrial invasion due to advanced gestational age. This prevents fertility-sparing surgery. In this case report, uterus-preserving treatment of placenta previa and placenta increta in a 38-week pregnant patient with a history of four previous cesarean sections, was presented in the light of the literature, in whom bleeding control was achieved via bipolar energy device.

CASE REPORT

A 30-year-old 38/0 weeks pregnant patient presented to our clinic with suspicion of amniotic band referred from an external center. In her detailed history, it was learned that the pregnancy information was gravida 6, parity 4, abortion 1. All of her deliveries were performed by cesarean section. The patient reported that her last cesarean section had been complicated, and her physician had informed her that a T-incision was required for uterine entry.

On examination, the placenta was on the anterior wall, the distinction between myometrium and placenta was not clear, and the placenta was observed on the old cesarean scar line. There were large lacunes in the placenta and turbulent blood flow was observed in the lacunes. Fetal biometry was compatible with 38-39 weeks. Transvaginal ultrasonography showed that the cervical os was completely covered by the placenta. It was the patient's first presentation to our hospital. She had not been diagnosed with placenta previa or placental invasion anomaly before. Since the diagnosis of placental invasion anomaly was definite, no further examination (MRI) was required. The patient was diagnosed with placenta previa and placental invasion anomaly. Consent for hysterectomy was obtained and the necessary blood products were prepared and the operation was planned.

During the cesarean section, the uterus was adherent to the anterior abdominal wall and the intestines to the uterine fundus. The adhesions were dissected. A breech female baby weighing 3890 grams was delivered through Kerr incision. Placenta was observed as placenta increta invading the myometrium and placenta previa was observed to close the cervical opening (Figure 1). With the assistance of a vessel-

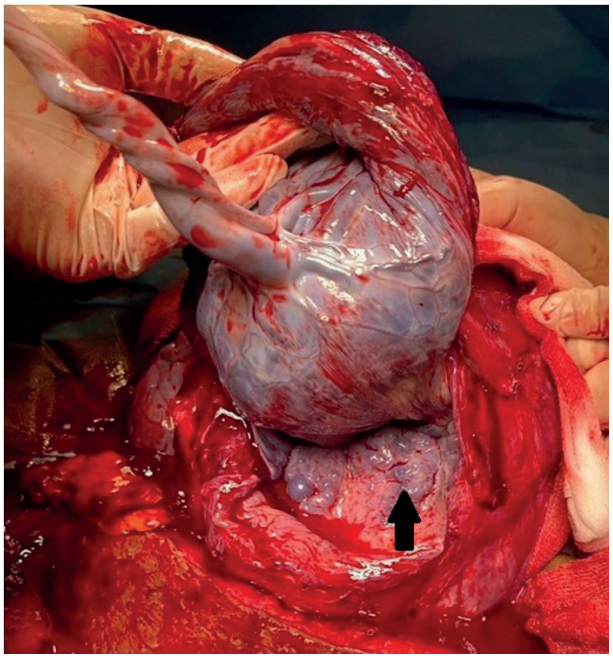


Figure 1. Intraoperative findings showing placenta previa completely covering the cervical os and placenta increta invading the myometrium (arrow sign).

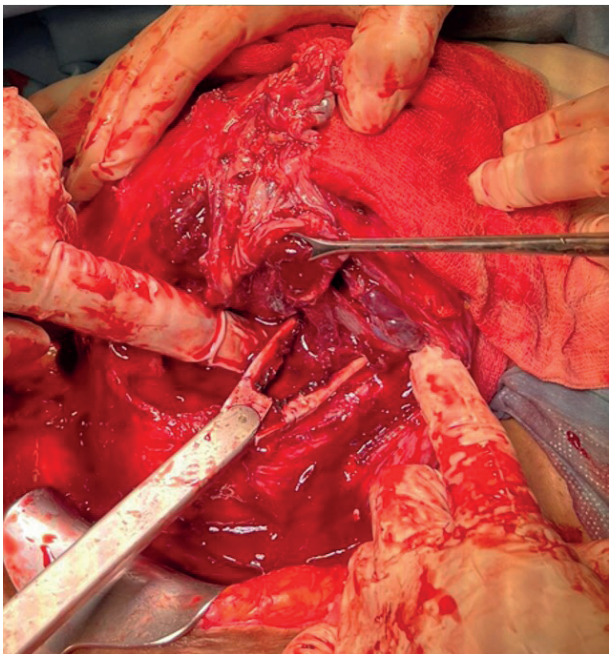


Figure 2. Separation of the placenta, which has invaded ½ of the myometrium, from the uterus by controlling bleeding with the help of bipolar energy device.

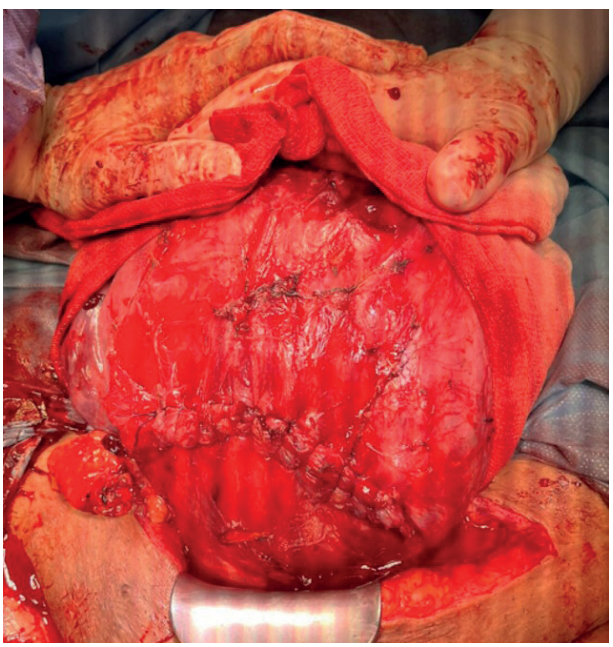


Figure 3. Intraoperative view of the Kerr incision line of the uterus, which is closed after bleeding control is achieved.

sealing device utilizing bipolar energy, the placenta was excised from the cavity via myometrial segment resection (Figure 2). The uterus was then repaired with continuous suturing (Figure 3). Salpingectomy was also performed. The abdominal folds were also closed according to the anatomy plan. The placenta was sent to pathology. Postpartum bleeding was controlled. No complications developed.

During the postoperative service follow-ups, the patient's vital signs were stable, the wound was dressed, there was no evidence for postpartum hemorrhage or need for blood product replacement. Intensive care unit admission was not required, and the postoperative course was uneventful. She was discharged with full recovery on the 3rd postoperative day. No additional problems were observed in the postoperative 10th day and postoperative 40th day controls. The patient started menstruating in the 5th month after birth. Her menstruation have remained regular. Prior to the inclusion of clinical details in this report, written informed consent was obtained from the patient.

DISCUSSION

Placental invasion anomalies, such as placenta increta, are rare but serious complications of pregnancy, with an incidence of one in 533 pregnancies (6). Preoperative management of placental invasion anomalies aims to minimize bleeding and reduce maternal morbidity and mortality through meticulous planning and patient counseling. Although it may be possible to stop bleeding with intraoperative medical and surgical treatments in cases of placenta acreta, cesarean hysterectomy may be inevitable, especially in cases of placenta increta and percreata, since the placenta is usually not detached (7-9). If the placenta is left in place, postpartum hemorrhage secondary to subinvolution may occur.

Cesarean delivery is preferred in cases with placenta accreta. Elective planning of the timing of delivery and being prepared for possible complications are very important in terms of maternal morbidity and mortality. It requires multidisciplinary teamwork including experienced obstetricians, anesthesiologists, urologists, and blood bank. Elective cesarean delivery is recommended at 34 to 35 weeks of gestation to reduce the likelihood of emergency cesarean delivery and other possible complications. In this case report, the patient underwent cesarean section at the 38th week because placental invasion anomaly was detected at the 38th week. This probably caused the invasion to progress from accreta to increta.

Various methods have been described for the management of placenta accreta (10,11). Treatment options ranges from cesarean hysterectomy, uterine preservation, and in some cases, adjuvant treatment with methotrexate, or more conservative treatment methods including leaving the placenta in situ or simply waiting for spontaneous resorption of the placenta. The conservative method was first described as follows: Arulkumaran et al. in 1986, systemic methotrexate was administered postnatally and the placental mass was expelled 11 days postnatally (12). Since then, a number of cases have been reported that were treated conservatively (13). Palacios Jaraquemada et al. described surgical methods of uterine preservation, which were successfully performed in 50 of 68 cases, and ten of the 42 successful cases followed up over a three-year period subsequently became pregnant and had uneventful cesarean sections (14).

The uterus can be preserved if the bleeding that may occur due to partial adhesion of the placenta to the uterus causing bleeding after placenta removal in cesarean section is controlled through medical and/or surgical treatments. Medical treatment (uterotonics, topical hemostatic agents), interventional radiologic treatments, mechanical methods such as balloon tamponade, and surgical interventions to stop bleeding can be performed (15).

Conservative surgery can be performed in selected cases where fertility is desired to be preserved. In this case, the uterus is preserved and fertilization continues by removing the entire placenta and the invaded myometrium without hysterectomy (16). In women whose placenta was left in the uterus, placental resorption occurred at an average of 13.5 weeks (between 4-60 weeks) (17). Undetectable levels of hCG in the blood do not guarantee complete resorption of the placental tissue left in situ. In our case, since all of the placental tissue was removed, postoperative hCG monitoring was not performed and methotrexate was not necessary. The reasons for the complete excision of placental tissue were that myometrial invasion was ½, the invasion was shallower in the area where the placenta covered the cervical os, and that intact myometrial tissue would remain after local excision.

A very common consequence of placental invasion is massive bleeding and the need for cesarean hysterectomy. Maternal morbidity and mortality also increase due to the risk of injury to the urinary tract and other abdominal organs.

To avoid postpartum hemorrhage and hysterectomy, bilateral hypogastric artery ligation, uterine artery ligation, and balloon tamponade placement are among the preferred methods (18). In our case, unlike the literature, bipolar energy device was used during the local excision procedure, thus providing bleeding control.

Hysterectomy may adversely influence women's overall well-being, affecting both physiological and psychological aspects of health (19,20). The uterus

represents a key component of feminine identity, contributing to perceptions of bodily integrity, self-worth, sexuality, and emotional stability. Consequently, surgical removal of the uterus can trigger significant emotional distress, manifesting as anxiety, stress, or depressive symptoms (19).

In the present case, although the 30-year-old patient did not desire future fertility, uterine preservation was prioritized to maintain her bodily integrity, sexual health, and psychological well-being. At the patient's request, salpingectomy was performed as part of the conservative surgical management (21).

The success of uterine sparing surgery and hysterectomy in placental invasion anomalies is summarized in the literature and classified in Table 1.

Uterine-sparing surgery and hysterectomy each offer distinct advantages and limitations in the management of placenta invasion anomalies. Uterine-sparing procedures provide the critical benefit of preserving both fertility and menstrual function, which is particularly valuable for women who wish to maintain their reproductive potential (10,22). These techniques also confer psychological benefits related to the retention of the uterus. However, they are not without risks, including the potential for infection, secondary postpartum hemorrhage, and the need for additional surgical interventions or extended follow-up (22,23). Conversely, hysterectomy represents a definitive treatment option, typically associated with a lower risk of recurrent bleeding and reduced likelihood of further surgical procedures (23,24). Nonetheless, it inevitably results in permanent infertility and may be accompanied by greater surgical morbidity, particularly in cases involving extensive placental invasion such as placenta percreta (23,24). As such, uterine-sparing strategies may be more appropriate for hemodynamically stable patients with focal placental invasion who desire future fertility, while hysterectomy is generally recommended for patients with severe invasion, uncontrolled bleeding, or completed childbearing (24). In conclusion, this case report demonstrates that uterine-sparing surgery can be a feasible option in carefully selected cases of placenta increta, emphasizing the importance of early diagnosis, multidisciplinary planning, and meticulous surgical technique.

Ethical approval

Written informed consent was obtained from the patient for the publication of the attached images.

Author contribution

Surgical and Medical Practices: EBE, AKTA, EKV; Concept: EBE, AKTA, EKV, MAE; Data Collection or Processing: EBE, AKTA, EKV, MAE; Analysis or Interpretation: EBE, AKTA, EKV, MAE; Literature Search: EBE, AKTA, EKV, MAE; Writing: EBE, AKTA, EKV. All authors reviewed the results and approved the final version of the article.

Source of funding

The authors declare the study received no funding.

Conflict of interest

The authors declare that there is no conflict of interest.

REFERENCES

1. Cunningham FG, Leveno KJ, Bloom SL, Dashe JS, Hoffman BL,Casey BM, Spong CY. Wiiliams Obstetrics. 2020: 775.
2. Rao KP, Belogolovkin V, Yankowitz J, Spinnato JA. Abnormal placentation: evidence-based diagnosis and management of placenta previa, placenta accreta, and vasa previa. Obstet Gynecol Surv. 2012; 67(8): 503-19. **[Crossref]**
3. Silver RM, Branch DW. Placenta Accreta Spectrum. N Engl J Med. 2018; 378(16): 1529-36. **[Crossref]**
4. Kale İ, Uluutku GB, Ergin B. Placenta previa and invasion anomalies; a retrospective analysis of cases in a tertiarty obstetric care center. The Journal of Gynecology - Obstetrics and Neonatology. 2021; 18(1): 662-70. **[Crossref]**
5. Rosenberg T, Pariente G, Sergienko R, Wiznitzer A, Sheiner E. Critical analysis of risk factors and outcome of placenta previa. Arch Gynecol Obstet. 2011; 284(1): 47-51. **[Crossref]**
6. Wu S, Kocherginsky M, Hibbard JU. Abnormal placentation: twenty-year analysis. Am J Obstet Gynecol. 2005; 192(5): 1458-61. **[Crossref]**
7. Zelop C, Nadel A, Frigoletto FD, Pauker S, MacMillan M, Benacerraf BR. Placenta accreta/percreta/increta: a cause of elevated maternal serum alpha-fetoprotein. Obstet Gynecol. 1992; 80(4): 693-4.

8. Hung TH, Shau WY, Hsieh CC, Chiu TH, Hsu JJ, Hsieh TT. Risk factors for placenta accreta. Obstet Gynecol. 1999; 93(4): 545-50. **[Crossref]**
9. Zelop CM, Harlow BL, Frigoletto FD, Safon LE, Saltzman DH. Emergency peripartum hysterectomy. Am J Obstet Gynecol. 1993; 168(5): 1443-8. **[Crossref]**
10. Kayem G, Davy C, Goffinet F, Thomas C, Clément D, Cabrol D. Conservative versus extirpative management in cases of placenta accreta. Obstet Gynecol. 2004; 104(3): 531-6. **[Crossref]**
11. Oyelese Y, Smulian JC. Placenta previa, placenta accreta, and vasa previa. Obstet Gynecol. 2006; 107(4): 927-41. **[Crossref]**
12. Arulkumaran S, Ng CS, Ingemarsson I, Ratnam SS. Medical treatment of placenta accreta with methotrexate. Acta Obstet Gynecol Scand. 1986; 65(3): 285-6. **[Crossref]**
13. Adair SR, Elamin D, Tharmaratnam S. Placenta increta; conservative management -- a successful outcome. Case report and literature review. J Matern Fetal Neonatal Med. 2004; 15(4): 275-8. **[Crossref]**
14. Palacios Jaraquemada JM, Pesaresi M, Nassif JC, Hermosid S. Anterior placenta percreta: surgical approach, hemostasis and uterine repair. Acta Obstet Gynecol Scand. 2004; 83(8): 738-44. **[Crossref]**
15. American College of Obstetricians and Gynecologists. ACOG Practice Bulletin: Clinical Management Guidelines for Obstetrician-Gynecologists Number 76, October 2006: postpartum hemorrhage. Obstet Gynecol. 2006; 108(4): 1039-47. **[Crossref]**
16. Garmi G, Goldman S, Shalev E, Salim R. The effects of decidual injury on the invasion potential of trophoblastic cells. Obstet Gynecol. 2011; 117(1): 55-9. **[Crossref]**
17. Sentilhes L, Ambroselli C, Kayem G, et al. Maternal outcome after conservative treatment of placenta accreta. Obstet Gynecol. 2010; 115(3): 526-34. **[Crossref]**
18. Karateke A, Küçükbaş M, Sozen H, Namazov A, Cakır S, Akdemir Y. Fertility sparing surgery on placenta invasion anomalies and placenta previa. Iran J Reprod Med. 2012; 10(3): 271-4.
19. Goudarzi F, Khadivzadeh T, Ebadi A, Babazadeh R. Women's interdependence after hysterectomy: a qualitative study based on Roy adaptation model. BMC Womens Health. 2022; 22(1): 40. **[Crossref]**
20. Grover B, Einerson BD, Keenan KD, et al. Patient-reported health outcomes and quality of life after peripartum hysterectomy for placenta accreta spectrum. Am J Perinatol. 2022; 39(3): 281-7. **[Crossref]**
21. Kaba M. Invasion anomaly or neovascularization? A new surgical approach for cesarean delivery in pregnant women with invasive placenta accreta spectrum disorder accompanied by placenta previa. Ginekol Pol. 2024; 95(2): 114-22. **[Crossref]**
22. Timmermans S, van Hof AC, Duvekot JJ. Conservative management of abnormally invasive placentation. Obstet Gynecol Surv. 2007; 62(8): 529-39. **[Crossref]**
23. Eller AG, Bennett MA, Sharshiner M, et al. Maternal morbidity in cases of placenta accreta managed by a multidisciplinary care team compared with standard obstetric care. Obstet Gynecol. 2011; 117(2 Pt 1): 331-7. **[Crossref]**
24. Jauniaux E, Ayres-de-Campos D, Langhoff-Roos J, Fox KA, Collins S, FIGO Placenta Accreta Diagnosis and Management Expert Consensus Panel. FIGO classification for the clinical diagnosis of placenta accreta spectrum disorders. Int J Gynaecol Obstet. 2019; 146(1): 20-4. **[Crossref]**

NORTHWESTERN MEDICAL JOURNAL

Northwestern Med J. 2025;5(4):274-280
DOI: 10.54307/NWMJ.2025.155

CASE REPORT

Acute cervical intervertebral disc herniation after traffic accident: the importance of rapid diagnosis with magnetic resonance imaging in the emergency department

Baycan Kuş[1], Gözde Yılmaz[1], Fatma Hançer Çelik[1], Necmi Baykan[1], Merve İrem Atıcı[1]

[1]Department of Emergency Medicine, Kayseri City Hospital, Kayseri, Türkiye

The study has been presented at The International Emergency Medicine Academic Congress in Konya in 2024.

Cite as: Kuş B, Yılmaz G, Hançer Çelik F, Baykan N, Atıcı Mİ. Acute cervical intervertebral disc herniation after traffic accident: the importance of rapid diagnosis with magnetic resonance imaging in the emergency department. Northwestern Med J. 2025;5(4):274-280.

ABSTRACT

The cervical vertebrae (C1-C7) are the most commonly injured region of the spine because they are the most flexible, mobile, and exposed part of the spinal column. Different studies have demonstrated that cervical spine injuries are more common in motor vehicle accidents. An acute intervertebral disc herniation may develop due to hyperextension of the cervical region in trauma-related conditions such as motor vehicle accidents. Rapid and definitive diagnosis is possible with magnetic resonance imaging (MRI). In acute cervical spine injury, MRI is the gold standard for the surgeon to decide on issues such as the need for surgery, the type of approach, the need for instrumentation, the level of decompression, and the need for re-surgery in case of insufficient decompression, as well as the presence of edema or hematoma.

In this case report, a patient with plegia is presented who had an acute intervertebral disc herniation without cervical spine fracture due to hyperextension of the cervical region during a motor vehicle accident and whose diagnosis was possible with early consultation and MRI. Computed tomography and X-ray imaging were normal. The patient's clinical status could not have been adequately elucidated without the performance of MRI. As a matter of fact, according to the spinal cord trauma guidelines, neurological recovery is faster and more favorable in cases operated on within 24 hours, as observed in our case.

This article was written to emphasize that relying solely on Computed tomography or X-ray may be insufficient, and that MRI should not be avoided when clinically indicated. To identify and monitor the presence and underlying causes of spinal cord hematoma, edema, and stenosis, MRI is required.

Keywords: Cervical spine injury, emergency department, intervertebral disc displacement, magnetic resonance imaging, traffic accident

Corresponding author: Baycan Kuş **E-mail:** drcan86@gmail.com
Received: 09.10.2024 **Accepted:** 01.09.2025 **Published:** 26.10.2025

INTRODUCTION

The intervertebral discs function as shock-absorbing structures that facilitate the distribution of axial loads across the spinal column. When the compression forces exceed the absorption capacity of the disc, the annulus fibrosus ruptures, the nucleus pulposus protrudes into the vertebral canal, and as a result, spinal nerve or spinal cord compression develops. Disc herniation occurs, resulting in the sudden onset of neck pain, paraspinal muscle spasm, and radicular arm pain and weakness (1). In a study, it was determined that 23% of the herniations secondary to trauma occurred in the cervical spine. Of these, 23% were in C3–C4, 21% in C4–C5, and 23% in C5–C6 (2).

In this case report, a patient with plegia is presented who had an acute intervertebral disc herniation without cervical spine fracture due to hyperextension of the cervical region during a motor vehicle accident and whose diagnosis was possible with early consultation and magnetic resonance imaging (MRI). Computed tomography (CT) and X-ray imaging were normal. The patient's clinical status could not have been adequately elucidated without the performance of magnetic resonance imaging. The aim of this case report is to highlight the importance, as emphasized in the literature, of rapid MRI diagnosis and surgical intervention within 24 hours in improving the prognosis of patients who develop acute neurological deficits following trauma.

CASE REPORT

The patient was a 54-year-old male with no known disease. He was admitted to the emergency department by ambulance because he had rear-ended a truck at a slow speed while driving under the influence of alcohol. At the time of admission, his blood pressure was 122/81 mmHg, pulse rate was 86/min, SO2: 99%, and temperature was 36.5°C. The total score of the Glasgow Coma Scale (GCS) was 15. He was conscious, cooperative, and oriented. Pupils were isochoric, and the pupillary light reflex was intact bilaterally. Eye movements were exact in all directions. He had no speech disorder. Strength was 4/5 in the right upper limb, 0/5 in the left upper limb, and 5/5 in the lower extremities. There was a sensory deficit in the left upper extremity. There were no other pathological examination findings. The patient persistently stated that he was fine before the accident, but after the accident, he could not move his left arm in particular. CT and X-ray imaging were performed for the pathological findings. Cranial CT, thorax CT, cervical spine CT, and X-ray were normal (Figure 1, Figure 2). There were no pathological findings in the clavicle, scapula, shoulder joints, or upper extremity bones to explain the loss of strength. Diffusion MRI was performed on the assumption that the patient may have had a traffic accident after a cerebrovascular accident. There was no diffusion-restricted area in the brain. The possibility of an acute intervertebral disc herniation occurring due to trauma was considered, and approximately twelve hours after admission to the emergency department, a consultation was requested from the neurosurgery department. A cervical MRI was performed at the suggestion of the consultant and urgently interpreted by the radiology department.

MRI interpretation: Diffuse bulging obliterating the anterior subarachnoid space and causing significant compression of the spinal cord and central broad-based protrusion are observed in the C3-4 disc. The neural foramen is narrowed. At this level, a segment of approximately 25 mm in length in the spinal cord shows a T2 slightly hyperintense faint limited area that may be compatible with myelopathy (Figure 3).

After the diagnosis of the patient was established by MRI, he was quickly re-consulted with the neurosurgery department and underwent emergency surgery. After one month, the motor examination of the patient was 5/5 in the right upper extremity and 4/5 in the left upper extremity; there was no sensory deficit or pathological reflex.

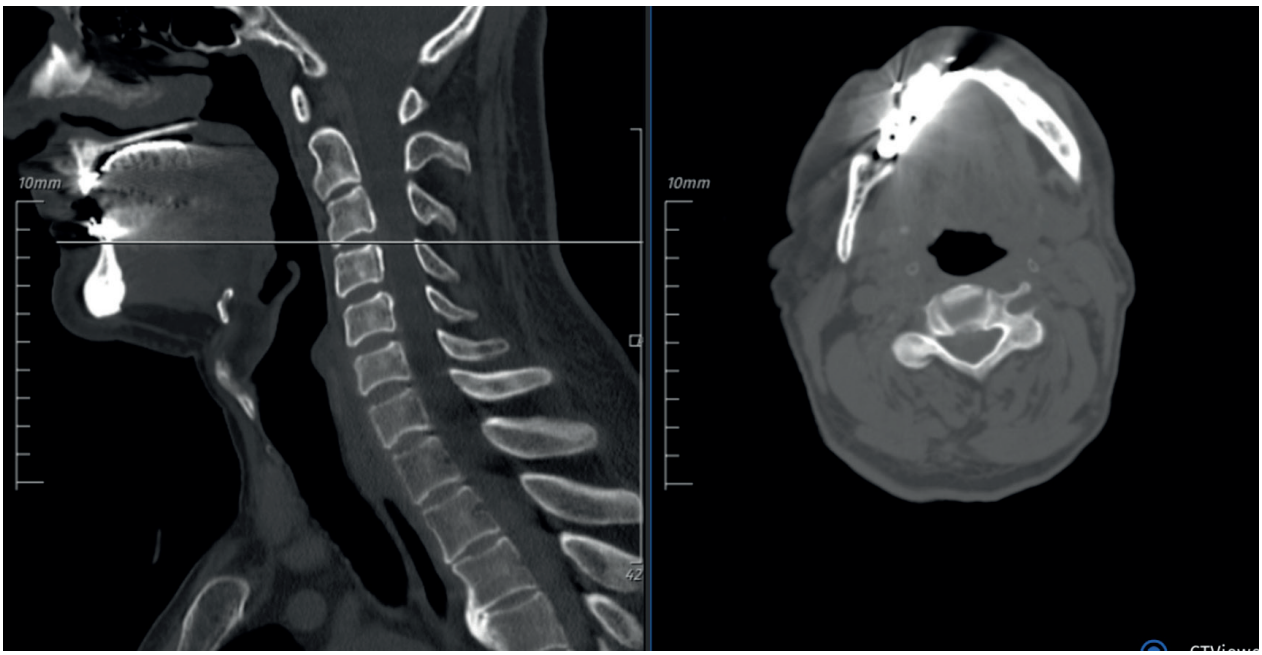


Figure 1. Cervical CT image (Normal).

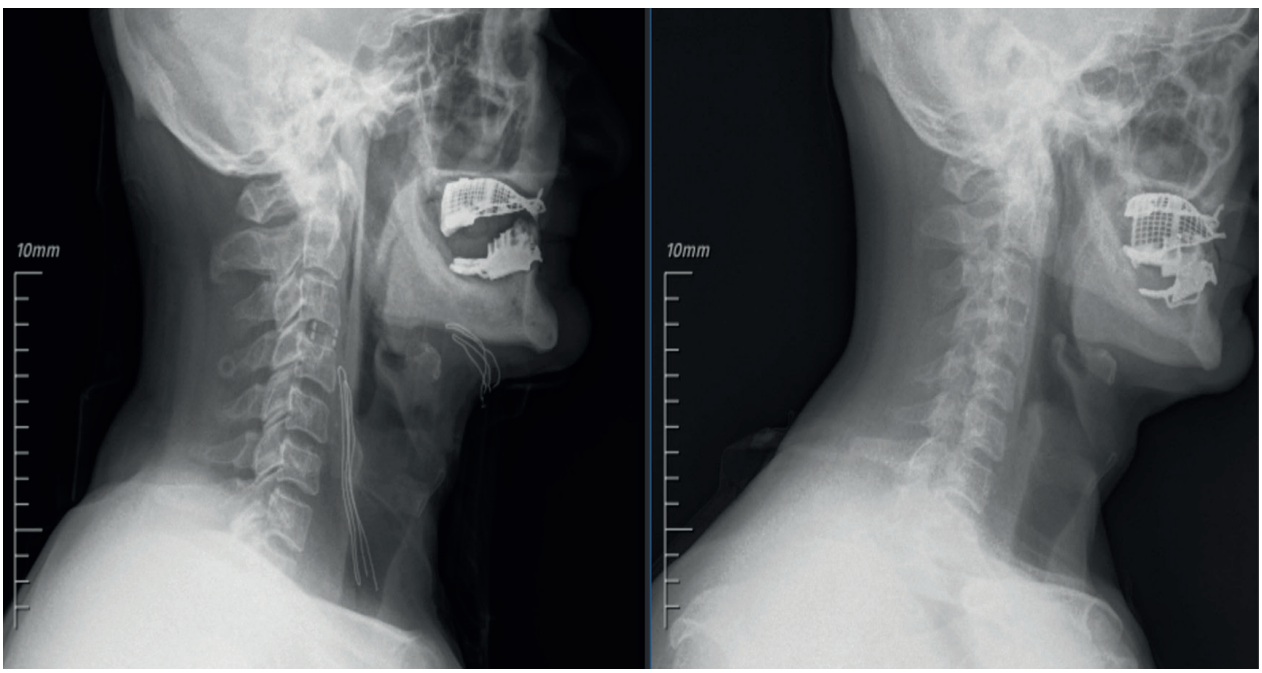


Figure 2. Cervical X-ray images postoperatively (left) and one month later (right) (Normal).

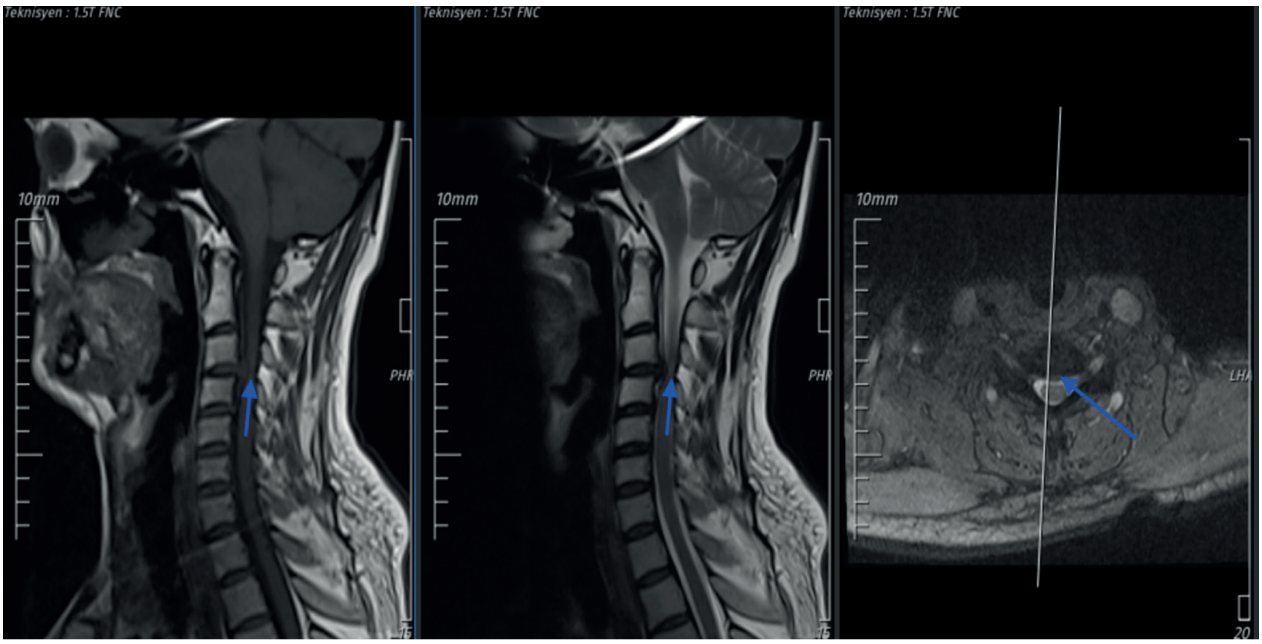


Figure 3. Cervical MRI Image (Diffuse bulging obliterating the anterior subarachnoidal space and causing significant compression of the spinal cord and central broad-based protrusion are observed in the C3-4 disc).

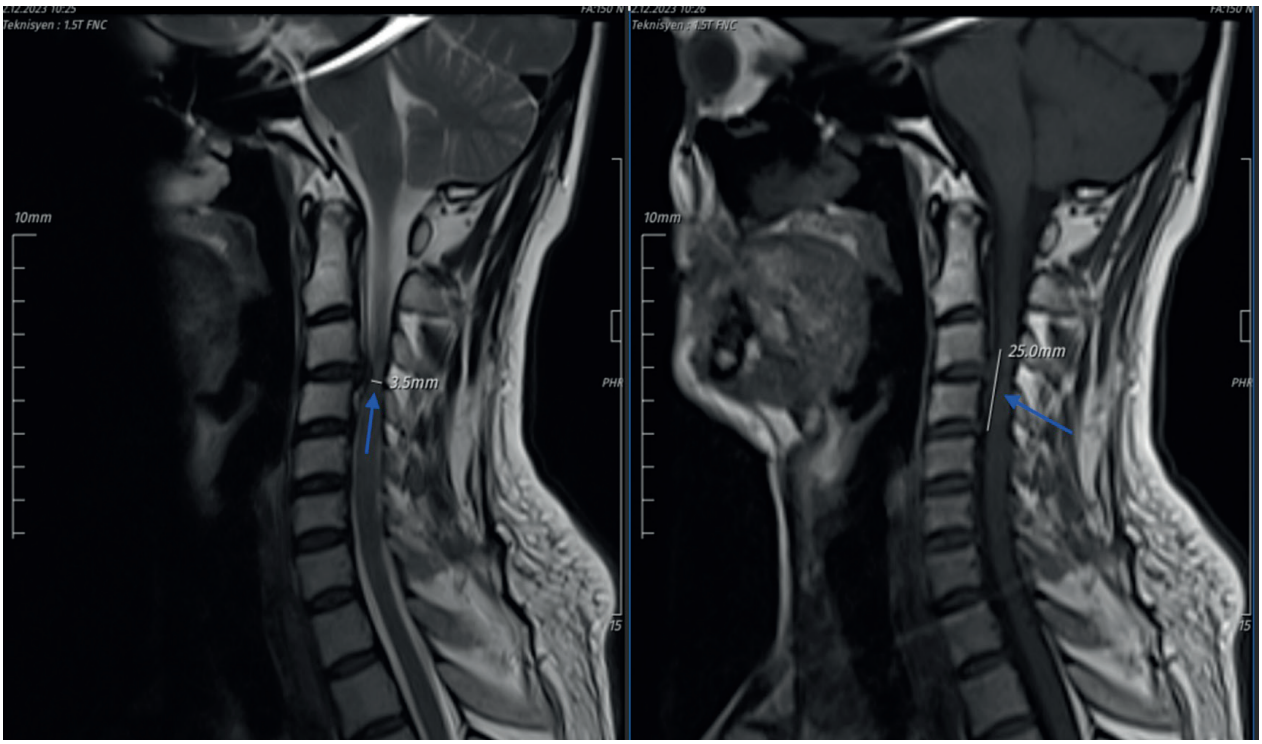


Figure 4. Cervical MRI image (IMLL:25 mm, channel diameter at MSCC level:3.5 mm).

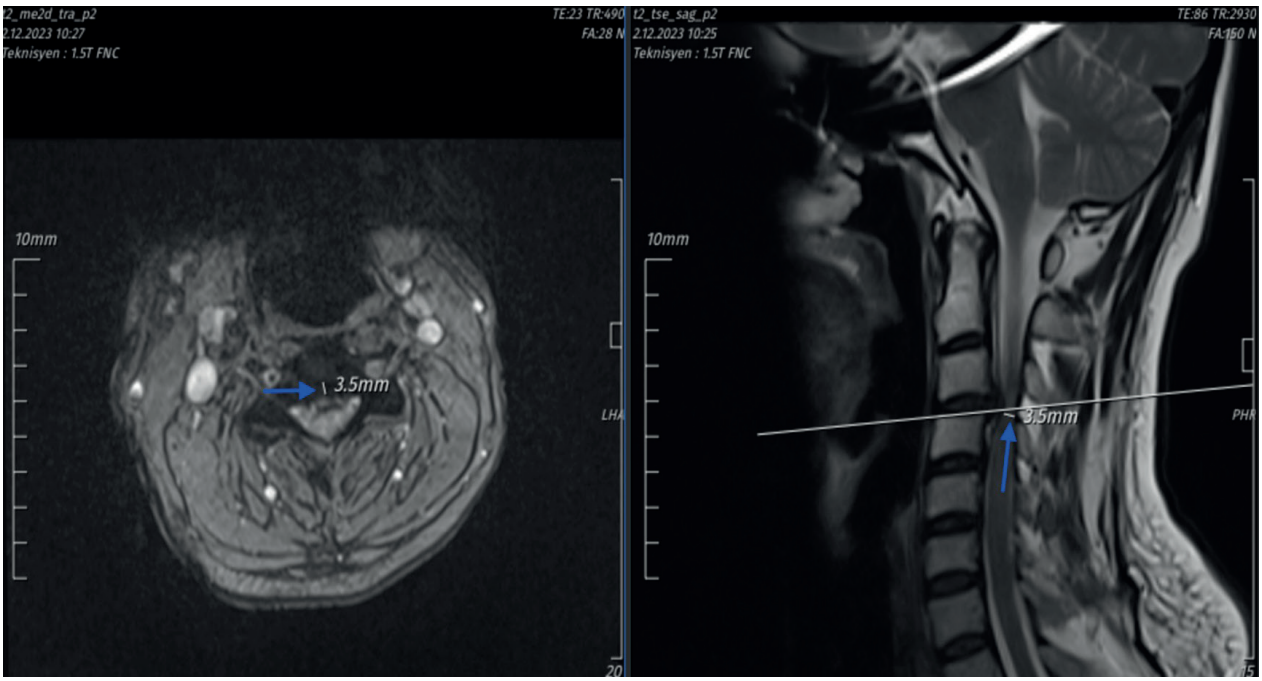


Figure 5. Cervical MRI image (channel diameter at MSCC level:3.5 mm).

DISCUSSION

Considering the multiple axes of motion, vertebrae can be injured by various mechanisms, and a number of different injury patterns can be identified (3,4). The cervical vertebrae (C1-C7) are the most commonly injured region of the spine because they are the most flexible, mobile, and exposed part of the spinal column. Most injuries occur at the C2 level and from C5 to C7 (5). Different studies have demonstrated that cervical spine injuries (CSI) are more common in motor vehicle accidents. Cervical spine injuries represent a significant proportion of spinal traumas, frequently observed in emergency departments. Globally, they account for approximately 60% of all spinal injuries, with road traffic accidents being the leading cause, responsible for over half of the cases (>50%) (6,7).

CT is the first imaging method preferred to evaluate the cervical spine in trauma. CT is more sensitive and specific than other radiological imaging modalities for assessing the cervical spine in trauma patients and can be performed promptly (8). However, MRI for soft tissue and spinal cord injuries have a higher sensitivity rate than CT. The sensitivity of CT is 88.6%, and its specificity is 99%. MRI has a sensitivity of 89,8%, and specificity of 99,2%. Although CT is highly sensitive in identifying cervical spine trauma, MRI plays a crucial role in detecting clinically relevant lesions that may be missed on CT, particularly in patients with persistent symptoms (9-11).

CSI is a devastating condition and a common cause of disability and death, especially in young people (disability ratio 20%, mortality ratio 5.1%) (12,13). An acute intervertebral disc herniation may develop due to hyperextension of the cervical region in trauma-related conditions such as motor vehicle accidents. Rapid and definitive diagnosis is possible with MRI (14). In acute CSI, MRI is the gold standard for the surgeon to decide on issues such as the need for surgery, the type of approach, the need for instrumentation, the level of decompression, and the need for re-surgery in case of insufficient decompression, as well as the presence of edema or hematoma (11,15,16). A recent study investigated the relationship between the length of the spinal cord intramedullary lesion (IMLL), the diameter of the canal at the maximum spinal cord compression level (MSCC), and the presence of intramedullary hemorrhage on MRI and prognosis. Shorter spinal

cord lesion (<6.5 mm), larger diameter of the canal at the MSCC level (>5.5 mm), and the absence of intramedullary hemorrhage were associated with better prognosis (17). In this case, although the canal diameter was narrow (3.5 mm) and the length of the IMLL lesion (25 mm) was longer, a good prognosis was observed due to the rapid diagnosis and the operation within 24 hours (Figure 4, Figure 5).

As a matter of fact, according to spinal cord trauma guidelines, neurological recovery is faster and more favorable in cases operated within 24 hours, as observed in our case (18- 21). In Türkiye, MRI scanners are available in nearly every province and in most large districts; in fact, some provinces are equipped with multiple MRI units. Performing MRI in patients with a clear indication, as in our case, has a favorable impact on prognosis. Therefore, this article was written to emphasize that relying solely on CT or X-ray may be insufficient, and that MRI should not be avoided when clinically indicated. To identify and monitor the presence and underlying causes of spinal cord hematoma, edema, and stenosis, MRI is required.

CONCLUSION

In motor vehicle accidents and even other traumas that cause hyperextension in the cervical region, cervical MRI should be performed in the presence of neurological deficit findings, even if CT and X-ray imaging are normal. When spinal cord injury is detected, early surgical intervention should be ensured by consulting with the neurosurgery department immediately. Thus, the patient's paralysis can be prevented by acting on-site and, when necessary, selected advanced imaging techniques.

Ethical approval

This study is not experimental or clinical research. Because it was a case report, ethical approval was not needed. Written informed consent was obtained from the patient for the publication of the case report.

Author contribution

Surgical and Medical Practices: BK, GY, FHÇ, NB, MİA; Concept: BK, GY, FHÇ, NB, MİA; Design: BK, GY, FHÇ, NB, MİA; Data Collection or Processing: BK, GY, FHÇ, NB, MİA; Analysis or Interpretation: BK, GY, FHÇ, NB, MİA; Literature Search: BK, GY, FHÇ, NB, MİA; Writing: BK, GY, FHÇ, NB, MİA. All authors reviewed the results and approved the final version of the article.

Source of funding

The authors declare the study received no funding.

Conflict of interest

The authors declare that there is no conflict of interest.

REFERENCES

1. Yamaguchi JT, Hsu WK. Intervertebral disc herniation in elite athletes. Int Orthop. 2019; 43(4): 833-40. [Crossref]
2. Gray BL, Buchowski JM, Bumpass DB, Lehman RA, Mall NA, Matava MJ. Disc herniations in the national football league. Spine (Phila Pa 1976). 2013; 38(38): 1934-8. [Crossref]
3. Amorosa LF, Vaccaro AR. Current concepts in cervical spine trauma. Instr Course Lect. 2014; 63: 255-62.
4. Syre P, Petrov D, Malhotra NR. Management of upper cervical spine injuries: a review. J Neurosurg Sci. 2013; 57(3): 219-40.
5. Greenbaum J, Walters N, Levy PD. An evidenced-based approach to radiographic assessment of cervical spine injuries in the emergency department. J Emerg Med. 2009; 36(1): 64-71. [Crossref]
6. Wang H, Zhang Y, Xiang Q, et al. Epidemiology of traumatic spinal fractures: experience from medical university-affiliated hospitals in Chongqing, China, 2001-2010. J Neurosurg Spine. 2012; 17(5): 459-68. [Crossref]
7. Beeharry MW, Moqeem K, Rohilla MU. Management of Cervical Spine Fractures: A Literature Review. Cureus. 2021; 13(4): e14418. [Crossref]
8. Bailitz J, Starr F, Beecroft M, et al. CT should replace three-view radiographs as the initial screening test in patients at high, moderate, and low risk for blunt cervical spine injury: a prospective comparison. J Trauma. 2009; 66(6): 1605-9. [Crossref]
9. Rutsch N, Amrein P, Exadaktylos AK, et al. Cervical spine trauma - Evaluating the diagnostic power of CT, MRI, X-Ray and LODOX. Injury. 2023; 54(7): 110771. [Crossref]

10. Mathew M, Mezue WC, Chikani MC, Jimoh AO, Uche EO, Mathew MB. Correlation of Quantitative MRI Parameters with Neurological Outcome in Acute Cervical Spinal Cord Injury. J West Afr Coll Surg. 2021; 11(1): 5-10. [Crossref]

11. Rutges JPHJ, Kwon BK, Heran M, Ailon T, Street JT, Dvorak MF. A prospective serial MRI study following acute traumatic cervical spinal cord injury. Eur Spine J. 2017; 26(9): 2324-32. [Crossref]

12. Häske D, Lefering R, Stock JP, Kreinest M, TraumaRegister DGU. Epidemiology and predictors of traumatic spine injury in severely injured patients: implications for emergency procedures. Eur J Trauma Emerg Surg. 2022; 48(3): 1975-83. [Crossref]

13. Kreinest M, Ludes L, Türk A, Grützner PA, Biglari B, Matschke S. Analysis of prehospital care and emergency room treatment of patients with acute traumatic spinal cord injury: a retrospective cohort study on the implementation of current guidelines. Spinal Cord. 2017; 55(1): 16-9. [Crossref]

14. Henninger B, Kaser V, Ostermann S, et al. Cervical Disc and Ligamentous Injury in Hyperextension Trauma: MRI and Intraoperative Correlation. J Neuroimaging. 2020; 30(1): 104-9. [Crossref]

15. Freund P, Seif M, Weiskopf N, et al. MRI in traumatic spinal cord injury: from clinical assessment to neuroimaging biomarkers. Lancet Neurol. 2019; 18(12): 1123-35. [Crossref]

16. Ghaffari-Rafi A, Peterson C, Leon-Rojas JE, et al. The Role of Magnetic Resonance Imaging to Inform Clinical Decision-Making in Acute Spinal Cord Injury: A Systematic Review and Meta-Analysis. J Clin Med. 2021; 10(21): 4948. [Crossref]

17. Dobran M, Aiudi D, Liverotti V, et al. Prognostic MRI parameters in acute traumatic cervical spinal cord injury. Eur Spine J. 2023; 32(5): 1584-90. [Crossref]

18. Sánchez JAS, Sharif S, Costa F, Rangel JAIR, Anania CD, Zileli M. Early Management of Spinal Cord Injury: WFNS Spine Committee Recommendations. Neurospine. 2020; 17(4): 759-84. [Crossref]

19. Tarawneh AM, D'Aquino D, Hilis A, Eisa A, Quraishi NA. Can MRI findings predict the outcome of cervical spinal cord Injury? a systematic review. Eur Spine J. 2020; 29(10): 2457-64. [Crossref]

20. Seif M, Curt A, Thompson AJ, Grabher P, Weiskopf N, Freund P. Quantitative MRI of rostral spinal cord and brain regions is predictive of functional recovery in acute spinal cord injury. Neuroimage Clin. 2018; 20: 556-63. [Crossref]

21. Fehlings MG, Tetreault LA, Wilson JR, et al. A Clinical Practice Guideline for the Management of Patients With Acute Spinal Cord Injury and Central Cord Syndrome: Recommendations on the Timing (≤24 Hours Versus >24 Hours) of Decompressive Surgery. Global Spine J. 2017; 7(Suppl 3): 195S-202S. [Crossref]

NORTHWESTERN MEDICAL JOURNAL

www.nwmedj.org

ISSN: 2979-9538