ISSN: 1305-9319 e-ISSN: 1305-9327 Number: 3 Volume: 21 Year: 2025

MEDICAL JOURNAL OF BAKIRKÖY

www.bakirkoytip.org

ISSN: 1305-9319 e-ISSN: 1305-9327 Number: 3 Volume: 21 Year: 2025

MEDICAL JOURNAL OF BAKIRKÖY

Medical Journal of Bakırköy of is an official scientific Journal of University of Health Sciences Türkiye, Bakırköy Dr. Sadi Konuk Training and Research Hospital (BMJ)

It is published quarterly as 4 issues every year (March, June, September, December).

Please refer to the journal's webpage **(https://bakirkoymedj.org/)** for "Ethical Policy" and "Instructions to Authors".

The editorial and publication processes of the journal are shaped in accordance with the guidelines of the ICMJE, WAME, CSE, COPE, EASE, and NISO. The journal conforms with the Principles of Transparency and Best Practice in Scholarly Publishing (doaj.org/bestpractice). The Medical Journal of Bakırköy is indexed in **Web of Science; Emerging Sources Citation Index (ESCI), TUBITAK ULAKBIM TR Index, Scopus, EBSCO – Academic Search Complete, J-Gate, Embase, CINAHL Complete, ScopeMed, WorldCat, Infobase, Ulrich's Database, Ideal Online, Türkiye Atıf Dizini, Türk Medline** and **CNKI**.

The journal is published online.

Owner: University of Health Sciences Türkiye, Bakırköy Dr. Sadi Konuk Training and Research Hospital

Responsible Manager: Musa Çırak

Publisher Contact
Address: Molla Gürani Mah. Kaçamak Sk. No: 21/1 34093 İstanbul, Türkiye
GSM: +90 530 177 30 97
E-mail: info@galenos.com.tr/yayin@galenos.com.tr
Web: www.galenos.com.tr **Publisher Certificate Number:** 14521

Printing Date: September 2025
ISSN: 1305-9319 **E-ISSN:** 1305-9327
International scientific journal published quarterly.

Advisory Board

Advisory Board

Advisory Board

Advisory Board

Advisory Board

Contents

Researches

Contents

Case Report

Medical Journal of Bakırköy

Med J Bakirkoy 2025;21(3):225-231
DOI: 10.4274/BMJ.galenos.2025.2024.4-7

Research

The Morphometrical Characteristics of the Proximal Tibia in the Anatolian Population

Anadolu Toplumunda Tibia Proksimalinin Morfometrik Özellikleri

Meryem Uçak, Betül Sevindik, Büşra Pirinç, Zeliha Fazlıoğulları

Selçuk University Faculty of Medicine, Department of Anatomy, Konya, Türkiye

ABSTRACT

Objective: Our study aimed to analyze the proximal tibia measurements of the Anatolian population to improve the design of appropriate prostheses in knee arthroplasties.

Methods: Specific morphometric measurements, including anteroposterior and mediolateral measurements of the condyles, tibial plateau, and proximal tibial circumference, were performed on 68 dry human tibias from the Anatolian population using digital calipers, and compared by statistical methods.

Results: Our study revealed significant differences between the proximal tibia morphometric measurements of the Anatolian population and other populations. Mean proximal tibia circumference, tibial plateau, and tibial tuberosity distances were significantly higher compared to the control group.

Conclusion: Population-specific data are important for optimizing tibial prosthesis design. Personalized treatments based on tibial morphometric measurements can better adapt tibial components in knee arthroplasties and improve individuals' functionality. Further studies are needed to increase the accuracy of tibial morphometric data with computed tomography support, and minimize age and sex differences.

Keywords: Tibial morphometry, arthroplasty, Anatolian population

ÖZ

Amaç: Çalışmamızın amacı, diz artroplastilerinde uygun protezlerin tasarımını geliştirmek için Anadolu popülasyonunun proksimal tibia ölçümlerini analiz etmektir.

Gereç ve Yöntem: Anadolu popülasyonuna ait 68 kuru insan tibiası dijital kumpas kullanılarak kondillerin anteroposterior ve mediolateral ölçümleri, tibial plato ve proksimal tibia çevresi dahil olmak üzere spesifik morfometrik ölçümleri yapılmış ve istatistiksel metodlarla karşılaştırılmıştır.

Bulgular: Çalışmamız, Anadolu popülasyonunun proksimal tibia morfometrik ölçümleri ile diğer popülasyonlar arasında önemli farklılıklar olduğunu ortaya koymuştur. Ortalama proksimal tibia çevresi, tibial plato ve tuberositas tibia mesafeleri anlamlı olarak daha yüksekti.

Sonuç: Popülasyona özgü veriler, tibial protez tasarımının optimizasyonu için önemlidir. Tibial morfometrik ölçümlere dayalı kişiselleştirilmiş tedaviler, diz artroplastilerinde tibial bileşenlerin daha iyi adaptasyonunu sağlayabilir ve bireylerin işlevselliğini artırabilir. Bilgisayarlı tomografi desteği ile tibial morfometrik verilerin doğruluğunun artırılarak, yaş ve cinsiyet farklılıklarının en aza indirildiği ileri çalışmalara ihtiyaç vardır.

Anahtar Kelimeler: Tibial morfometri, artroplasti, Anadolu popülasyonu

Address for Correspondence: Betül Sevindik MD, Selçuk University Faculty of Medicine, Department of Anatomy, Konya, Türkiye
E-mail: drbetulsevindik@gmail.com **ORCID ID:** orcid.org/0000-0003-1287-5544

Cite as: Uçak M, Sevindik B, Pirinç B, Fazlıoğulları Z. The morphometrical characteristics of the proximal tibia in the Anatolian population. Med J Bakirkoy. 2025;21(3):225-231

Received: 24.04.2024
Accepted: 04.10.2024
Publication Date: 03.09.2025

INTRODUCTION

The knee joints are susceptible to different structural and traumatic pathologies. Deformity in the proximal tibia, which is a part of the knee joint, disrupts the stabilization of the joint and impairs the overall quality of life. In the proximal tibia, damage on the upper articular facet may cause pain and limitation during joint movement (1). Total knee arthroplasty (TKA) is a treatment method preferred for patients in later stages to improve joint functionality and relieve pain. In arthroplasty, the tibia is more predisposed to complications than the femur. To obtain an optimal response from TKA, an anthropometrically optimized knee prosthesis is required (2,3). In the literature, there are studies on the morphometry of the tibia in different populations (4,5). Recent anthropometric studies on the tibia have shown that the current tibial implants used for TKA are not entirely appropriate for all populations due to interracial differences (6-9). Appropriate dimensioning of the tibial component in TKA is essential to maximize weight distribution proximal to the tibia. If the tibial component is small, loosening of the prosthesis may occur in the early postoperative period due to inadequate tibia support. A small tibial component increases the risk associated with unicondylar knee arthroplasty. If the implant is larger than needed, the protruding part may cause pain in the patient by irritating soft tissues (10). Our study aims to increase the success of surgical procedures by determining the proximal tibia parameters in the Anatolian population. Thus, the aim is to reduce postoperative pain, unnecessary reoperations, and financial and psychological losses. An increase in the success of surgical results will improve patients' quality of life. We think the data obtained will guide procedures such as knee replacement involving the proximal tibia.

METHODS

In this study, 68 human dry tibiae from adults of unknown age and sex, from the Anatolian population, without bone pathology were analyzed. All of these bones were sourced from the dried bone collection at the Department of Anatomy, Selçuk University Faculty of Medicine. It has obtained ethical approval from the Local Ethics Committee of Selçuk University (approval no: 2023/469, date: 13.10.2023). A digital caliper (Insize 1108/Suzhou, People's Republic of China) with a measuring range of 0-150 mm and an accuracy of 0.03 mm was utilized for parameter measurements. Caliper calibration was performed before each measurement. An authorized expert conducted all measurements twice, and the average of the two measurements was calculated. The detailed measurements are illustrated in Figure 1.

This study involving human participants was conducted in accordance with the ethical standards established by the institutional and national committees, following the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Since data for this study were obtained from the dry bone collection of our faculty, there is also no informed consent form.

Statistical Analysis

The data collected from the study were evaluated using the IBM SPSS Statistics 22 (IBM Corp., Armonk, NY, USA) software. A value of $p<0.05$ was considered significant. The minimum, maximum, mean, and standard deviation of the data were calculated. Normal distribution of the data was assessed using skewness and kurtosis tests. An independent samples t-test was used to compare the right and left tibia data. Data of the medial and lateral condyles were compared with the Pearson's correlation test.

RESULTS

Our study includes descriptive statistics of proximal tibia measurements of the Anatolian population. The minimum, maximum, mean, and standard deviation values of the measured bones were calculated from 68 dry bones (30 left and 38 right). When the skewness and kurtosis values of the proximal tibia measurements were analyzed, it was seen that they fit the normal distribution (Table 1). The data that will contribute to the analysis and explanation of the differences between the right and left sides of the proximal tibia are shown in Table 2.

Left-right proximal tibia measurements were compared. T-test results showed that the mean differences between the groups were not significant for "AB" ($t=-0.35$, $p=0.73$), "CD" ($t=0.78$, $p=0.44$), "EF" ($t=-1.04$, $p=0.30$), "GH" ($t=0.12$, $p=0.90$), and "IJ" ($t=0.19$, $p=0.85$), "KL" ($t=-0.19$, $p=0.85$), "MN" ($t=0.94$, $p=0.35$), "circle" ($t=0.001$, $p=0.99$), "tibial tuberosity (TT) length" ($t=-0.84$, $p=0.41$). However, there was a statistically significant difference according to the t-test results for the regions "OP" ($t=-2.85$, $p=0.01$) and "RS" ($t=-3.03$, $p=0.001$) (Table 3). The mean difference for "AB-EF" was 5.88 and the standard deviation was 3.25. The anteroposterior diameter of the medial condyle was significantly larger than the anteroposterior diameter of the lateral condyle ($t=14.93$, $p<0.001$).

The mean value of the "CD-GH" measurement difference was 1.66, and the standard deviation was 2.22. The mediolateral diameter of the medial condyle was significantly larger than the mediolateral diameter of the lateral condyle ($t=6.16$, $p<0.001$).

Figure 1. **(a)** Anteroposterior measurement of medial condyle (AB), **(b)** mediolateral measurement of medial condyle (CD), **(c)** anteroposterior measurement of lateral condyle (EF), **(d)** mediolateral measurement of lateral condyle (GH), **(e)** tibial plateau width measurement (IJ), **(f)** anteroposterior measurement of the intercondylar area (KL), **(g)** mediolateral measurement of the intercondylar area at the anterior end (MN), **(h)** mediolateral measurement of the intercondylar area in the middle narrow part (OP), **(i)** mediolateral measurement of the intercondylar area at the posterior end (RS), **(j)** circumference of the proximal tibia (Circle) **(k)** length between tibial plateau and tibial tuberosity (TT length) **(ı)** anterior view of the tibia

Table 1. Descriptive statistics of the proximal tibia measurement

Parameters	Min	Max	Mean	SD	Skewness	SD	Kurtosis	SD
AB	31.9	53.06	42.22	4.29	0.23	0.29	-0.10	0.57
CD	14.59	29.51	21.63	2.67	0.42	0.29	1,46	0.57
EF	28.03	46.79	36.33	3.73	0.15	0.29	0.04	0.57
GH	12.04	25.12	19.97	2.50	-0.20	0.29	0.27	0.57
IJ	52.09	78.76	67.49	5.57	-0.30	0.29	0.14	0.57
KL	37.68	58.55	46.73	4.65	0.32	0.29	-0.43	0.57
MN	18.46	36.16	25.73	3.50	0.03	0.29	0.18	0.57
OP	8.72	15.46	11.45	1.66	0.41	0.29	-0.46	0.57
RS	11.54	22.42	15.76	2.59	0.36	0.29	-0.72	0.57
Circle*	15.75	23.3	19.76	1.60	-0.05	0.29	-0.40	0.57
TT length	30.02	62.95	46.01	8.40	-0.09	0.29	-0.96	0.57

*Parameter value is cm, others are mm. AB: Anteroposterior measurement of medial condyle, CD: Mediolateral measurement of medial condyle, EF: Anteroposterior measurement of lateral condyle, GH: Mediolateral measurement of lateral condyle, IJ: Tibial plateau width measurement, KL: Anteroposterior measurement of the intercondylar area, MN: Mediolateral measurement of the intercondylar area at the anterior end, OP: Mediolateral measurement of the intercondylar area in the middle narrow part, RS: Mediolateral measurement of the intercondylar area at the posterior end, Circle: Circumference of the proximal tibia, TT length: Length between tibial plateau and tibial tuberosity, Min: Minimum, Max: Maximum, SD: Standard deviation, n: Number

Table 2. Left-right side comparison data of the proximal tibia

Parameters	Side	n	Mean	SD	SE
AB	Left	30	42.01	3.78	0.69
	Right	38	42.37	4.69	0.76
CD	Left	30	21.91	2.55	0.46
	Right	38	21.41	2.77	0.45
EF	Left	30	35.81	3.50	0.64
	Right	38	36.74	3.89	0.63
GH	Left	30	20.01	2.09	0.38
	Right	38	19.93	2.81	0.45
IJ	Left	30	67.63	4.45	0.81
	Right	38	67.38	6.37	0.03
KL	Left	30	46.12	4.01	0.73
	Right	38	47.21	5.10	0.83
MN	Left	30	26.17	3.08	0.56
	Right	38	25.38	3.81	0.62
OP	Left	30	10.84	1.50	0.27
	Right	38	11.93	1.64	0.26
RS	Left	30	14.78	1.93	0.35
	Right	38	16.53	2.81	0.45
Circle*	Left	30	19.76	1.29	0.23
	Right	38	19.76	1.83	0.30
TT length	Left	30	45.04	8.38	1.53
	Right	38	46.77	8.46	1.37

*Parameter value is cm, others are mm. AB: Anteroposterior measurement of medial condyle, CD: Mediolateral measurement of medial condyle, EF: Anteroposterior measurement of lateral condyle, GH: Mediolateral measurement of lateral condyle, IJ: Tibial plateau width measurement, KL: Anteroposterior measurement of the intercondylar area, MN: Mediolateral measurement of the intercondylar area at the anterior end, OP: Mediolateral measurement of the intercondylar area in the middle narrow part, RS: Mediolateral measurement of the intercondylar area at the posterior end, Circle: Circumference of the proximal tibia, TT length: Length between tibial plateau and tibial tuberosity, SD: Standard deviation, SE: Standard error, n: Number

Table 3. Independent sample t-test for left-right proximal tibia measurements

Parameters	t-test for equality of means						
	t	df	p-value	Mean difference	SE	95% confidence interval	
						Low	Up
AB	-0.35	65.96	0.73	-0.36	1.03	-2.41	1.69
CD	0.78	64.41	0.44	0.5	0.65	-0.79	1.8
EF	-1.04	64.84	0.30	-0.94	0.9	-2.73	0.86
GH	0.12	65.8	0.90	0.07	0.59	-1.11	1.26
IJ	0.19	65.14	0.85	0.25	1.31	-2.37	2.88
KL	-0.99	66	0.33	-1.09	1.11	-3.3	1.12
MN	0.94	65.95	0.35	0.78	0.84	-0.88	2.45
OP	-2.85	64.6	0.01	-1.09	0.38	-1.85	-0.33
RS	-3.03	64.89	0.001	-1.75	0.58	-2.9	-0.59
Circle	0.001	65.29	0.99	0	0.38	-0.76	0.76
TT length	-0.84	62.64	0.41	-1.72	2.05	-5.83	2.38

AB: Anteroposterior measurement of condylus medialis, CD: Mediolateral measurement of condylus medialis, EF: Anteroposterior measurement of condylus lateralis, GH: Mediolateral measurement of condylus lateralis, IJ: Tibial plateau width measurement, KL: Anteroposterior measurement of the area intercondylaris, MN: Mediolateral measurement of the area intercondylaris at the anterior end, OP: Mediolateral measurement of the area intercondylaris in the middle narrow part, RS: Mediolateral measurement of the area intercondylaris at the posterior end, Circle: Circumference of the proximal tibia, TT length: Length between tibial plateau and tuberositas tibia, t: t-value, df: Degrees of freedom, SE: Standart error

For the "circumference-TT length" measurement difference, the mean was -26.24 and the standard deviation was 7.63. The distance between the circle of the proximal tibia and the tibial plateau-TT was statistically significant and greater (t=-28.32, p<0.001).

DISCUSSION

There are few studies on proximal tibia measurements in the literature. Proximal tibia measurements aim to reveal differences between societies and sexes. Today, most knee implants are produced to fit the anatomical characteristics of European and North American populations. Due to anthropometric differences in the knee joint, there is an incompatibility between knee implants and the resected surfaces. This incompatibility leads to recurrence of postoperative pain, complications, and loss of quality of life (10). To resolve this incompatibility, the morphometric characteristics of different populations should be considered in tibial implant construction. Our study provides morphometric characteristics of the proximal part of the tibia in the Anatolian population.

In one of the limited number of studies in the literature (Table 4), Gandhi et al. (11) examined 100 dried human tibias from Indian populations. They measured the AB of the medial condyle as 45.23±4.34 mm and the CD of the medial condyle as 28.44±2.8 mm; the EF of the lateral condyle was 38.91±3.69 mm, and the GH of the lateral condyle was 27.4±2.95 mm. Servien et al. (12) examined the AB value of the medial condyle as 50.80±3.3 mm, the CD value of the medial condyle as 28.80±2.5 mm, the EF value of the lateral condyle as 47.20±3.3 mm and the GH value of the lateral condyle as 29.30±2.4 mm with computed tomography (CT) in 37 French patients. According to the data obtained by Zalawadia and Patel (13) from 120 human dry tibiae belonging to the Indian (Gujarat) population the AB value of the medial condyle was 42.07 mm, the CD value of the medial condyle was 27.26 mm, the EF value of the lateral condyle was 36.94 mm, and the GH value of the lateral condyle was 28.07 mm. In the study by Pradhan et al. (14), which examined 200 dry tibiae from Eastern India, AB and EF were found to be 42.72 mm and 39.44 mm, respectively. Bilkay et al.'s (15) study, conducted on 44 dry tibia bones in the Turkish population, determined that AB and EF were 41.9 mm and 36.5 mm, respectively. In the study of Toy and Secgin (16), AB, CD, EF, and GH values were found to be 47.51 mm, 30.18 mm, 43.28 mm, and 32.79 mm, respectively, in the results of 33 dry tibia bones in the Turkish population. In Shojaolsadati et al.'s (17) study, 200 proximal tibiae in the Anatolian population were examined by magnetic resonance imaging, and AB and EF were determined to be 47.1 mm and 38.5 mm, respectively. In one of the recent studies, Akdemir Aktaş et al. (4) measured the AB value of the medial condyle as 39.76±4.10 mm, the CD value of the medial condyle as 23.27±2.63 mm, the EF value of the lateral condyle as 34.72±3.51 mm, and the GH value of the lateral condyle as 21.83±2.52 mm in their study with 57 human dry tibiae from the Turkish population.

Our study's AB, CD, EF, and GH measurements were lower than those of the Indian and French populations. We think that this result may be due to ethnic differences. However, the study by Servien et al. (12) used CT data. The large size of the measurements compared to dry bone studies is a reasonable finding. Our study's values for anteroposterior diameter are consistent with the literature. In the measurements, the AB value of the

Table 4. Comparison of morphometric measurements of the proximal tibia

	AB	CD	EF	GH	IJ	KL	MN	OP	RS	Circle*	TT length
Servien et al. (12)	50.80	28.80	47.2	29.30	-	-	-	-	-	-	-
Gandhi et al. (11)	45.23	28.44	39.41	27.4	-	46.01	23.79	6,92	6,92	-	-
Bilkay et al. (15)	41.9	-	36.5	-	70.2	-	-	-	-	-	-
Zalawadia and Patel (13)	42.07	27.26	36.94	28.07	70.78	-	-	-	-	-	-
Shojaolsadati et al. (17)	47.1	-	38.5	-	77	-	-	-	-	-	-
Toy and Secgin (16)	47.51	30.18	43.28	32.79	-	-	-	-	-	-	-
Akdemir Aktaş et al. (4)	39.76	23.27	34.72	21.83	65.14	41.62	24.86	21.27	21.36	-	-
Pradhan et al. (14)	42.72	-	39.44	-	67.91	42.73	-	-	-	19.37	33.48
Our study	42.22	21.63	36.33	19.97	67.49	46.73	25.73	11.45	15.76	19.76	46.01

*Parameter value is cm, others are mm. AB: Anteroposterior measurement of medial condyle, CD: Mediolateral measurement of medial condyle, EF: Anteroposterior measurement of lateral condyle, GH: Mediolateral measurement of lateral condyle, IJ: Tibial plateau width measurement, KL: Anteroposterior measurement of the intercondylar area, MN: Mediolateral measurement of the intercondylar area at the anterior end, OP: Mediolateral measurement of the intercondylar area in the middle narrow part, RS: Mediolateral measurement of the intercondylar area at the posterior end, Circle: Circumference of the proximal tibia, TT length: Length between tibial plateau and tibial tuberosity

medial condyle was significantly higher than the EF value of the lateral condyle. Similarly, the mediolateral diameter (CD) of the medial condyle was significantly greater than the mediolateral diameter (GH) of the lateral condyle ($p<0.05$). These results are a possible consequence of the asymmetry of the medial and lateral tibial condyles. However, when TKA prostheses are examined, they are mostly designed symmetrically. This symmetrical production approach is cost-oriented. However, when the postoperative disability and health problems are taken into consideration, the damages incurred may far exceed the the prosthesis production costs. This difference between the medial and lateral condyles must be taken into consideration when developing knee prostheses (18).

Tibial plateau width (IJ) was measured as 65.14±4.70 mm by Akdemir Aktaş et al. (4). Our IJ data were higher than those reported by Akdemir Aktaş et al. (4). Zalawadia and Patel (13) found the mean IJ value to be 70.78±2.40 mm. Our IJ data were lower than this study's data. The IJ value was found to be 67.91 mm, 70.2 mm, and 77 mm in the studies of Pradhan et al. (14), Bilkay et al. (15), and Shojaolsadati et al. (17), respectively. Our study's IJ value is 67.49 mm, which is lower than these studies'. The KL value of the intercondylar area was reported as 46.01±3.62 mm by Gandhi et al. (11) and 41.62±4.27 mm by Akdemir Aktaş et al. (4). While the KL value we found, 46.73 mm, was similar to that of Gandhi et al. (11), it was larger than that of Akdemir Aktaş et al. (4). The MN value of the intercondylar area at the anterior end was found to be 23.79 mm by Gandhi et al. (11) and 24.86 mm by Akdemir Aktaş et al. (4). We found the MN value to be 25.73 mm. IJ, KL, and MN measurements were consistent with the literature. It can be considered that standard production in prosthesis design does not make a significant difference for IJ, KL, and MN regions, and these regions are safe for complications.

The OP value of the intercondylar area in the middle narrow part was reported as 6.92 mm by Gandhi et al. (11), 21.27 mm by Akdemir Aktaş et al. (4), and 11.45 mm in our study. The RS value of the intercondylar area at the posterior end was reported as 6.92 mm by Gandhi et al. (11), was reported as 21.36 mm by Akdemir Aktaş et al. (4), and was reported as 15.76 mm in our study. Our OP and RS results were inconsistent with the literature data. Accordingly, OP and RS regions suggest that tibia dimensions may differ significantly regardless of ethnicity.

When comparing measurements between studies, the expected differences between CT and dry bone measurements were not evaluated separately. In addition, when we compare our data with the previous study of the Anatolian population, the differences can be attributed to the fact that Anatolia is located on a migration route due to its geopolitical position and is home to a mobile-heterogeneous population. However, the accessibility of dry bone is becoming increasingly difficult, and the absence of sex differences to account for data variability presents an important limitation. Tibial morphometric measurements are worthwhile for personalized prosthetic treatments.

Circle and TT length measurements in the proximal tibia have been evaluated in limited studies. In Pradhan et al.'s (14) study, the circle and TT length measurements were found to be 19.37 cm and 33.48 mm, respectively. While the circle value was compatible with our study, the TT length was found to be higher. According to our measurements, the mean circle was significantly greater than the mean TT length. Taking the circumference and TT length as the basis for designing prostheses will provide a more appropriate prosthesis for better adaptation to the tibia.

Study Limitations

A limitation of our study was the lack of age and gender data due to the dry bone method.

CONCLUSION

A better understanding of the anatomy of the proximal tibia is essential for the design of appropriate tibial prostheses. The average KL value of the intercondylar area, the OP value of the intercondylar area in the middle narrow part, and the RS value of the intercondylar area at the rear end, differ from those of the same population, supporting the necessity of a personalized prosthesis. Patient-specific prostheses may resolve incompatibilities soon. The difference between the anteroposterior and mediolateral dimensions of the medial and lateral condyles should be considered when designing knee prostheses. More comprehensive studies evaluating the proximal tibia with CT in terms of sex and age differences can be planned.

ETHICS

Ethics Committee Approval: It has obtained ethical approval from the Local Ethics Committee of Selçuk University (approval no: 2023/469, date: 13.10.2023).

Informed Consent: Since data for this study were obtained from the dry bone collection of our faculty, there is also no informed consent form.

Acknowledgments

The authors thank the individuals who contributed bones to the study.

FOOTNOTES

Authorship Contributions

Surgical and Medical Practices: B.P., Concept: Z.F., Design: Z.F., Data Collection or Processing: M.U., Analysis or Interpretation: B.S., B.P., Z.F., Literature Search: M.U., B.S., Writing: M.U.

Conflict of Interest: No conflict of interest was declared by the authors.

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

REFERENCES

1. Desdicioğlu K. Articulatio genu'nun morfolojik özellikleri. Med J Süleyman Demirel Uni. 2008;15:45-52.
2. Erkocak OF, Kucukdurmaz F, Sayar S, Erdil ME, Ceylan HH, Tuncay I. Anthropometric measurements of tibial plateau and correlation with the current tibial implants. Knee Surg Sports Traumatol Arthrosc. 2016;24:2990-7.
3. Cheng CK, Lung CY, Lee YM, Huang CH. A new approach of designing the tibial base plate of total knee prostheses. Clin Biomech. 1999;14:112-7.
4. Akdemir Aktaş H, Ülkir M, Günenç Beşer C, Demiryürek D. Reappraisal of the proximal tibia anatomy in Turkish population. Surg Radiol Anat. 2023;45:263-70.
5. Gupta C, Kumar J, Kalthur SG, D'souza AS. A morphometric study of the proximal end of the tibia in South Indian population with its clinical implications. Saudi J Sports Med. 2015;15:166-9.
6. Ho WP, Cheng CK, Liau JJ. Morphometrical measurements of resected surface of femurs in Chinese knees: correlation to the sizing of current femoral implants. Knee. 2006;13:12-4.
7. Urabe K, Miura H, Kuwano T, Matsuda S, Nagamine R, Sakai S, et al. Comparison between the shape of resected femoral sections and femoral prostheses used in total knee arthroplasty in Japanese patients: simulation using three-dimensional computed tomography. J Knee Surg. 2003;16:27-33.
8. Vaidya SV, Ranawat CS, Aroojis A, Laud NS. Anthropometric measurements to design total knee prostheses for the Indian population. J Arthroplasty. 2000;15:79-85.
9. Yip DK, Zhu Y, Chiu KY, Ng TP. Distal rotational alignment of the Chinese femur and its relevance in total knee arthroplasty. J Arthroplasty. 2004;19:613-9.
10. Küçükdurmaz F, Tuncay I, Elmadag M, Tuncer N. Morphometry of the medial tibial plateau in Turkish knees: correlation to the current tibial components of unicompartmental knee arthroplasty. Acta Orthop Traumatol Turc. 2014;48:147-51.
11. Gandhi S, Singla RK, Kullar JS, Suri RK, Mehta V. Morphometric analysis of upper end of tibia. J Clin Diagnostic Res. 2014;8:AC10-13.
12. Servien E, Saffarini M, Lustig S, Chomel S, Neyret P. Lateral versus medial tibial plateau: morphometric analysis and adaptability with current tibial component design. Knee Surg Sports Traumatol Arthrosc. 2008;16:1141-5.
13. Zalawadia AZ, Patel SM. Morphometric study of upper end of tibia in Gujrat region and its clinical implication in knee arthroplasty. Int J Anat Res. 2018;6:4871-5.
14. Pradhan RR, Korukonda S, Patro GC, Prusti JS. Morphometric analysis of the proximal end of the dry adult tibia: a cross-sectional study from eastern India. J Clin Diagn Res. 2023;17:AC01-04.
15. Bilkay C, Dursun A, Özgüner G, Öztürk K, Candan B, Sulak O. Morphometric features of the proximal tibia in Turkish society and its clinical significance. Int J Basic Clin Med. 2016;4:109-17.
16. Toy S, Secgin Y. Determination of dry tibia bone morphometry by photo analysis: dry tibia bone morphometry. Int J Curr Med Biol Sci. 2022;2:165-70.
17. Shojaolsadati P, Yüzbasioglu N, Nalbant A, Ormeci T, Albay S, Ortug A, et al. The magnetic resonance imaging evaluation of morphometry of the distal femur and proximal tibia on adult Anatolian population. J Anat Soc India. 2020;69:9-14.
18. Koh YG, Nam JH, Chung HS, Lee HY, Kang KT. Morphologic difference and size mismatch in the medial and lateral tibial condyles exist with respect to gender for unicompartmental knee arthroplasty in the Korean population. Knee Surg Sports Traumatol Arthrosc. 2020;28:1789-96.

Medical Journal of Bakırköy

Med J Bakirkoy 2025;21(3):232-239
DOI: 10.4274/BMJ.galenos.2025.2024.11-2

The Effect of an Exercise Programme for a Healthy Pregnancy

Sağlıklı Bir Gebelik İçin Egzersiz Programının Etkisi

Abdurrahim Yıldız[1], Özge Fidan[1], Tuba Tarnı[1], Buse Aktaş[1], Yasemin Tanrıkulu[2]

[1]Sakarya University of Applied Sciences, Faculty of Health Sciences, Department of Physiotherapy and Rehabilitation, Sakarya, Türkiye
[2]Sakarya University Training and Research Hospital, Department of Physiotherapy and Rehabilitation, Sakarya, Türkiye

ABSTRACT

Objective: This study aims to reduce musculoskeletal pain and increase physical activity capacity during pregnancy, thereby enhancing daily living activities. It aims to manage the increased musculoskeletal pain and decreased physical activity capacity during pregnancy with exercise and to determine the benefits on activities of daily living.

Methods: The study was conducted with 42 pregnant women between 16-34 weeks gestation who participated in a 6-week exercise program at the Pregnancy School of Sakarya Training and Research Hospital and Physiotherapy and Rehabilitation Application and Research Center. Short Form-36 (SF-36) was used for quality of life, Visual Analogue Scale (VAS) for pain severity, Pregnancy Physical Activity Questionnaire for physical activity levels, and McGill Pain Questionnaire for pain characteristics.

Results: Post-intervention analyses showed that pain was reduced and quality of life significantly improved. SF-36 scores showed significant improvements in the sub-parameters of physical role, emotional well-being, social functioning, and general health. VAS scores showed significant reductions in pain during activity, rest, and at night. Participants also showed significant improvements in the home/care and sport/exercise subdomains of physical activity.

Conclusion: Regular physical activity during pregnancy effectively reduces pain levels and improves quality of life. The study supports the inclusion of structured exercise programs in antenatal care, highlighting the benefits of physical activity in managing pain and improving the overall well-being of pregnant women. This approach not only facilitates daily activities but also helps reduce complications during and after labour.

Keywords: Exercise, pain, pregnancy, physical activity, quality of life

ÖZ

Amaç: Bu çalışma, hamilelik sırasında kas-iskelet ağrısını azaltmayı ve fiziksel aktivite kapasitesini artırmayı, böylece günlük yaşam aktivitelerini geliştirmeyi amaçlamaktadır. Gebelikte artan kas-iskelet sistemi ağrısı ve azalan fiziksel aktivite kapasitesinin egzersiz ile düzenlenmesi ve günlük yaşam aktiviteleri üzerine faydalarının belirlenmesi amaçlanmıştır.

Gereç ve Yöntem: Çalışma Sakarya Eğitim ve Araştırma Hastanesi Gebelik Okulu ve Fizyoterapi ve Rehabilitasyon Uygulama ve Araştırma Merkezi'nde, 6 haftalık egzersiz programına katılan 16-34 gebelik haftaları arasındaki 42 gebe ile yürütüldü. Değerlendirmelerde yaşam kalitesi için Kısa Form-36 (SF-36), ağrı şiddeti için Görsel Analog Ölçeği (GAS), fiziksel aktivite düzeyleri için Gebelik Fiziksel Aktivite Anketi ve ağrı özellikleri için McGill Ağrı Anketi kullanılmıştır.

Bulgular: Müdahale sonrasında yapılan analizler, ağrının azaldığını ve yaşam kalitesinin anlamlı derecede iyileştiğini göstermiştir. SF-36 skorları fiziksel rol, duygusal refah, sosyal işlevsellik ve genel sağlık alt parametrelerinde anlamlı iyileşmeler göstermiştir. GAS skorları aktivite, dinlenme ve gece ağrılarında önemli azalmalar olduğunu göstermiştir. Katılımcılar ayrıca fiziksel aktivitenin ev/bakım ve spor/egzersiz alt alanlarında da anlamlı iyileşmeler göstermiştir.

Sonuç: Hamilelik sırasında düzenli fiziksel aktivite ağrı seviyelerini etkili bir şekilde azaltır ve yaşam kalitesini artırır. Çalışma, yapılandırılmış egzersiz programlarının doğum öncesi bakıma dahil edilmesini desteklemekte, fiziksel aktivitenin ağrıyı yönetmede ve hamile kadınların genel refahını iyileştirmedeki faydalarını vurgulamaktadır. Bu yaklaşım sadece günlük aktiviteleri kolaylaştırmakla kalmaz, aynı zamanda doğum sırasında ve sonrasında komplikasyonların azaltılmasına da yardımcı olur.

Anahtar Kelimeler: Egzersiz, ağrı, gebelik, fiziksel aktivite, yaşam kalitesi

Address for Correspondence: Lec, Abdurrahim Yıldız, MD, Sakarya University of Applied Sciences Faculty of Health Sciences, Department of Physiotherapy and Rehabilitation, Sakarya, Türkiye
E-mail: abdurrahimyildiz@subu.edu.tr **ORCID ID:** orcid.org/0000-0002-6049-0705

Cite as: Yıldız A, Fidan Ö, Tarnı T, Aktaş B, Tanrıkulu Y. The effect of an exercise programme for a healthy pregnancy. Med J Bakirkoy. 2025;21(3):232-239

Received: 11.11.2024
Accepted: 11.02.2025
Publication Date: 03.09.2025

INTRODUCTION

Pregnancy is the term used to describe the period during which a foetus develops in a woman's uterus. Pregnancy is a sensitive period in the mother's life during which she experiences significant physiological and psychological changes (1). Hormonal and physical changes during this period have significant effects on many systems of the body, especially the musculoskeletal, cardiovascular, and respiratory systems (2). These changes manifest as physiological adaptations such as joint laxity, increased cardiac output, and metabolic adjustments (2). These factors can negatively affect activities of daily living, physical activity (PA) capacity, and pain levels. Exercise programmes can help in adapting to these changes and improve quality of life (3). Pilates and various exercises can be performed according to the physiological changes of pregnancy. Regular training has been shown to strengthen the pelvic floor muscle and increase its structural integrity and function. Pelvic floor muscle contraction exercises are part of modern Pilates. Pilates exercise during pregnancy is a safe method to shorten the duration of the active and second stages of labour, reduce birth pain and increase the mother's satisfaction with the birth process (4).

Perceived quality of life is a key component of perinatal health, influenced by physical, psychological, and social factors (5). Potential mechanisms of the PA and quality of life/health relationship include PA-induced changes in neurotransmitters in the brain and endogenous opioids known to be associated with depression, anxiety, and other mood constructs (6-8). The critical role of PA in health has been consistently demonstrated in numerous studies. PA is associated with at least a 20-30% reduction in premature mortality and with a reduced risk of developing more than 25 chronic medical conditions (6). PA improves lifestyle and is recommended for women during pregnancy as a way to limit complications that may occur (7). Especially because the responsibilities of being a parent start before birth, it indicates pregnant women enter a psychologically and physically difficult process (7). Physiological changes during pregnancy, including weight gain, intra-abdominal pressure, hormonal fluctuations, and fluid retention, significantly increase the risk of musculoskeletal disorders and pain (8). Changes in the body during pregnancy can increase the intensity of pain and make it worse if there is a pre-existing history of pain (8). Musculoskeletal, rheumatological, neuropathic, and pelvic-abdominal pain syndromes can be classified into groups. Pelvic girdle pain (PGP) and low back pain (LBP) occur in many pregnant women and can also have a negative impact on their quality of life (9). During pregnancy, abdominopelvic musculoskeletal issues, particularly lower back strain, are the primary source of pain. LBP occurs in pregnant women due to weight gain, laxity in the ligaments, and the growth of the uterus, causing mechanical tension. Apart from pain, the changes that occur during pregnancy also affect the quality of sleep, as they affect fatigue and energy (10).

This study was planned to reduce the difficulties experienced during pregnancy and to facilitate activities of daily living. PA facilitates life in every sense before and after pregnancy. Facilitates birth and reduces complications (11). In our country, the difficulty of access to services for pregnant women due to the high cost in private sectors, the partially inadequate facilities for pregnant women in public institutions, and the inadequacy of the existing facilities in pregnancy education were considered one of the main factors in this study. Based on these reasons, exercise training, especially in pregnancy, is of great importance. During the training of the pregnant women, the physical conditions were examined, hormonal and mental disorders, and changes occurring in the respiratory systems were examined. It was planned to create exercise programmes for these conditions during pregnancy and to make the necessary follow-ups. During our study, we aim to convey the importance of pregnancy education to pregnant individuals, follow their processes with exercise, and provide better quality of life.

METHODS

Sample of the Research

Our study was carried out with pregnant individuals attending the Pregnancy School of Sakarya Training and Research Hospital and Physiotherapy and Rehabilitation Application and Research Center between October 2023 and January 2024. We included 42 people in the study because there was a possibility of participant dropout. The exercise program was conducted for 45 minutes a day, 3 days a week for 6 weeks. The inclusion criteria were pregnant women between 16-34 weeks who were volunteering to participate in the study. Exclusion criteria included the presence of significant heart disease, risk of preterm delivery, pregnancy intoxication, pregnancy-induced hypertension, and significant lung disease such as chronic obstructive pulmonary disease.

The study was conducted in accordance with the rules of the Helsinki Declaration. Ethical approval for the study was obtained from the Sakarya University of Applied Sciences Ethics Committee (approval no: E-26428519-044-89268,

date: 07.07.2023). The study was explained to the patients face to face, and verbal consent was obtained.

The exercise program outlines an exercise routine designed for pregnant women. First, low-impact general warm-ups were performed to provide mobility in the main joints. Afterwards, aerobic activities requiring moderate effort were performed to increase the heart rate along with breathing exercises, using sports equipment. Muscle strengthening exercises were performed with resistance bands and light weights to work different muscle groups of the body. Another part of the program consisted of simple coordination and balance exercises with sports equipment. Kegel exercises were performed to strengthen the pelvic floor muscles, and the program ended with stretching and relaxation movements (12,13).

Data Collection Tools

Short Form-36 (SF-36): It is a scoring system used to evaluate eight subcategories. These parameters are: physical role, bodily pain, physical functioning, energy, social functioning, emotional role, general health, change in health. There is a scale from 0 to 100 for each subcategory. An increase in the scoring result indicates an increase in quality of life (14).

Visual Analogue Scale (VAS): It is a reliable and valid scale that measures pain intensity. The scale consists of a 10 cm long horizontal line. According to VAS, pain intensity is graded between 0 and 10. The patient marks a point on the line that accurately reflects their pain. The intensity of the pain increases as the number he/she marks gets closer. The distance, usually measured in millimeters, is reported as points. In pregnancy, back, low back, knee, and hip pain are very high, and the pregnant woman marks the degree of pain with this scale (15).

Pregnancy Physical Activity Questionnaire (PPAQ): This questionnaire is a 35-question scale that assesses the level of PA during pregnancy. A total of 32 activities are evaluated in these 35 questions. The scale has 4 sub-dimensions: (housework and care), occupation, (sports and exercise), and (transport and sedentary activities). The validity and reliability of the Turkish version were validated and tested by Tosun et al. (16).

McGill-Melzack Pain Questionnaire (MPQ): Mainly used by patients to indicate the subjective level of pain. The questionnaire consists of three main classes of word descriptors: sensory, emotional, and evaluative. It also includes an intensity scale and other items to determine the level and characteristics of the pain experienced. The questionnaire is designed to provide quantitative measures of clinical pain that can be treated statistically (17).

Statistical Analysis

G*Power analysis program (version 3.1.9.7; Heinrich Heine University, Düsseldorf, Germany) was used to determine the number of people to be included in the study. Using the light PA values, from the reference article, it was found that a total of 38 people should be included with an effect size of 0.489 at a margin of error of 0.05 and a power of 0.90. Due to the possibility of participants leaving the study unfinished, it was planned to include 46 people, which is 10 percent more than the original 42 people (18). The SPSS version 27.0 program (IBM Corp., Armonk, NY, USA) was used for data analysis. In data analysis, descriptive characteristics were presented as number, percentage, median, interquartile range and minimum-maximum values. The normality assumption of continuous variables was evaluated by paired t-tests and Pearson correlation tests. Since the data were not normally distributed, the relationship between PPAQ and SF-36 should be evaluated by a more appropriate method than Pearson correlation analysis. The strength of the correlation was categorised as weak (rho=0.00-0.24), moderate (rho=0.25-0.49), strong (rho=0.50-0.74) and very strong (rho=0.75-1.00).

RESULTS

The mean age of the pregnant women who participated in the study was 27.88±4.67 years. The mean body mass index (BMI) before and during pregnancy were 23.23±3.11 and 25.97±3.22. The mean gestational weeks was 25.57±4.06. 92.8% of the pregnant women included in the study are having their first child. 69% of the participants stated that they had planned pregnancies, and 31% stated that they had unplanned pregnancies. While 81% of the pregnant women participating in the study did not have a history of miscarriage, 19% had experienced a miscarriage previously. 11.9% of the pregnant women included in the study had previously been at risk of miscarriage. Of the women surveyed, 88.1% did not have a risk of miscarriage. 42.9% of the pregnant women participating in the study had premature births, while 54.8% did not have premature births. 80.9% of the pregnant women did not have a chronic disease. In the other participants, 4.8% had familial Mediterranean fever, 4.8% had anxiety disorder, 4.8% had anaemia, 2.4% had hypertension, 2.4% had panic attacks, and 2.4% had thyroid disease.

When we evaluated the correlation of the sub-dimensions of the SF-36 quality of life questionnaire with pain and BMI, a significant correlation was found between BMI and the sub-dimensions of general health (p=0.016) and health change (p=0.043). A significant correlation was found between the total score of the MPQ and the sub-dimensions of physical

function (p<0.001), physical health role limitation (p<0.001), energy/fatigue (p=0.026), and pain (p<0.001). When we looked at the correlation with VAS parameters, a significant correlation was found between health change (p=0.001) the VAS night score, as well as between pain (p=0.046) and the VAS night score (see Table 1).

When VAS activity, nighttime, and resting parameters of the participants were compared before and after exercise, statistically significant differences were found in all three parameters (p<0.001). Comparison of the values before and after treatment for the parameters of pain intensity, pain characteristic, pain-time relationship, and total questionnaire score of the MPQ showed a statistically significant difference in all parameters (p<0.001, p<0.001, p=0.044, p<0.001, respectively) (see Table 2).

In our study, a statistically significant difference was found in the comparison of the physical role subparameter (p=0.008), emotional well-being subparameter (p=0.014), social function subparameter (p=0.011), pain subparameter (p<0.001), and general health subparameter (p=0.002), before and after the application of the SF-36 questionnaire. This questionnaire was to evaluate daily living activities. No statistically significant difference was found in the physical, the emotional role, fatigue, and the health change sub-parameters of the SF-36 questionnaire (see Table 2).

When we looked at the evaluation of PPAQ sub-parameters, we found statistically significant differences in household/caregiving (p=0.043), sports/exercise (p<0.001), and total PA score (p<0.001). There was no statistically significant difference in the sub-dimension of occupational activities (see Table 2). Additionally, the mean pre- and post-exercise PA scores and pain scores are given in Figure 1.

Table 1. Correlation between SF-36 subdomains, pain and BMI

	BMI		MPQ total		VAS-resting		VAS-activity		VAS-nignt	
	r	p	r	p	r	p	r	p	r	p
Physical functioning	-0.080	0.613	-0.663	**<0.001***	0.190	0.228	-0.038	0.811	-0.227	0.157
Role limitations due to physical health	0.165	0.296	-0.527	**<0.001***	0.012	0.938	0.039	0.805	-0.287	0.066
Role limitations due to emotional problems	0.036	0.823	-0.253	0.106	0.203	0.198	-0.107	0.500	-0.211	0.181
Energy/fatigue	-0.230	0.143	-0.344	**0.026***	0.017	0.916	-0.030	0.849	-0.105	0.507
Emotional well-being	-0.063	0.691	-0.223	0.156	0.154	0.331	0.194	0.219	-0.129	0.414
Social functioning	0.119	0.455	-0.192	0.224	-0.058	0.717	0.224	0.154	-0.124	0.432
Pain	0.240	0.125	-0.519	**<0.001**	-0.045	0.777	-.131	0.409	-0.494	**0.001**
General health	0.368	**0.016***	0.087	0.584	0.048	0.764	0.119	0.451	0.091	0.568
Health change	-0.314	**0.043***	-0.152	0.338	-0.230	0.143	-0.167	0.290	-0.310*	**0.046***

*: Pearson correlation test, statistical significance p<0.05.
MPQ: McGill-Melzack Pain Questionnaire, VAS: Visual Analogue Scale, BMI: Body mass index, SF-36: Short Form-36

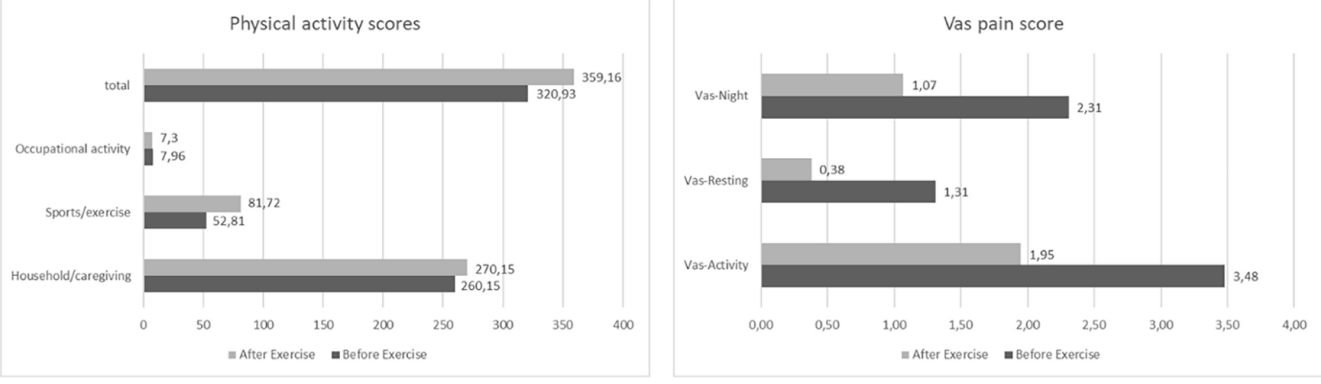

Figure 1. Mean of physical activity and pain scores before and after exercise
VAS: Visual Analogue Scale

Table 2. Comparison of post-exercise pregnancy physical activity questionnaire assessment

	Mean differences±SD	95% CI of the difference			
		Lower	Upper	t	p
VAS-activity	1.50±1.30	1.10	1.90	7.327	**<0.001***
VAS-resting	0.90±1.30	0.50	1.30	4.715	**<0.001***
VAS-night	1.20±1.50	0.80	1.70	5.310	**<0.001***
SF-36					
Physical functioning	-0.60±3.50	-1.70	0.00	-1.094	0.281
Role limitations due to physical health	-6.70±15.60	-11.60	-1.80	-2.782	**0.008***
Role limitations due to emotional problems	0.70±9.00	-2.10	3.00	494	0.624
Energy/fatigue	-1.90±6.60	-4.00	1.00	-1.864	0.069
Emotional well-being	-2.10±5.30	-4.00	0.00	-2.577	**0.014***
Social functioning	-4.80±11.70	-8.00	-1.00	-2.662	**0.011***
Pain	-9.30±11.00	-12.80	-6.00	-5.528	**<0.001***
General health	-1.80±3.50	-2.90	-0.70	-3.344	**0.002***
Health change	-2.90±10.00	-6.00	0.20	-1.901	0.064
MPQ					
Pain intensity	0.93±1.46	0.48	1.38	4.136	**<0.001***
Pain characteristic	4.43±3.09	3.47	5.39	9.302	**<0.001***
Pain-time relationship	-0.19±0.59	-0.38	0.00	-2.077	**0.044***
PPAQ					
Household/caregiving	-10.00±30.98	-19.65	-0.34	-2.092	**0.043***
Sports/exercise	-28.90±20.99	-35.44	-22.36	-8.924	**<0.001***
Occupational activity	0.67±5.84	-1.15	2.49	0.739	0.464
Total	-38.23±37.07	-49.79	-26.68	-6.684	**<0.001***

*: Paired t-test, statistical significance p<0.05.
MPQ: McGill-Melzack Pain Questionnaire, PPAQ: Pregnancy Physical Activity Questionnaire, SF-36: Short Form-36, SD: Standard deviation, CI: Confidence interval, VAS: Visual Analogue Scale

DISCUSSION

In this study, we aimed to evaluate the functional and psychological effects of exercise in pregnancy and its contribution to activities of daily living. The mean age of our participants, when their sociodemographic characteristics were compared with the literature, was 27.40 years. In our study, 69% of the pregnant women were aged 26-30, while 31% were over 30. According to the Türkiye 2018 Demographic and Health Survey (TDHS) data, the highest fertility rate was reported between the ages of 25-29 (19). In other studies, the age range of pregnant women is typically between 18 to 40 years (20). In our study, the lower age limit was 20 years. Among the pregnant women included in our study, 19.2% had chronic diseases. 4.8% of our participants had familial Mediterranean fever, 4.8% had anxiety disorder, 4.8% had anaemia, 2.4% had hypertension, 2.4% had panic attack, 2.4% had thyroid disease. In comparison, the prevalence of chronic diseases among pregnant women was reported as 8.1% in a study conducted in Ordu, and 2.4% in another study in Edirne (21). These differences may stem from regional variations and the differing characteristics of the institutions where the studies were conducted. the studies were conducted. When the education levels of our participants were analysed, it was found that 52.4% were university graduates. According to TDHS data, 59% of women in Türkiye received high school and lower level education (19). In our study, the level of education was higher than the TDHS average. In a study conducted at University of Health Sciences Türkiye, İzmir Tepecik Training and Research Hospital, 27.9% of pregnant women reported that they had previously been informed about pregnancy exercises (22). Similarly, a study conducted in Karaman reported that 61.1% of pregnant women had information

about which exercises would be beneficial during pregnancy and that they received counselling (23). In our study, the rate of counselling reached 50%, which is relatively high.

In our study, it was observed that increasing PA decreased pain levels and significantly improved quality of life. Similar findings have been reported in the literature. While PA during pregnancy can reduce the intensity and severity of pain, it is not entirely effective in preventing its occurrence (10,24). Pain during pregnancy is very common, affecting between 20% and 80% of people and various parts of the body (25). PA during pregnancy improves common pregnancy-related conditions such as LBP and PGP. It provides an easier and more comfortable pregnancy process (23). Pain during pregnancy reduces quality of life and limits daily activities. This study shows that this effect can be reduced by PA (26). Apart from the use of medication to reduce pain, PA is a treatment option with many benefits that will improve quality of life during pregnancy (27). Rodríguez-Díaz et al. (28) demonstrated that 8 weeks of Pilates exercises significantly reduced labour pain. While many studies have reported decreased levels of PA (29), further research is needed to understand the implications of this decline. Oktaviani (30) showed in a study of 40 pregnant women that Pilates exercises could effectively suppress pain. However, conflicting findings also exist. For example, one study reported no significant impact of Pilates on pain reduction (31). Mazzarino et al. (32) found insufficient evidence to support the effectiveness of Pilates in relieving LBP during pregnancy. Nevertheless, Mendo and Jorge's (33) meta-analysis stated that Pilates was beneficial against pain during pregnancy. PA, which has proven health benefits and continuously enhances individuals' quality of life, is critical for pregnant women. However, the parameters used to evaluate PA outcomes are also significant. In conclusion, given the considerable physical changes experienced during pregnancy due to the growing foetus, it may be challenging to prevent pain or discomfort. Nevertheless, we believe that increasing PA levels can effectively reduce these pains and improve overall well-being.

The benefits of exercise during pregnancy encompass both physical and mental capacities. It is known from previous studies that regular exercise in pregnant women reduces their stress and increases their self-awareness (34). However, the relationship between stress and exercise during pregnancy is not yet fully understood. Still, there is a clear association between low exercise frequency and higher levels of stress-related symptoms (35). Perales et al. (36) investigated exercise and depression reduction during pregnancy in a study of 167 women. Participants were randomly assigned to the exercise group and completed 60-minute sessions of supervised PA three times per week throughout pregnancy. Compared to the control group, women in the exercise intervention group scored significantly lower on a depression scale at the end of the study (36). In our study, it was also found that regular exercise improves mental status and emotional well-being. In our study, when the SF-36 emotional well-being section was analyzed before and after treatment, a significant difference was found. Ünver and Aylaz (37), in his 2014 study, found that exercise had a direct positive effect on the quality of life in pregnant women. In our study, no significant difference was found in the SF-36 energy parameter of pregnant women before and after exercise.

Study Limitations

This study highlights the benefits of structured exercise during pregnancy but has several limitations. The small sample size limits generalizability, and the lack of a control group makes it difficult to attribute improvements solely to the intervention. Additionally, the six-week program may not reflect long-term effects. Future studies should include larger, more diverse samples, use control groups to strengthen causal inferences, and extend intervention durations to assess sustained benefits. Incorporating psychological and social factors, objective measurements, and standardized exercise protocols can provide more comprehensive and clinically applicable findings.

CONCLUSION

This study demonstrates that regular PA during pregnancy significantly reduces pain levels and improves quality of life, supporting the integration of structured exercise programs into antenatal care. These findings align closely with the mission of advancing maternal health by promoting evidence-based clinical practices. By enhancing the physical and emotional well-being of pregnant women, exercise facilitates daily activities, reduces complications during and after delivery, and contributes to a healthier pregnancy process.

In the context of public health, the adoption of such programs can serve as a preventive strategy to address common pregnancy-related challenges, such as musculoskeletal pain and limited PA. Given the barriers many women face in accessing antenatal care, especially in resource-constrained settings, implementing structured exercise programs as part of routine prenatal care could help bridge gaps in maternal health services.

These findings can guide health authorities in integrating structured exercise programs into prenatal care policies,

promoting maternal and neonatal health. Future research should build on this foundation by exploring longer intervention periods, larger sample sizes, and diverse populations to better understand the broader implications of exercise on maternal health and its role in public health initiatives.

ETHICS

Ethics Committee Approval: The study was conducted in accordance with the rules of the Helsinki Declaration. Ethical approval for the study was obtained from the Sakarya University of Applied Sciences Ethics Committee (approval no: E-26428519-044-89268, date: 07.07.2023).

Informed Consent: The study was explained to the patients face to face, and verbal consent was obtained.

Acknowledgments

We would like to thank all participants in our study, and TÜBİTAK for their support.

FOOTNOTES

Authorship Contributions

Concept: A.Y., Ö.F., T.T., B.A., Design: A.Y., Ö.F., T.T., B.A., Y.T., Data Collection or Processing: Ö.F., T.T., B.A., Y.T., Analysis or Interpretation: A.Y., Literature Search: A.Y., Ö.F., T.T., B.A., Y.T., Writing: A.Y., Ö.F., T.T., B.A.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The study was supported by TÜBİTAK (project no: 2209-A) a project.

REFERENCES

1. Kazma JM, van den Anker J, Allegaert K, Dallmann A, Ahmadzia HK. Anatomical and physiological alterations of pregnancy. J Pharmacokinet Pharmacodyn. 2020;47:271-85.
2. Soma-Pillay P, Nelson-Piercy C, Tolppanen H, Mebazaa A. Physiological changes in pregnancy. Cardiovasc J Afr. 2016;27:89-94.
3. Marquez DX, Aguiñaga S, Vásquez PM, Conroy DE, Erickson KI, Hillman C, et al. A systematic review of physical activity and quality of life and well-being. Transl Behav Med. 2020;10:1098-109.
4. Berghella V, Saccone G. Exercise in pregnancy! Am J Obstet Gynecol. 2017;216:335-7.
5. Wadephul F, Glover L, Jomeen J. Conceptualising women's perinatal well-being: a systematic review of theoretical discussions. Midwifery. 2020;81:102598.
6. Aguilar Cordero MJ, Rodríguez Blanquet R, Sánchez García JC, Sánchez López AM, Baena García L, López Contreras G. Influencia del programa SWEP (Study Water Exercise Pregnant) en los resultados perinatales: protocolo de estudio. Nutr Hosp. 2016;33:162-76.
7. Yang Z, Zhu Z, Wang C, Zhang F, Zeng H. Association between adverse perinatal outcomes and sleep disturbances during pregnancy: a systematic review and meta-analysis. J Matern Fetal Neonatal Med. 2022;35:166-74.
8. Harskamp-van Ginkel MW, Ierodiakonou D, Margetaki K, Vafeiadi M, Karachaliou M, Kogevinas M, et al. Gestational sleep deprivation is associated with higher offspring body mass index and blood pressure. Sleep. 2020;43:zsaa110.
9. Marquez DX, Bustamante EE, Bock BC, Markenson G, Tovar A, Chasan-Taber L. Perspectives of Latina and non-Latina white women on barriers and facilitators to exercise in pregnancy. Women Health. 2009;49:505-21.
10. Kandru M, Zallipalli SN, Dendukuri NK, Linga S, Jeewa L, Jeewa A, et al. Effects of conventional exercises on lower back pain and/or pelvic girdle pain in pregnancy: a systematic review and meta-analysis. Cureus. 2023;15:e42010.
11. Santos-Rocha R. Exercise and sporting activity during pregnancy. Cham: Springer; 2019.
12. Barakat R. An exercise program throughout pregnancy: Barakat model. Birth Defects Res. 2021;113:218-26.
13. Mazzarino M, Kerr D, Morris ME. Pilates program design and health benefits for pregnant women: a practitioners' survey. J Bodyw Mov Ther. 2018;22:411-7.
14. Kocyigit H, Aydemir O, Fisek G, Olmez N, Memiş A. Validity and reliability of Turkish version of Short form 36: a study of a patients with romatoid disorder. Drug Therapy. 1999;12:102-6.
15. Aydın OU, Aydın EU. Comparison of VAS scores recorded by nurse vs surgeon: a case-control study. J Surg Med. 2018;2:74-7.
16. Tosun OC, Solmaz U, Ekin A, Tosun G, Mutlu EK, Okyay E, et al. The Turkish version of the pregnancy physical activity questionnaire: cross-cultural adaptation, reliability, and validity. J Phys Ther Sci. 2015;27:3215-21.
17. Melzack R, Katz J. The McGill Pain Questionnaire: appraisal and current status. In: Turk DC, Melzack R, editors. Handbook of pain assessment. New York (NY): The Guilford Press; 1992. p. 152-68.
18. Brett KE, Wilson S, Ferraro ZM, Adamo KB. Self-report Pregnancy Physical Activity Questionnaire overestimates physical activity. Can J Public Health. 2015;106:e297-302.
19. Hacettepe Üniversitesi Nüfus Etütleri Enstitüsü (HÜNE). 2018 Türkiye Nüfus ve Sağlık Araştırması. Ankara: Hacettepe Üniversitesi; 2019.
20. Arslan B, Arslan A, Kara S, Öngel K, Mungan MT. Risk factors for pregnancy anxiety and depression: assessment in 452 cases. Anatol J Med. 2011;21:79-84.
21. Kaya Zaman F, Özkan N, Toprak D. Depression and anxiety in pregnancy. Konuralp Med J. 2018;10:20-5.
22. Balsak D, Yıldırım Y, Avcı ME, Töz E, Gültekin E, Kayhan K, et al. The evaluation of knowledge and behavior of pregnant women living in the Aegean region about pregnancy exercises. Balkan Med J. 2007;2007:200-4.
23. Göker A, Yanıkkerem E, Baydar Ö. Determination of the physical activity and quality of life in pregnant women. Celal Bayar Univ Sag Bil Enst Derg. 2021;8:315-22.
24. Diez-Buil H, Hernandez-Lucas P, Leirós-Rodríguez R, Echeverría-García O. Effects of the combination of exercise and education in the treatment of low back and/or pelvic pain in pregnant women: systematic review and meta-analysis. Int J Gynaecol Obstet. 2024;164:811-22.
25. Sward L, Manning N, Murchison AB, Ghahremani T, McCaulley JA, Magann EF. Pelvic girdle pain in pregnancy: a review. Obstet Gynecol Surv. 2023;78:349-57.
26. McDonald HM, Sherman KA, Kasparian NA. Factors associated with psychological distress among Australian women during pregnancy. Pers Individ Dif. 2021;172:110577.

27. Aparicio VA, Marín-Jiménez N, Flor-Alemany M, Acosta-Manzano P, Coll-Risco I, Baena-García L. Effects of a concurrent exercise training program on low back and sciatic pain and pain disability in late pregnancy. Scand J Med Sci Sports. 2023;33:1201-10.
28. Rodríguez-Díaz L, Ruiz-Frutos C, Vázquez-Lara JM, Ramírez-Rodrigo J, Villaverde-Gutiérrez C, Torres-Luque G. Effectiveness of a physical activity programme based on the Pilates method in pregnancy and labour. Enferm Clin. 2017;27:271-7.
29. Adrien N. Active lifestyle among pregnant women in Burundi. Am J Sports Sci. 2018;6:74.
30. Oktaviani I. Pilates workouts can reduce pain in pregnant women. Complement Ther Clin Pract. 2018;31:349-51.
31. Hall DC, Kaufmann DA. Effects of aerobic and strength conditioning on pregnancy outcomes. Am J Obstet Gynecol. 1987;157:1199-203.
32. Mazzarino M, Kerr D, Wajswelner H, Morris ME. Pilates method for women's health: systematic review of randomized controlled trials. Arch Phys Med Rehabil. 2015;96:2231-42.
33. Mendo H, Jorge M. Pilates method and pain in pregnancy: a systematic review and metanalysis. Braz J Pain. 2021;4:276-82.
34. Beattie J, Hall H, Biro MA, East C, Lau R. Effects of mindfulness on maternal stress, depressive symptoms and awareness of present moment experience: a pilot randomised trial. Midwifery. 2017;50:174-83.
35. Sinclair I, St-Pierre M, Elgbeili G, Bernard P, Vaillancourt C, Gagnon S, et al. Psychosocial stress, sedentary behavior, and physical activity during pregnancy among Canadian women: relationships in a diverse cohort and a nationwide sample. Int J Environ Res Public Health. 2019;16:5150.
36. Perales M, Refoyo I, Coteron J, Bacchi M, Barakat R. Exercise during pregnancy attenuates prenatal depression: a randomized controlled trial. Eval Health Prof. 2015;38:59-72.
37. Ünver H, Aylaz R. The effect of physical exercise on quality of life in pregnancy. Turk Klin J Gynecol Obstet. 2017;27:71-8.

Medical Journal of Bakırköy

Med J Bakirkoy 2025;21(3):240-245
DOI: 10.4274/BMJ.galenos.2025.2024.11-10

Research

Assessment of Vitamin D Sufficiency in Different Age Groups of Children in Türkiye Using the Bhattacharya Method

Türkiye'deki Farklı Yaş Gruplarındaki Çocukların Vitamin D Yeterliliğinin Bhattacharya Yöntemi ile Değerlendirilmesi

Fatma Hande Karpuzoğlu, Serhat Kılınç, Abdurrahman Fatih Aydın

İstanbul University, İstanbul Faculty of Medicine, Department of Medical Biochemistry, İstanbul, Türkiye

ABSTRACT

Objective: Vitamin D plays a role in regulating mineral balance, particularly calcium and phosphate, and in promoting bone health. It is essential to have reliable age-specific reference values for vitamin D in order to make accurate clinical decisions. This study aimed to establish reference ranges for 25-hydroxyvitamin D [25(OH)D] levels in children aged 1-19 years in Türkiye using the Bhattacharya method and to assess vitamin D status across different age groups.

Methods: Data from 243,906 [25(OH)D] measurements, collected from hospital laboratory records between 2019 and 2023, were analyzed.

Results: The reference ranges for serum [25(OH)D] levels vary by age and gender. For children aged 1-2 years, the range is 24.5-56.6 ng/mL, decreasing progressively to 16-42.9 ng/mL for ages 7-10 years. For women aged 11-15 years, the range is 14.8-35.5 ng/mL, and for ages 16-19 years, it is 15-39 ng/mL. For male patients, the ranges are slightly higher: 15.6-40.7 ng/mL for ages 11-15 years and 15.8-39.3 ng/mL for ages 16-19 years. The study revealed that the prevalence of [25(OH)D] deficiency levels below 12 ng/mL (considered deficient based on the institute of medicine criteria) increased with age among children and adolescents. This trend underscores a notable rise in deficiency rates after early childhood, peaking during adolescence.

Conclusion: The results provide significant insights into the vitamin D status of the pediatric population in Türkiye. The findings emphasise a decline in vitamin D levels with age and indicate that the highest prevalence of deficiency is seen in adolescents.

Keywords: Parathyroid hormone (PTH), vitamin D, 25-hydroxyvitamin D [25(OH)D], Bhattacharya method, indirect method, reference interval

ÖZ

Amaç: D vitamini, başta kalsiyum ve fosfat olmak üzere mineral dengesinin düzenlenmesinde ve kemik sağlığının geliştirilmesinde rol oynamaktadır. Doğru klinik kararlar verebilmek için D vitamini için yaşa özgü güvenilir referans değerlere sahip olmak önemlidir. Bu çalışmanın amacı, Bhattacharya yöntemini kullanarak Türkiye'de 1-19 yaş arası çocuklarda 25-hidroksivitamin D [25(OH)D] düzeyleri için referans aralıkları oluşturmak ve farklı yaş gruplarında D vitamini durumunu değerlendirmektir.

Gereç ve Yöntem: 2019-2023 yılları arasında hastane laboratuvar kayıtlarından toplanan 243.906 [25(OH)D] ölçümünden elde edilen veriler analiz edilmiştir.

Bulgular: Serum [25(OH)D] düzeyleri için referans aralıkları yaşa ve cinsiyete göre değişmektedir. Bir ile iki yaş arası çocuklar için bu aralık 24,5-56,6 ng/mL'dir ve 7-10 yaş arasında kademeli olarak 16-42,9 ng/mL'ye düşmektedir. On bir ile on beş yaş arası kadınlar için aralık 14,8-35,5 ng/mL ve 16-19 yaş için 15-39 ng/mL'dir. Erkek hastalar için aralıklar biraz daha yüksektir: 11-15 yaş için 15,6-40,7 ng/mL ve 16-19 yaş için 15,8-39,3 ng/mL. Çalışma, 12 ng/mL'nin altındaki (tıp enstitüsü kriterlerine göre eksik kabul edilen) [25(OH)D] eksikliği seviyelerinin yaygınlığının çocuklar ve ergenler arasında yaşla birlikte arttığını ortaya koymuştur. Bu eğilim, erken çocukluk döneminden sonra ergenlik döneminde zirve yapan eksiklik oranlarında kayda değer bir artışın altını çizmektedir.

Sonuç: Sonuçlar, Türkiye'deki pediatrik nüfusun D vitamini durumuna ilişkin önemli bilgiler sağlamaktadır. Bulgular, yaşla birlikte D vitamini seviyelerinde bir düşüş olduğunu vurgulamakta ve en yüksek eksiklik prevalansının ergenlerde görüldüğüne işaret etmektedir.

Anahtar Kelimeler: Paratiroid hormonu (PTH), vitamin D, 25-hidroksivitamin D [25(OH)D], Bhattacharya yöntemi, indirekt yöntem, referans aralığı

Address for Correspondence: Fatma Hande Karpuzoğlu, Asst. Prof., İstanbul University, İstanbul Faculty of Medicine, Department of Medical Biochemistry, İstanbul, Türkiye
E-mail: handekarpuzoglu@yahoo.com.tr **ORCID ID:** orcid.org/0000-0001-9603-5838

Cite as: Karpuzoğlu H, Kılınç S, Aydın AF. Assessment of vitamin D sufficiency in different age groups of children in Türkiye using the Bhattacharya method. Med J Bakirkoy. 2025;21(3):240-245

Received: 28.11.2024
Accepted: 11.02.2025
Publication Date: 03.09.2025

INTRODUCTION

Vitamin D plays a crucial role in regulating mineral metabolism, maintaining calcium and phosphate homeostasis, and promoting bone and overall health (1). Accurate interpretation of clinical decisions requires reliable reference values for vitamin D specific to age (2,3). Parathyroid hormone (PTH) enhances calcium reabsorption and inhibits phosphate reabsorption in the kidneys. Furthermore, it enhances the synthesis of 1,25-dihydroxyvitamin D, which subsequently elevates the gastrointestinal absorption of calcium (1,4). Vitamin D is a fat-soluble vitamin that plays a role in regulating the balance of calcium and phosphorus and in enhancing bone development (5). Vitamin D deficiency increases the risk of several health problems, including rickets, bone fractures, various types of cancer, cardiovascular disease, diabetes, autoimmune and metabolic disorders, infections due to immune-system dysfunction, and even certain neuropsychiatric disorders (6). It is therefore crucial to ascertain the levels of vitamin D present in the population. The most effective laboratory test for the assessment of vitamin D levels in the body is the measurement of 25-hydroxyvitamin D [25(OH)D] (7).

Reference range studies allow us to comprehend the normal distribution and variations of laboratory test results within a population, which contributes to the evaluation of health status. This, in turn, assists in making more precise clinical decisions and diagnoses (8).

The EP28-A3C guideline proposes that population-based reference intervals can be established through either direct sampling of individuals from a healthy population (termed the direct sampling approach) or indirect sampling, also known as data mining techniques (9).

The results of data mining analysis demonstrate reference intervals that are highly comparable to those obtained from direct reference interval studies for a diverse range of analytes (10). Bhattacharya is a data mining technique that can be used to analyse large databases of patient results. It facilitates the stratification of subjects into multiple age and sex partitions without compromising statistical power, following the application of rigorous exclusion criteria (11,12).

The aim of this study was to determine reference intervals for [25(OH)D] levels in children of different age groups in Türkiye using the Bhattacharya method. Therefore, this study aims to evaluate the vitamin D status in these age groups, and provide insights into its distribution and possible deficiencies.

METHODS

Study Design and Data Collection

In our study, we calculated the reference range for [25(OH)D] (n=243,906) using hospital-based data from raw laboratory records registered in the laboratory information system (LIS) of the Central Biochemistry Laboratory of İstanbul University, İstanbul Faculty of Medicine between 2019 and 2023. The study was not designed with a specific protocol but used pre-selected patients, and specific exclusion criteria were applied for the included patients (4). Only data of participants aged between 1 and 19 years were included. The study excluded data from patients older than 19 years (n=100,102). Patients below one year old were excluded from the study due to insufficient laboratory data (n=62). Hospitalized and follow-up patients (n=51,021) were also excluded from the study. Additionally, data exceeding the detection limits of the analysis methods ([25(OH)D] levels >1000 or <5 ng/mL) were excluded from the study without statistical analysis (n=1,383). Vitamin D, calcium, and PTH measurements were requested together in the trial, and patients with calcium and PTH levels outside the reference range were excluded (n=21,170). Measurement of [25(OH)D] plasma levels was performed using a high-performance liquid chromatography (HPLC) device (Thermo Ultimate 3000, USA) with a ultraviolet detector. The vitamin D clinical research HPLC kit (Recipe Chemicals Instruments Munich, Germany) was employed for the analysis.

The [25(OH)D] test is routinely conducted in our laboratory. Prior to analysing patients' samples, an internal quality control test is performed using two samples with known specific concentrations representing normal and pathological conditions.

Ethical Approval and Informed Consent

The study was conducted in accordance with the ethical principles of the Declaration of Helsinki. Approval was obtained from the İstanbul University, İstanbul Faculty of Medicine Clinical Research Ethics Committee (approval no: 13, date: 12.07.2024). As this was a retrospective study using anonymised data from the LIS, informed consent was not required.

Statistical Analysis

Visual assessment of the distribution of plasma [25(OH)D] levels was performed using a boxplot graph (Figure 1). One-way ANOVA was used for partitioning based on age and sex comparisons, with results showing $p<0.05$ being considered statistically significant (13).

The reference ranges for vitamin D hormone were calculated using the Bhattacharya method. The data were sorted into

subgroups in ascending order with equal intervals. The total frequency of the data was then divided into 20 equally spaced classes, with a bin size of 2.0, for [25(OH)D]. The logarithm of the counts (represented as fi) and the Δlog (fi) values were calculated for transitions from one bin to another. To minimize the influence of random fluctuations, we utilized the 5-point Savitzky-Golay smoothing procedure, resulting in a smoother curve. Additionally, following the suggestion of Oosterhuis et al. (12), we introduced a weighting factor to improve the precision of the Bhattacharya analysis. After plotting the derivative of the frequency of the measured values against the logarithm of the concentration, we calculated the intercept (a) and slope (b) of the linear relationship. This was done only for the linear portion of the relationship where the coefficient of determination (R^2) exceeded 0.99 (as shown in Figure 2) (11,13). However, if the fit was not satisfactory, we applied a Box-Cox transformation. The parameter λ for the transformation was selected based on the best fit to the data using https://www.statology.org/box-cox-transformation-excel/ (Figure 3). The midpoint, upper, and lower reference limits (URL and LRL) were calculated and a data plot was generated using Microsoft Excel.

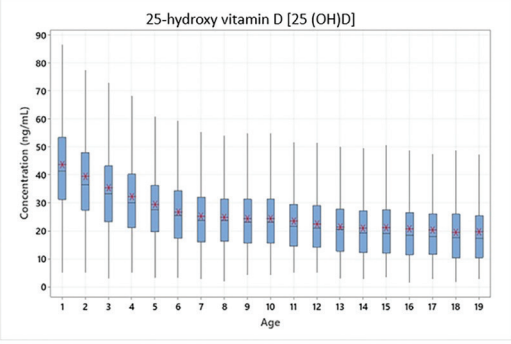

Figure 1. Distribution of serum [25(OH)D] concentration according to age. The median values are shown as horizontal lines and mean values are indicated with asterisks
[25(OH)D]: 25-hydroxyvitamin D

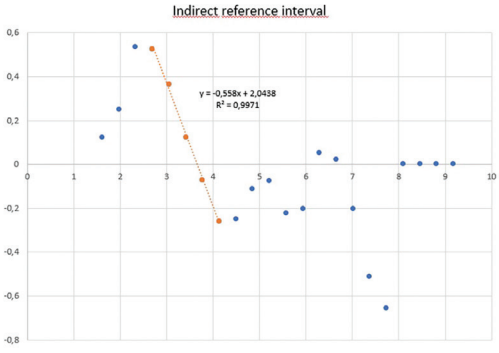

Figure 2. [25 (OH)D] Bhattagram dataset for females aged 11-14. The Gaussian component of the dataset was calculated using the slope and y-intercept of the line of best fit for these data points
[25(OH)D]: 25-hydroxyvitamin D

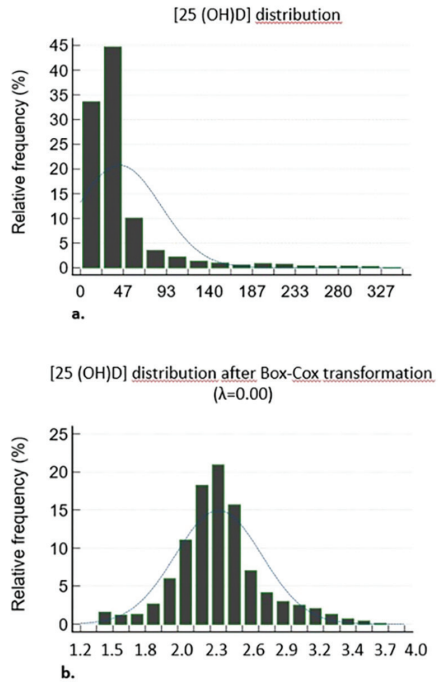

Figure 3. The distribution of a skewed [25 (OH)D] before (a) and after (b) applying the Box-Cox transformation
[25(OH)D]: 25-hydroxyvitamin D

RESULTS

After applying exclusion criteria, 70,168 vitamin D data points were included in the study. Vitamin D data from 29,921 males and 40,247 females aged 1 to 19 years were used in the study. The reference intervals for the [25(OH)D] test were calculated jointly for males and females across the first 10 age groups. For the age groups 11-14 and 15-19, separate calculations were conducted for both sexes.

The reference ranges for serum [25(OH)D] levels in children aged 1-2 years are 24.5-56.6 ng/mL and 22.4-56.3 ng/mL in spring/summer and autumn/winter, respectively. For 3-year-olds, the ranges are 21.5-55.7 ng/mL and 20.3-55.6 ng/mL; for 4-year-olds, 19.1-49.7 ng/mL and 18.6-48.2 ng/mL; for 5-year-olds, 17.0-47.3 ng/mL and 16.8-47.8 ng/mL; for 6-year-olds, 16.6-48.0 ng/mL and 16.3-46.2 ng/mL; and for children aged 7-10 years, 16.0-42.9 ng/mL and 15.7-43.0 ng/mL, respectively.

For females aged 11-15 years, the reference ranges are 14.8-35.5 ng/mL and 14.5-34.2 ng/mL in spring/summer and autumn/winter and for those aged 16-19 years, 15.0-39.0 ng/mL and 15.0-38.6 ng/mL, respectively.

In males aged 11-15 years, the ranges are 15.6-40.7 ng/mL in spring/summer and 15.4-40.3 ng/mL in autumn/winter, and for those aged 16-19 years, 15.8-39.3 ng/mL in spring/summer and 15.5-39.8 ng/mL in autumn/winter, respectively (Table 1).

Table 1. Reference ranges for the 25-hydroxyvitamin D [25 (OH)D], stratified by age, sex and season

[25(OH)D] (ng/mL)						
Age groups	**Gender**	**Season**	**n**	**Midpoint**	**Lower limit (LL)**	**Upper limit (UL)**
1-2 age	Both	Spring/Summer	1313	40.5	24.5	56.6
		Autumn/Winter	1402	39.4	22.4	56.3
3 age	Both	Spring/Summer	1232	38.6	21.5	55.7
		Autumn/Winter	1438	37.9	20.3	55.6
4 age	Both	Spring/Summer	1554	34.4	19.1	49.7
		Autumn/Winter	1611	33.4	18.6	48.2
5 age	Both	Spring/Summer	1589	32.2	17.0	47.3
		Autumn/Winter	1712	32.3	16.8	47.8
6 age	Both	Spring/Summer	1479	32.3	16.6	48
		Autumn/Winter	1519	31.25	16.3	46.2
7-10 age	Both	Spring/Summer	3810	29.5	16.0	42.9
		Autumn/Winter	4356	29.4	15.7	43.0
11-15 age	Female	Spring/Summer	3014	24.0	14.8	35.5
		Autumn/Winter	5112	24.4	14.5	34.2
11-15 age	Male	Spring/Summer	3788	28.2	15.6	40.7
		Autumn/Winter	3987	27.9	15.4	40.3
16-19 age	Female	Spring/Summer	6500	27.0	15.0	39
		Autumn/Winter	7310	26.8	15.0	38.6
16-19 age	Male	Spring/Summer	7407	27.8	15.8	39.3

DISCUSSION

This study establishes sex- and age-specific reference intervals for vitamin D in children and adolescents in Türkiye. The results show that vitamin D levels vary according to both age and sex in individuals aged 1-18 years. In addition, the study evaluates vitamin D status in the child population using the calculated reference intervals.

Vitamin D levels are influenced by several factors, including geographical location, seasonal variation, ethnicity, skin pigmentation, clothing, diet and sunscreen use (14). Vitamin D deficiency should be assessed using established cut-off values rather than reference ranges. The 2010 Institute of Medicine (IOM) guidelines define vitamin D deficiency as less than 12 ng/mL (30 nmol/L), vitamin D insufficiency as 12-19 ng/mL, and vitamin D sufficiency as 20 ng/mL (50 nmol/L) and above (15). The recommendations of the Endocrine Society Committee categorise vitamin D levels as follows: levels below 20 ng/mL are considered vitamin D deficiency, levels between 21 and 29 ng/mL indicate vitamin D insufficiency, and levels at or above 30 ng/mL are considered sufficient (with a preferred range of 40-60 ng/mL). In addition, levels above 150 ng/mL are considered indicative of vitamin D toxicity (16).

In our study, we established reference ranges for vitamin D [25(OH)D] to assess vitamin D status in the pediatric population. These reference ranges were calculated separately for each age group between 1 and 6 years. In the first three years, we found that vitamin D levels were within the IOM target range. The Turkish Ministry of Health recommends providing 400 IU/day of vitamin D to all infants to prevent deficiency and promote bone health. This recommendation, established in 2005, should be followed until the child is at least one year old, preferably extending until the age of three (17). We believe that the Ministry of Health's vitamin D supplementation program for infants has been successfully implemented in this age group. However, in our study, we found that [25(OH)D] levels decreased in children aged three and older. Our study found that vitamin D levels decrease with age, while PTH hormone levels increase. This trend may be due to increased vitamin D supplementation during early childhood. PTH and vitamin D interact in a complex feedback system to regulate calcium levels. PTH strongly stimulates the synthesis of vitamin D in the kidney, while vitamin D exerts a negative feedback effect on the release of PTH (1,18).

In particular, we observed that the reference range was lowest in the 11-15 age group. When evaluating serum

[25(OH)D] levels according to age groups, we obtained the following results: for one year, 2%; for two years, 3.3%; for three years, 5%; for four years, 7.4%; for five years, 7.8%; for six years, 10%, for 7-10 years, 13%; for 11-15 years, 25.8%; for 16-18 years, 23.9% of the results indicated [25(OH)D] levels below 12 ng/mL, which we considered deficient according to the IOM. Similarly, Yakarış et al. (19) demonstrated a decline in vitamin D levels with age in children between 2012 and 2019. The study conducted by Yeşiltepe-Mutlu et al. (20) on [25(OH)D] levels in the Turkish population found that the distribution of [25(OH)D] levels for the age range of 1-18 years is consistent with our findings. Similarly, we observed that participants aged 1-10 years had higher [25(OH)D] levels, compared to other age groups. However, in contrast to their study, we calculated separate reference ranges for each age group in the first 6 years. As in our country, vitamin D deficiency in chilhood is common worlwide (21).

In our study, we investigated the seasonal variation of [25(OH)D] levels. Our findings revealed that levels exhibited seasonal fluctuations. We observed that [25(OH)D] levels were highest within the reference range during the summer, and lowest during the winter and spring seasons (Figure 4). Previous studies have also shown seasonal variations, and these findings are consistent with our study (20,22,23).

The Bhattacharya method assumes that most patients who have come to the hospital are not actually ill, and suggests that statistical techniques can be used to distinguish between healthy and pathological data (10-12).

Baadenhuijsen and Smit (11) state that the Bhattacharya technique requires a significant amount of data to be applied effectively. This is necessary to avoid large statistical fluctuations and to identify the linear part of the distribution. The authors recommend collecting more than 1500 values for each analyte. The Bhattacharya algorithm assumes that a significant portion of an unselected general population can be regarded as "normal" It also assumes that there is only a partial overlap between the "healthy" portion and the abnormal (either high or low) portion (12).

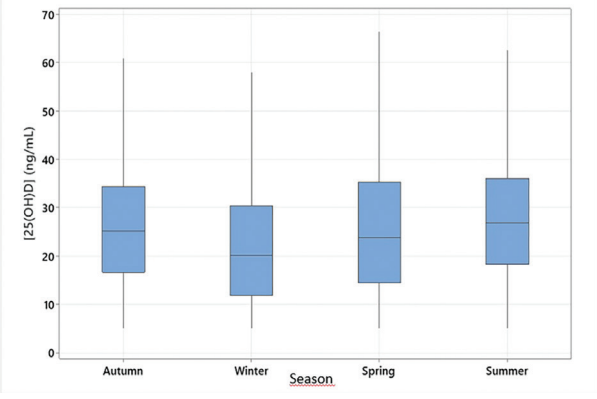

Figure 4. Seasonal distribution of [25(OH)D] levels
[25(OH)D]: 25-hydroxyvitamin D

The Bhattacharya method is a graphical technique used to determine a Gaussian distribution in a dataset. The linear segment of the obtained graph, where the R^2 value should be greater than 0.99, includes the data to be used for reference value analysis. The data between the values that establish the lower and upper limits of the curve are utilized for reference value analysis (10).

Study Limitations

This study has some limitations. Firstly, we did not use a reference population to establish reference ranges, so our study lacks comparison between reference intervals created by both direct and indirect methods. In addition, due to the unavailability of complete patient data, we could not apply all intended exclusion criteria. In particular, we did not exclude patients without a registered diagnosis from the study.

CONCLUSION

The use of the Bhattacharya method proved an effective means of calculating reliable reference intervals, distinguishing between healthy and pathological data, and identifying critical trends in [25(OH)D] levels. These findings highlight the importance of age-specific reference intervals for paediatric populations, thus enhancing the accuracy of clinical evaluations related to bone health and mineral metabolism. Furthermore, the findings of this study emphasize the necessity for effective monitoring and supplementation strategies to address vitamin D deficiency in children and adolescents.

ETHICS

Ethics Committee Approval: Approval was obtained from the İstanbul University, İstanbul Faculty of Medicine Clinical Research Ethics Committee (approval no: 13, date: 12.07.2024).

Informed Consent: Retrospective study.

FOOTNOTES

Authorship Contributions

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

Conflict of Interest: No conflict of interest was declared by the authors.

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

REFERENCES

1. Khodadadiyan A, Rahmanian M, Shekouh D, Golmohammadi M, Ghaedi A, Bazrgar A, et al. Evaluating the effect of vitamin D supplementation on serum levels of 25-hydroxy vitamin D, 1,25-dihydroxy vitamin D, parathyroid hormone and renin–angiotensin–aldosterone system: a systematic review and meta-analysis of clinical trials. BMC Nutr. 2023;9:132.
2. Yalla N, Bobba G, Guo G, Stankiewicz A, Ostlund R. Parathyroid hormone reference ranges in healthy individuals classified by vitamin D status. J Endocrinol Invest. 2019;42:1353-60.
3. Farrell C-JL, Nguyen L, Carter AC. Parathyroid hormone: data mining for age-related reference intervals in adults. Clin Endocrinol (Oxf). 2018;88:311-7.
4. Leung EKY. Parathyroid hormone. Adv Clin Chem. 2021;101:41-93.
5. Wimalawansa SJ, Razzaque MS, Al-Daghri NM. Calcium and vitamin D in human health: hype or real? J Steroid Biochem Mol Biol. 2018;180:4-14.
6. Berretta M, Quagliariello V, Bignucolo A, Facchini S, Maurea N, Di Francia R, et al. The multiple effects of vitamin D against chronic diseases: from reduction of lipid peroxidation to updated evidence from clinical studies. Antioxidants. 2022;11:1090.
7. Bilezikian JP, Formenti AM, Adler RA, Binkley N, Bouillon R, Castro ML, et al. Vitamin D: dosing, levels, form, and route of administration: does one approach fit all? Rev Endocr Metab Disord. 2021;22:1201-18.
8. Ozarda Y. Reference intervals: current status, recent developments and future considerations. Biochem Med. 2016;26:5-16.
9. Clinical and Laboratory Standards Institute (CLSI). Defining, establishing, and verifying reference intervals in the clinical laboratory; approved guideline—third edition. CLSI document EP28-A3c. Wayne, PA: CLSI; 2008.
10. Jones GRD, Haeckel R, Loh TP, Sikaris K, Streichert T, Katayev A, et al. Indirect methods for reference interval determination: review and recommendations. Clin Chem Lab Med. 2019;57:20-9.
11. Baadenhuijsen H, Smit JC. Indirect estimation of clinical chemical reference intervals from total hospital patient data: application of a modified Bhattacharya procedure. J Clin Chem Clin Biochem. 1985;23:829-39.
12. Oosterhuis WP, Modderman TA, Pronk C. Reference values: Bhattacharya or the method proposed by the IFCC? Ann Clin Biochem. 1990;27:359-65.
13. Martinez-Sanchez K, Marques-Garcia F, Ozarda Y, Blanco A, Brouwer N, Canalias F, et al. Big data and reference intervals: rationale, current practices, harmonization and standardization prerequisites and future perspectives of indirect determination of reference intervals using routine data. Adv Lab Med. 2020;2:9-25.
14. Gözüoğlu G, Gökçe Ş, Aslan A, Kurugöl Z, Koturoğlu G. Seasonal variation of vitamin D via several parameters in adolescents. J Pediatr Res. 2018:171.
15. Vieth R, Holick MF. The IOM-endocrine society controversy on recommended vitamin D targets: in support of the endocrine society position. InVitamin D. 2018:1091-107.
16. Michael MD, Holick F. Vitamin D status: measurement, interpretation, and clinical application. Ann Epidemiol. 2009;19:73-8.
17. Gül İ, Gür E, Erener Ercan T, Can G. The effect of vitamin D prophylaxis on 25-OH vitamin D levels in children. Turk Arch Pediatr. 2021;56:618-23.
18. Geserick M, Vogel M, Eckelt F, Schlingmann M, Hiemisch A, Baber R, et al. Children and adolescents with obesity have reduced serum bone turnover markers and 25-hydroxyvitamin D but increased parathyroid hormone concentrations results derived from new pediatric reference ranges. Bone. 2020;132:115-24.
19. Yakarış AB, Öner C, Şimşek EE, Çetin H. Frequency of vitamin D deficiency in children admitted to pediatrics outpatient clinics: a hospital-based study. J Turk Fam Physician. 2022;13:12-20.
20. Yeşiltepe-Mutlu G, Aksu ED, Bereket A, Hatun, Ş. Vitamin D status across age groups in Turkey: results of 108,742 samples from a single laboratory. J Clin Res Pediatr Endocrinol. 2020;12:248-55.
21. van Schoor N, de Jongh R, Lips P. Worldwide vitamin D status. Feldman and Pike's Vitamin D. 2024:47-75.
22. Çaykara B, Öztürk G, Mutlu HH. Vitamin D deficiency to the age and season of Turkish patients admitted to family medicine: a retrospective hospital-based study in Istanbul. Haydarpaşa Numune Med J. 2022;62:290.
23. Aşpdemır M. Vitamin D deficiency status in Turkey: a meta-analysis. Int J Med Biochem. 2019;2:1-3.

Medical Journal of Bakırköy

Med J Bakirkoy 2025;21(3):246-251
DOI: 10.4274/BMJ.galenos.2025.2024.11-15

Research

A Comprehensive Analysis of Prognostic Factors in Head and Neck Squamous Cell Carcinoma: Experience from a Single-center

Baş ve Boyun Skuamöz Hücreli Kanserlerinde Prognostik Faktörlerin Kapsamlı Analizi: Tek Merkez Deneyimi

Gülşen Pınar Soydemir, Elif Eda Özer, Meltem Kırlı Bölükbaş, Metin Figen, Esengül Koçak Üzel

University of Health Sciences Türkiye, Bakırköy Dr. Sadi Konuk Training and Research Hospital, Clinic of Radiation Oncology, İstanbul, Türkiye

ABSTRACT

Objective: The aim of this study is to evaluate the prognostic factors of adjuvant or definitive radiotherapy (RT) and/or chemoradiotherapy (CRT) in patients with squamous cell head and neck (HN) tumor treated in our clinic.

Methods: The prognostic factors in the patients who were treated between February 2017 and May 2021 were retrospectively evaluated. A total of 78 patients diagnosed with HN cancer were included in the study. RT was applied to tumor/tumor lodge ± lymphatics at a dose of 54-70 gray. The prognostic factors, side effects, and overall survival in patients were noted and evaluated.

Results: Of a total of 78 patients, 15 (19.2%) were female and 63 (80.8%) were male. The most common tumor location in the patients was larynx (53.8%), followed by oral cavity (24.4%), and oropharynx (15.4%). Twenty-seven (34.6%) patients were in the T2 stage. Additionally, most of the patients were N0 (39 patients, 50%), and 29 (37.2%) patients were N2. Forty-three (55.1%) patients underwent surgery. Forty-three (55.1%) patients received adjuvant RT. Concomitant chemotherapy with RT was administered to 46 (59%) patients. In all groups, significant differences were found in hemoglobin and platelets before and after RT. Borderline significant in white blood cells. N stage of the tumor, smoking habit, and tumor localization were found to be significant for survival.

Conclusion: In the treatment of HN cancers, disease-free survival and a functional life in which organs at risk are protected as much as possible are aimed. RT/CRT is a highly toxic, long-term, and organ-preserving therapy. One of the main goals is to provide a survival advantage by increasing local control and protecting patients from side effects.

Keywords: Head and neck tumors, squamous cell cancer, concurrent chemoradiotherapy, prognostic factors

ÖZ

Amaç: Bu çalışmanın amacı, kliniğimizde skuamöz hücreli baş ve boyun (BB) tümörü tanısı ile tedavi edilen hastalarda adjuvan veya definitif radyoterapi (RT) ve/veya kemoradyoterapinin (KRT) prognostik faktörlerini değerlendirmektir.

Gereç ve Yöntem: Şubat 2017-Mayıs 2021 tarihleri arasında tedavi edilen hastalarda prognostik faktörler geriye dönük olarak değerlendirildi. Çalışmaya BB kanserli toplam 78 hasta dahil edildi. Tümör/tümör loju ± lenfatiklere 54-70 gray dozda radyasyon tedavisi uygulandı. Hastalardaki prognostik faktörler, yan etkiler ve genel sağkalım not edildi ve değerlendirildi.

Bulgular: Toplam 78 hastanın 15'i (%19,2) kadın, 63'ü (%80,8) erkekti. Hastalarda en sık tümör yerleşimi larinks (%53,8) iken, bunu oral kavite (%24,4) ve orofarenks (%15,4) izlemekteydi. Yirmi yedi (%34,6) hasta T2 evresindeydi ve hastaların çoğu N0 (39 hasta, %50) ve 29 (%37,2) hasta N2 idi. Kırk üç (%55,1) hasta ameliyat edildi. Kırk üç (%55,1) hastaya adjuvan RT uygulandı. Kırk altı (%59) hastada RT ile eş zamanlı kemoterapi uygulandı. Tüm gruplarda RT öncesi ve sonrası hemoglobin ve trombositlerde anlamlı farklılıklar bulundu. Beyaz kan hücrelerinde sınırda anlamlılık bulundu. Tümörün N evresi, sigara içme alışkanlığı ve tümör lokalizasyonu sağkalım için önemli bulundu.

Sonuç: BB kanserlerinin tedavisinde hastalıksız sağkalım ve risk altındaki organların mümkün olduğunca korunduğu fonksiyonel bir yaşam amaçlanmaktadır. RT/KRT oldukça toksik, uzun süreli, organ koruyucu bir tedavidir. Ana hedeflerden biri yan etkilerle lokal kontrolü artırarak sağkalım avantajı sağlamak ve hastaları yan etkilerden korumaktır.

Anahtar Kelimeler: Baş ve boyun tümörleri, skuamöz hücreli kanser, kemoradyoterapi, prognostik faktörler

Address for Correspondence: Gülşen Pınar Soydemir Assoc. Prof., University of Health Sciences Türkiye, Bakırköy Dr. Sadi Konuk Training and Research Hospital, Clinic of Radiation Oncology, İstanbul, Türkiye
E-mail: gulpin28@gmail.com **ORCID ID:** orcid.org/0000-0001-7758-8760

Cite as: Soydemir GP, Özer EE, Kırlı Bölükbaş M, Figen M, Koçak Üzel E. A comprehensive analysis of prognostic factors in head and neck squamous cell carcinoma: experience from a single-center. Med J Bakirkoy. 2025;21(3):246-251

Received: 05.12.2024
Accepted: 11.02.2025
Publication Date: 03.09.2025

INTRODUCTION

Head and neck (HN) squamous cell carcinoma (SCC) is the seventh most common malignancy (1). There are 550,000 patients diagnosed with HN tumors (HNTs) worldwide each year; 300,000 of these patients lose their lives. Around 90% of all HNTs are SCC (2). The disease most commonly occurs as SCC that spreads from the mucosal lining of the upper respiratory-digestive tract, typically in the oral cavity, larynx, or pharynx (3). The disease is highly correlated with standard of living. The incidence increases as alcohol and/or cigarette consumption increases. Sixty percent of patients with HNT test positive for human papilloma virus (4).

A multidisciplinary approach is important to decide on treatment. The choice of treatment is based on histopathological features, tumor localization, and patient-related factors. Especially in its early stages, radiotherapy (RT) and surgery, alone or in combination, can eliminate regional disease (3). In terms of survival, there is no difference between these two methods (5).

SCC HNT with stage 3-4 are at high risk for regional recurrence and distant metastasis. Usually, combined treatment modalities are used. Treatment modalities include surgery, RT and chemotherapy (CT). Generally, RT is used after surgery or definitively, with or without CT. CT can be used as an induction. Surgery and/or RT following induction CT, definitive chemoradiotherapy (CRT), or postoperative therapy have been shown to improve local control and survival (6,7).

In this study, we aimed to evaluate the prognostic factors of squamous cell HN cancers, which were treated with adjuvant or definitive RT and/or CRT in our clinic.

METHODS

This study was approved by the Clinical Research Ethics Committee of the University of Health Sciences Türkiye, Bakırköy Dr. Sadi Konuk Training and Research Hospital (approval no: 2022-09, date: 09.05.2022).

A total of 78 patients diagnosed with HN cancer, who were treated in our clinic between February 2017 and May 2021, were included in the study. This is a retrospective study, and all participants provided written informed consent.

Patient inclusion criteria were as follows: oriented and cooperative patients aged 18-80 years diagnosed with HNT. In the study, the criteria for administering adjuvant RT and/or CT, determining CT schemes, assessing lymph node involvement, evaluating extracapsular extension (ECE), surgical margin positivity/proximity, and addressing other histopathological risk factors [such as lymphatic vessel invasion (LVI), perineural invasion (PNI), and blood vessel invasion] were examined.

During the pre-RT evaluation process, history and physical examinations of the patients were performed. Positron emission tomography (PET) images, magnetic resonance imaging (MRI), and/or computed tomography images were taken. Complete blood count and blood biochemistry were evaluated for all patients before treatment. Laboratory tests were repeated weekly during treatment.

To create a RT plan, each patient underwent a special thermoplastic mask fixation in the supine position. The section thickness was taken as 2.5 mm for tomography images. Patients' planning computed tomography images were fused with pretreatment MRI and/or PET-computed tomography images for tumor localization and nodal involvement. Lymphatic areas were determined according to the disease indication in the Radiation Therapy Oncology Group (RTOG) HN atlas.

Monte Carlo planning analysis with 6 megavolt photon energy was used. Volumetric arc therapy plans were made on a linear accelerator device for RT. In a total of five fractions per week from Monday to Friday, with a daily fraction dose of 2 gray (Gy), an RT dose of 54 Gy was administered to prophylactic neck lymphatics, 60 Gy to involved neck lymphatics, and 66-70 Gy to tumor and/or tumor lodge. The patients were included in the treatment by performing cone-beam computed tomography every other day.

Concomitant CT with RT was given weekly or every 3 weeks. Cisplatin, carboplatin or cetuximab was used as CT agents. Cisplatin CT was administered at a dose of 75-100 mg/m^2 every 3 weeks, or 40 mg/m^2 per week.

Patients were checked weekly during RT and subsequently every 3 months following the first 6 weeks after RT. Side effects observed within 90 days from the start of RT were considered as early side effects. Those observed 90 days after RT were considered late side effects. Side effects were scored according to the American RTOG criteria (https://en.wikibooks.org/wiki/Radiation_Oncology/Toxicity_grading/RTOG).

Statistical Analysis

Statistical analyses were performed using SPSS software, version 15.0 (IBM Corp., Armonk, NY, USA). The variables were investigated using visual (histograms, probability plots) and analytical methods (Kolmogorov-Simirnov/Shapiro-Wilk test) to determine whether they are normally distributed. In our study, RT and CT treatments, as well

as descriptive analyses, were presented using means and standard deviations for normally distributed variables (hematological toxicity variables). Paired Student's t-test was used to compare the measurements at two time points (pre-treatment and post-treatment) for hematological variables (lymphocyte, hemoglobin, thrombocyte, white blood cell, neutrophil). The statistical significance value of $p<0.05$ was accepted as significant. The effect of tumor stage (T, N), age, gender, smoking history, RT technique, and tumor location on the survival of squamous cell HN cancer was investigated using the log-rank test. The general characteristics of patients were noted. Frequency percentages were calculated for categorical variables.

RESULTS

Of a total of 78 patients, 15 (19.2%) were female and 63 (80.8%) male. The mean age of all patients is 60 years. The most common tumor location in the patients was larynx (53.8%) followed by oral cavity (24.4%) and oropharynx (15.4%). Twenty-seven (34.6%) patients were in the T2 stage, 18 (23.1%) patients were in T4, 17 (21.8%) patients were in T1, and 16 (20.5%) patients were in T3 stage. Most of the patients were N0 (39 patients, 50%) and 29 (%) patients were N2. Most of the patients were smokers (70 patients, 89.7%). 43 (55,1%) patients underwent surgery. Forty-three (55.1%) patients received adjuvant RT; 35 (44.9%) patients received definitive RT. In the evaluation of pathological findings of the operated patients, it was observed that 16 (20.5%) of patients had ECE. 27 (34.6%) patients had no ECE and 35 (44.9%) were unknown. Twenty-eight (35.9%) patients were PNI-negative and 12 (15.4%) patients were PNI-positive. Twenty-one (26.9%) patients were LVI negative and 14 (17.9%) patients were LVI-positive. Surgical margins were close or positive in 7 (8.9%) patients. Recurrence was seen in 11 (14.1%) of the patients. Metastasis was seen in 5 (6.4) of the patients. 10 (12.8%) patients developed the need for hospitalization during the treatment. Characteristics of all patients are shown in Table 1.

Radiation therapy at a dose of 54-70 Gy was applied to the tumor/tumor lodgewith or without lymphatics. Ten (12.8%) patients underwent three-dimensional conformal RT and 68 (87.2%) patients underwent intensity-modulated RT. Concomitant CT with RT was performed in 46 (59%) patients. Thirty-eight (48.8%) patients received cisplatin, 3 (3.8%) received carboplatin and 5 (6.4%) received cetuximab. Twenty-five (32.1%) patients underwent cisplatin CT at a dose of 75-100 mg/m^2 every 3 weeks, and 21 (26.9%) patients received cisplatin CT at a dose of 40 mg/m^2 per week.

Table 1. Patient properties

		n	%
Gender	**Female**	15	19.2
	Male	63	80.8
Comorbidity	**Yes**	33	42.3
	No	34	43.6
	Unknown	11	14.1
Tumor location	**Larynx**	42	53.8
	Oral cavity	19	24.4
	Hipopharynx	5	6.4
	Orofarenks	12	15.4
T stage	**T1**	17	21.8
	T2	27	34.6
	T3	16	20.5
	T4	18	23.1
N stage	**N0**	39	50
	N1	10	12.8
	N2	29	37.2
Smoke	**No**	8	10.3
	Smoker	70	89.7
Operation	**Yes**	43	55.1
	No	35	44.9
Surgical margins	**Positive**	7	8.9
	Negative	36	46.2
ECE	**Yes**	16	20.5
	No	27	34.6
	Unknown	35	44.9
BVI	**Yes**	9	11.5
	No	15	19.2
	Unknown	54	69.2
LVI	**Yes**	14	17.9
	No	21	26.9
	Unknown	43	55.1
PNI	**Yes**	12	15.4
	No	28	35.9
	Unknown	38	48.7
RT technique	**3DCRT**	10	12.8
	IMRT	68	87.2
RT status	**Adjuvant**	43	55.1
	Definitive	35	44.9
Concomitant CT	**Yes**	46	59
	No	32	41
CT agent	**Cisplatin**	38	48.8
	Carboplatin	3	3.8
	Cetuximab	5	6.4
CT plan	**Triweekly**	25	32.1
	Weekly	21	26.9

Table 1. Continued

		n	%
Recurrence	**Yes**	11	14.1
	No	67	85.9
Metastasis	**Yes**	5	6.4
	No	73	93.6
Hospitalization	**Yes**	10	12.8
	No	68	87.2
TPN	**Yes**	5	6.4
	No	73	93.6
ONS	**Yes**	59	75.6
	No	19	24.4
IV support	**Yes**	22	28.2
	No	56	71.8

CT: Chemotherapy, RT: Radiotherapy, ECE: Extracapsular extension, BVI: Blood vessel invasion, LVI: Lymphatic vessel invasion, PNI: Perineural invasion, 3DCRT: Three-dimensional conformal radiotherapy, IMRT: Intensity modulated radiotherapy, TPN: Total parenteral nutrition, ONS: Oral nutrition solution, IV: Intravenous

The overall survival rate for three years is 59% in all data. The side effects on skin, oral mucosa, and esophagus were noted. 48.7% of the patients had grade 3 skin reaction; 80.8% of the patients had grade 3 mucositis; 14.1% had grade 2 mucositis; and 71.8% had grade 3 esophagitis, and 21.8% had grade 2 esophagitis. A complete blood count was performed to assess the patients' hematological toxicity. 84.6% of patients had grade 1 hematological toxicity and 15.4% of patients had grade 2 toxicity. The distribution of side effects is given in Table 2.

When hematological toxicity was compared in all groups, significant differences in hemoglobin and platelets were found before and after RT. Borderline significant in white blood cells (Table 3). A significant difference was found in hemoglobin and platelets when hematologic toxicity was compared between groups in those who received CRT alone (Table 4).

When survival analysis was performed according to patient characteristics, N stage of the tumor, smoking habit, and tumor localization were found to be significant. T stage of the tumor was found to be borderline significant (Table 5).

DISCUSSION

Local treatment methods such as surgery and or RT may be preferred alone for definitive treatment in early-stage diseases (8). In addition, concomitant CRT may be the most appropriate approach in cases where surgical intervention is restricted due to the anatomical location of the tumor (9). For locally advanced resectable diseases, surgery and adjuvant RT with or without CT are accepted as a standard treatment.

Table 2. Side effect

	Grade	n (%)
Skin reaction	0	2 (2.6)
	2	38 (48.7)
	3	38 (48.7)
Mucositis	0	2 (2.6)
	2	11 (14.1)
	3	63 (80.8)
	4	2 (2.6)
Esophagitis	0	3 (3.8)
	1	2 (2.6)
	2	17 (21.8)
	3	56 (71.8)
Haematological toxicity	0	0 (0%)
	1	66 (84.6%)
	2	12 (15.4%)
	3	0 (0%)

Table 3. Hematological toxicity in all groups

	Mean	SD	Number of patient	p-value
Pre-RT hemoglobin	12.8995	1.29388	78	0.000
Post-RT hemoglobin	12.2022	1.19837	78	
Pre-RT neutrophil	5.6335	3.09142	78	0.098
Post-RT neutrophil	4.5286	2.23554	78	
Pre-RT white blood cell	8.4519	3.36864	78	0.052
Post-RT white blood cell	6.2701	2.47090	78	
Pre-RT lymphocyte	1.9150	0.67887	78	0.102
Post-RT lymphocyte	1.0968	1.73616	78	
Pre-RT platelet	302.2758	85.68495	78	0.000
Post-RT platelet	266.8949	69.36562	78	

T-test paired samples statistical. RT: Radiotherapy, SD: Standard deviation

Meta-analyses have shown increased locoregional disease control and survival rates with CRT (10,11). For patients with inoperable locally advanced HN cancer, high-dose cisplatin-based concomitant CRT remains standard of care (12). In our study, 20.5% of patients were T3, 23.1% were T4, 12.8% were N1, and 37.2% were N2.

It is known that CRT treatment modality is associated with increased toxicity and is less tolerated in patients with poorer performance (12). Also, it increases the risk of mucositis, hematological suppression, and dermatitis (13). Cisplatin-based concomitant CRT improved patient survival but also increased toxicities such as gastrointestinal, haematological,

Table 4. Hematological toxicity only in CRT patients

	Mean	SD	Number of patient	p-value
Pre-RT hemoglobin	12.7524	1.11535	46	0.000
Post-RT hemoglobin	11.9822	1.17008	46	
Pre-RT neutrophil	5.8370	3.63124	46	0.108
Post-RT neutrophil	4.2220	2.24300	46	
Pre-RT white blood cell	8.6226	3.79959	46	0.066
Post-RT white blood cell	5.8461	2.47285	46	
Pre-RT lymphocyte	1.8354	.56599	46	0.498
Post-RT lymphocyte	0.8022	.32004	46	
Pre-RT platelet	299.7546	83.82116	46	0.005
Post-RT platelet	251.1913	72.39278	46	

T-test paired samples statistical. CRT: Chemoradiotherapy, RT: Radiotherapy, Std: Standard deviation

Table 5. Univariate analysis of patient characteristics

	3-year survival (%)	p-value
Age		
<50 (11) ≥50 (67)	52% 60%	0.575
Gender		
Male (63) Female (15)	58% 57%	0.89
T stage		
1 (17) 2 (27) 3 (16) 4 (18)	75% 70% 74% 19%	0.065
N stage		
N0 (39) N1 (10) N2 (29)	83% 50% 23%	0.009
Smoking history		
Yes (70) No (8)	46% 100%	0.023
RT technique		
3DRT (10) ARC/IMRT (68)	37% 64%	0.148
Tumor location		
Larynx (42) Oral cavity (19) Oropharynx (12) Hypopharynx (5)	82% 43% 0 0	0.000

3DRT: Three-dimensional radiotherapy, RT: Radiotherapy, ARC/IMRT: Arc therapy/intensity-modulated radiotherapy

and renal (14-16). Adelstein et al. (14) reported that the rate of side effects of grade 3 and above was 89% in their study. In our study, 80.8% of patients had grade 3 mucositis, 14.1% had grade 2 mucositis, 71.8% had grade 3 esophagitis and 21.8% had grade 2 esophagitis. Also, 48.7% of patients had grade 3 skin reactions, and 48.7% had grade 2 skin reactions. 15.4% had grade 2 hematological toxicity, and 84.6% had grade 1 hematological toxicity.

The study by Wu et al. (12) showed that CT causes hematological damage. In our study, hemoglobin and platelets significantly decreased in patients receiving CRT (p=0.000 for hemoglobin and p=0.005 for platelets).

In the study of Alterio et al. (1), the tumor located in the larynx and nasopharynx was shown to be a positive prognostic factor compared to other localizations. They also identified tumors located in the hypopharynx and other median regions as a poor prognostic factor. In our study, tumor location was found to be significant for survival. Three-year survival rates: larynx 83%; oral cavity 19%; hypopharynx and oropharynx 0%. The p-value was found to be significant (p=0.000). The study by Riaz et al. (17) reported that having T3-T4 tumors was a poorer prognostic factor than T1-T2 tumors. In our study, 3 years' survival for T4 tumors was 19%. The p-value was found to be borderline significant (p=0.065). Also, in our study, N stage and smoking history were found to be significant for survival. The p-values are 0.009 for N stage and 0.023 for smokers.

Study Limitations

This analysis has some limitations. First, its retrospective and single-center nature limits the generalizability of the information obtained. The relatively small sample sizes provide potentially significant prognostic value in proportion to the overall power distribution. Furthermore, some pathological and clinical data, such as human papillomavirus status and p16 expression, may be relevant for predicting survival and treatment response in each patient.

CONCLUSION

In conclusion, the treatment of HN cancers aims to achieve disease-free survival and a functional life, in which organs at risk are protected as much as possible. RT/CRT is a highly toxic, long-term, organ-preserving therapy. One of the main goals is to provide survival advantage by increasing local control with and to protect patients from side effects.

ETHICS

Ethics Committee Approval: This study was approved by the Clinical Research Ethics Committee of the University of Health Sciences Türkiye, Bakırköy Dr. Sadi Konuk Training and Research Hospital (approval no: 2022-09, date: 09.05.2022).

Informed Consent: This is a retrospective study, and all participants provided written informed consent.

FOOTNOTES

Authorship Contributions

Surgical and Medical Practices: G.P.S., Concept: G.P.S., E.E.Ö., M.K.B., M.F., E.K.Ü., Design: G.P.S., Data Collection or Processing: G.P.S., M.K.B., M.F., Analysis or Interpretation: G.P.S., M.K.B., E.K.Ü., Literature Search: G.P.S., E.E.Ö., M.K.B., M.F., Writing: G.P.S., E.E.Ö., M.F., E.K.Ü.

Conflict of Interest: No conflict of interest was declared by the authors.

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

REFERENCES

1. Alterio D, Marvaso G, Ferrari A, Volpe S, Orecchia R, Jereczek-Fossa BA. Modern radiotherapy for head and neck cancer. Semin Oncol. 2019;46:233-45.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. CA Cancer J Clin. 2017;67:7-30.
3. Schwartz DL, Hayes DN. The evolving role of radiotherapy for head and neck cancer. Hematol Oncol Clin North Am. 2020;34:91-108.
4. Kreimer AR, Clifford GM, Boyle P, Franceschi S. Human papillomavirus types in head and neck squamous cell carcinomas worldwide: a systematic review. Cancer Epidemiol Biomarkers Prev. 2005;14:467-75.
5. Demiröz C, Özşahin EM. Skuamöz hücreli baş boyun kanserinde kemoradyoterapi. Uludağ Univ Tıp Fak Derg. 2011;37:61-5.
6. Atasoy BM, Dane F, Sarı M, Akgün Z, Yumuk PF, Turhal NS, et al. Lokal ileri evre skuamöz hücreli baş ve boyun kanserinde sisplatinle eş zamanlı kemoradyoterapi: yan etki ve uygulanabilirlik analizi. Turk J Oncol. 2008;23:1.
7. Bernier J, Cooper JS, Pajak TF, van Glabbeke M, Bourhis J, Forastiere A, et al. Defining risk levels in locally advanced head and neck cancers: a comparative analysis of concurrent postoperative radiation plus chemotherapy trials of the EORTC (#22931) and RTOG (# 9501). Head Neck. 2005;27:843-50.
8. Agrawal N, Ha PK. Management of early-stage laryngeal cancer. Otolaryngol Clin North Am. 2008;41:757-69.
9. Scher RL, Esclamado RM. Organ and function preservation: the role of surgery as the optimal primary modality or as salvage after chemoradiation failure. Semin Radiat Oncol. 2009;19:17-23.
10. Pignon JP, le Maître A, Bourhis J; MACH-NC Collaborative Group. Meta-analyses of chemotherapy in head and neck cancer (MACH-NC): an update. Int J Radiat Oncol Biol Phys. 2007;69(2 Suppl):S112-4.
11. Brizel DM, Albers ME, Fisher SR, Scher RL, Richtsmeier WJ, Hars V, et al. Hyperfractionated irradiation with or without concurrent chemotherapy for locally advanced head and neck cancer. N Engl J Med. 1998;338:1798-804.
12. Wu Q, Zhu C, Zhang S, Zhou Y, Zhong Y. Hematological toxicities of concurrent chemoradiotherapies in head and neck cancers: comparison among cisplatin, nedaplatin, lobaplatin, and nimotuzumab. Front Oncol. 2021;11:762366.
13. Bonner JA, Harari PM, Giralt J, Azarnia N, Shin DM, Cohen RB, et al. Radiotherapy plus cetuximab for squamous-cell carcinoma of the head and neck. N Engl J Med. 2006;354:567-78.
14. Adelstein DJ, Li Y, Adams GL, Wagner HJr., Kish JA, Ensley JF, et al. An intergroup phase III comparison of standard radiation therapy and two schedules of concurrent chemoradiotherapy in patients with unresectable squamous cell head and neck cancer. J Clin Oncol. 2003;21:92-8.
15. Blanchard P, Lee A, Marguet S, Leclercq J, Ng WT, Ma J, et al. Chemotherapy and radiotherapy in nasopharyngeal carcinoma: an update of the MAC-NPC metaanalysis. Lancet Oncol. 2015;16:645-55.
16. Chen QY, Wen YF, Guo L, Liu H, Huang PY, Mo HY, et al. Concurrent chemoradiotherapy vs radiotherapy alone in stage II nasopharyngeal carcinoma: phase III randomized trial. J Natl Cancer Inst. 2011;103:1761-70.
17. Riaz N, Hong JC, Sherman EJ, Morris L, Fury M, Ganly I, et al. A nomogram to predict loco-regional control after re-irradiation for head and neck cancer. Radiother Oncol. 2014;111:382-7.

Medical Journal of Bakırköy

Med J Bakirkoy 2025;21(3):252-258
DOI: 10.4274/BMJ.galenos.2025.2025.3-3

Research

Evaluating Surgical Outcomes of Supracondylar Humerus Fractures Through Patient and Parent Perspectives

Suprakondiler Humerus Kırıklarının Cerrahi Sonuçlarının Hasta ve Ebeveyn Perspektifinden Değerlendirilmesi

Vedat Öztürk[1], Başar Burak Çakmur[2], Nebi Köksal Yazgan[3], Malik Çelik[1], Alkan Bayrak[4], Altuğ Duramaz[1]

[1]University of Health Sciences Türkiye, Bakırköy Dr. Sadi Konuk Training and Research Hospital, Clinic of Orthopedics and Traumatology, İstanbul, Türkiye

[2]Bezylikdüzü State Hospital, Clinic of Orthopedics and Traumatology, İstanbul, Türkiye

[3]Hitit University Çorum Erol Olçok Training and Research Hospital, Department of Orthopedics and Traumatology, Çorum, Türkiye

[4]İstinye University Hospital, Medical Park Gaziosmanpaşa, Department of Orthopedics and Traumatology, İstanbul, Türkiye

ABSTRACT

Objective: This study aimed to evaluate the surgical outcomes of supracondylar humerus (SCH) fractures in children aged 8-18 years, focusing on patient and parental satisfaction.

Methods: A retrospective analysis was conducted on 105 pediatric patients (81 male and 24 female) treated surgically between 2018 and 2025. Data collected included demographic characteristics, fracture classification (Gartland), reduction type, surgical method, Baumann angle, capitellohumeral angle, healing time, follow-up duration, Visual Analog Scale (VAS) scores, Pediatric Quality of Life Inventory (PedsQL) scores, Client Satisfaction Questionnaire-8 (CSQ-8) scores, Flynn cosmetic and functional criteria, and complications. Subgroup analyses were performed to assess differences based on fracture classification and reduction type.

Results: The mean age of the patients was 10.16±1.9 years. Gartland type 2 fractures demonstrated significantly shorter healing times compared to type 3 fractures (p=0.05), and closed reductions were associated with shorter healing times than open reductions (p=0.04). Although PedsQL scores, were higher in Gartland type 2 fractures (95.3±8) and closed reductions (94.9±8.1) compared to type 3 fractures (93.8±8.4) and open reductions (92.4±9.1), these differences were not statistically significant. Subgroup analysis revealed significant associations between CSQ-8 scores and the presence of concomitant injuries (p=0.03) as well as the need for revision surgeries (p=0.028). Overall, surgical outcomes were satisfactory, with high PedsQL and low VAS scores across all groups.

Conclusion: Patient and parental satisfaction in SCH fracture treatment is influenced by concomitant injuries and revision surgery, despite generally favorable surgical outcomes.

Keywords: Supracondylar humerus fractures, pediatric trauma, patient and parental satisfaction

ÖZ

Amaç: Bu çalışmada, 8-18 yaş arası çocuklarda suprakondiler humerus (SKH) kırıklarının cerrahi sonuçlarının hasta ve ebeveyn memnuniyeti ile olan ilişkisi incelenmiştir.

Gereç ve Yöntem: 2018-2025 yılları arasında cerrahi olarak tedavi edilen 105 pediatrik hasta (81 erkek, 24 kız) üzerinde retrospektif bir analiz yapıldı. Toplanan veriler arasında demografik özellikler, kırık sınıflaması (Gartland), redüksiyon türü, cerrahi yöntem, Baumann açısı, kapitellohumeral açı, kaynama süresi, takip süresi, Görsel Analog Skala (GAS) skorları, Çocuk Yaşam Kalitesi Envanteri (PedsQL) skorları, Hasta Memnuniyeti Anketi-8 (CSQ-8) skorları, Flynn kozmetik ve fonksiyonel kriterleri ile komplikasyonlar yer aldı. Alt grup analizleri, kırık sınıflamasına ve redüksiyon türüne göre farklılıkları değerlendirmek için gerçekleştirildi.

Address for Correspondence: Vedat Öztürk MD, University of Health Sciences Türkiye, Bakırköy Dr. Sadi Konuk Training and Research Hospital, Clinic of Orthopedics and Traumatology, İstanbul, Türkiye
E-mail: dr.ozturkvedat@gmail.com **ORCID ID:** orcid.org/0000-0003-2412-9725

Cite as: Öztürk V, Çakmur BB, Yazgan NK, Çelik M, Bayrak A, Duramaz A. Evaluating surgical outcomes of supracondylar humerus fractures through patient and parent perspectives: a study in children aged 8-18 years. Med J Bakirkoy. 2025;21(3):252-258

Received: 11.03.2025
Accepted: 13.05.2025
Publication Date: 03.09.2025

Bulgular: Hastaların ortalama yaşı 10,16±1,9 yıl idi. Gartland tip 2 kırıklarında, tip 3 kırıklara kıyasla anlamlı derecede daha kısa kaynama süreleri gözlendi (p=0,05) ve kapalı redüksiyon uygulanan hastalarda açık redüksiyon yapılanlara kıyasla kaynama süreleri daha kısa bulundu (p=0,04). Gartland tip 2 kırıklarında (95,3±8) ve kapalı redüksiyon uygulanan hastalarda (94,9±8,1) PedsQL skorları, tip 3 kırıklar (93,8±8,4) ve açık redüksiyon (92,4±9,1) ile karşılaştırıldığında daha yüksek olmasına rağmen, bu farklılıklar istatistiksel olarak anlamlı değildi. Alt grup analizi, CSQ-8 skorlarının eşlik eden yaralanmaların varlığı (p=0,03) ve revizyon cerrahisi ihtiyacı (p=0,028) ile anlamlı ilişkili olduğunu ortaya koydu. Genel olarak, tüm gruplarda yüksek PedsQL ve düşük GAS skorları ile cerrahi sonuçlar tatmin edici bulundu.

Sonuç: SKH kırıklarının cerrahi tedavisinde hasta ve ebeveyn memnuniyeti, genellikle olumlu sonuçlara rağmen, eşlik eden yaralanmalar ve revizyon cerrahisi ihtiyacından etkilenmektedir.

Anahtar Kelimeler: Suprakondiler humerus kırıkları, pediatrik travma, hasta ve ebeveyn memnuniyeti

INTRODUCTION

Supracondylar humerus (SCH) fractures are among the most common elbow injuries in the pediatric population (1). These fractures typically occur due to mechanisms such as falls, trauma, or high-energy impacts, and require careful management due to their potential to cause serious neurovascular complications (2). The primary goals in the treatment of SCH fractures are to restore fracture stability, protect neurovascular structures, and optimize functional recovery (3).

The choice of treatment is generally determined by the type and severity of the fracture. According to the Gartland classification, type 1 fractures can often be managed conservatively, whereas type 2 and type 3 fractures typically require surgical intervention. Closed reduction and percutaneous pinning are considered the gold standard for the surgical management of these fractures, while open reduction and internal fixation are reserved for more complex fractures or cases where closed reduction fails (4).

The literature has explored the relationship between different fracture types, treatment outcomes, and patient satisfaction, focusing on patient perspectives and cosmetic results (5-9). However, studies evaluating the relationship between both the surgical treatment of SCH fractures and patient and parental satisfaction, as well as cosmetic outcomes, remain limited.

This study aims to evaluate the radiological, clinical, and cosmetic outcomes of the surgical treatment of SCH fractures. Additionally, it seeks to assess treatment outcomes from the perspectives of patients and their parents, and to investigate the associations between these outcomes and variables such as fracture type and treatment method.

METHODS

Study Design and Setting

This retrospective, single-center study evaluated pediatric patients who underwent surgical treatment for SCH fractures between 2018 and 2025. Ethical approval for the study was obtained from the Clinical Research Ethics Committee of the University of Health Sciences Türkiye, Bakırköy Dr. Sadi Konuk Training and Research Hospital (approval no: 2021-04-12, date: 15.02.2021).

Inclusion and Exclusion Criteria

The study included pediatric patients aged 8 to 18 years who presented with SCH fractures, underwent surgical treatment, and completed a minimum follow-up of 24 months at our institution. Patients capable of completing the Pediatric Quality of Life Inventory (PedsQL) questionnaire were eligible for inclusion. Exclusion criteria were incomplete medical records, follow-up at external institutions, pre-existing upper extremity deformities, metabolic bone disorders, or pathological fractures. A total of 105 patients who met the inclusion criteria were included in the analysis.

Data Collection

Demographic data (age, sex, injured side, dominant side, and mechanism of injury) were collected directly from patients during outpatient visits and confirmed using electronic medical records. Associated injuries (vascular injury, neurological injury, compartment syndrome), mechanisms of injury, and fracture classification based on the Gartland system were documented. Surgical treatment details (open or closed reduction, fixation method) and complications during or after surgery were recorded. Radiological outcomes, including Baumann angle and capitellohumeral angle, were assessed using the most recent radiographs obtained during follow-up. Clinical outcomes such as healing time, follow-up duration, and Visual Analog Scale (VAS) scores were also evaluated.

Surgical Technique

Under general anesthesia, patients were positioned supine, and pneumatic tourniquets were applied to control blood flow and ensure optimal surgical conditions. Closed reduction and percutaneous pinning were initially attempted in all cases. In successful closed reductions, fixation was achieved using crossed K-wires inserted laterally and/or medially, depending on fracture configuration. For cases where closed reduction was unsuccessful, the pneumatic tourniquet was inflated and a lateral mini-incision was made

for open reduction. Fixation was subsequently achieved using crossed K-wires, similar to the closed reduction group. At the end of the surgical procedure, a long-arm splint was applied to all patients (Figure 1).

Postoperative Care

Patients were monitored with a standard analgesia protocol during hospitalization. Depending on pain levels and soft tissue conditions, they were discharged within 24 to 48 hours postoperatively. All patients were called for a follow-up visit at the second postoperative week. For those who underwent open reduction, sutures were removed during this visit. At the fourth postoperative week, K-wires were removed, and passive and active range-of-motion exercises, as well as strengthening exercises, were initiated. Follow-up visits were scheduled at the second, third, sixth, and twelfth months postoperatively and annually thereafter. During the final follow-up visit, patient satisfaction was evaluated using PedsQL scores, and parental satisfaction was assessed using Client Satisfaction Questionnaire-8 (CSQ-8) scores. Additionally, objective cosmetic and functional outcomes were evaluated using Flynn cosmetic and functional criteria, and radiological outcomes were assessed by measuring the Baumann and capitellohumeral angles (Figure 2).

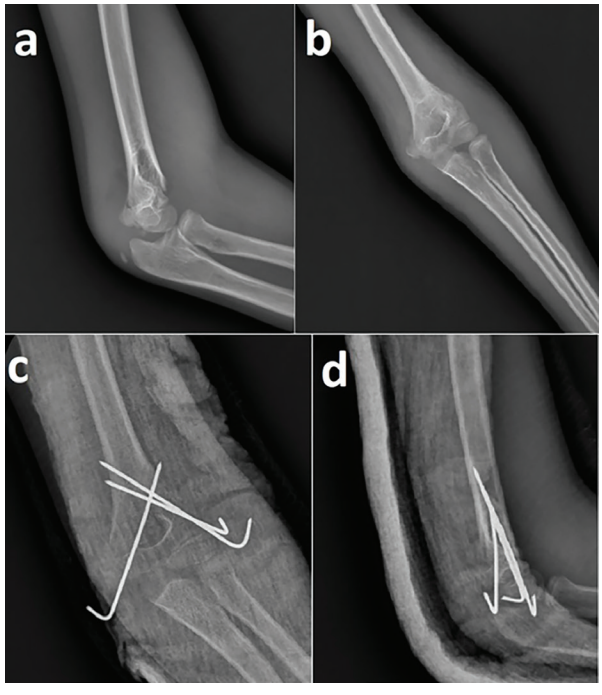

Figure 1. Preoperative and postoperative radiographic images of a 9-year-old male patient with a Gartland type 3 supracondylar humerus fracture. (a) Preoperative lateral elbow X-ray. (b) Preoperative anteroposterior (AP) elbow X-ray. (c) Postoperative AP elbow X-ray showing closed reduction and fixation with two lateral and one medial K-wire. (d) Postoperative lateral elbow X-ray

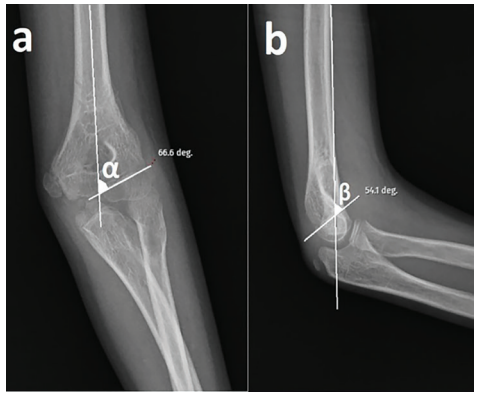

Figure 2. Anteroposterior (AP) and lateral elbow X-rays of the same patient taken at the 2-year follow-up. (a) Measurement of the Baumann angle on the AP elbow X-ray. (b) Measurement of the capitellohumeral angle on the lateral elbow X-ray

Primary and Secondary Outcomes

The primary outcomes of this study were healing time, radiological results (Baumann angle and capitellohumeral angle), patient satisfaction measured using PedsQL scores, and parental satisfaction evaluated using CSQ-8 scores. Secondary outcomes included functional and cosmetic results assessed using Flynn criteria, the occurrence of complications, the need for revision surgeries, and postoperative pain levels based on VAS scores. Additionally, correlations between CSQ-8, PedsQL, and Flynn criteria were analyzed to explore their interrelationships.

Statistical Analysis

Statistical analysis was performed using IBM SPSS 26 (Chicago, IL, USA). Descriptive statistical methods (minimum, maximum, and median values) were used to evaluate the study data. The normality of quantitative data was tested using the Kolmogorov-Smirnov test and graphical examinations. Independent sample t-tests were used to compare normally distributed quantitative variables between two groups, while the chi-square test was employed for qualitative independent data. A p-value of <0.05 was considered statistically significant.

RESULT

The demographic characteristics of the patients are summarized in Table 1. A total of 105 patients, aged 8-15 years, were included in the study (24 females and 81 males). Subgroup analyses based on fracture types and reduction methods revealed no statistically significant differences in Baumann angle, capitellohumeral angle, follow-up duration, or VAS scores among the groups ($p>0.05$). However, healing time was significantly shorter in Gartland type 2 fractures compared to type 3 fractures ($p=0.05$); and in cases treated

with closed reduction compared to open reduction (p=0.04).

The mean PedsQL scores were higher in Gartland type 2 fractures (95.3±8) compared to type 3 fractures (93.8±8.4), but this difference was not statistically significant (p>0.05). Similarly, PedsQL scores were higher in cases treated with closed reduction (94.9±8.1) compared to open reduction (92.4±9.1), but the difference was not statistically significant (p>0.05) (Table 2).

According to the Flynn criteria, there were no statistically significant differences in fracture classification, reduction type, or sex (p>0.05). When CSQ-8 scores were categorized into three subgroups, no statistically significant differences were found among dominant side, mechanism of injury, Gartland classification, surgical method, or sex (p>0.05) (Table 3). However, subgroup analysis based on CSQ-8 scores revealed a statistically significant difference between the intermediate and high CSQ-8 groups in terms of the presence of associated injuries (p=0.03). Similarly, revision surgeries were significantly more common in the intermediate CSQ-8 group compared to the high CSQ-8 group (p=0.028).

Key Findings

1) Healing time was shorter in cases of Gartland type 2 fractures and in cases treated with closed reduction.

2) Mean PedsQL scores were higher in Gartland type 2 fractures and cases treated with closed reduction, although the differences were not statistically significant.

3) Patients with associated injuries were significantly more common in the intermediate CSQ-8 group compared to the high CSQ-8 group.

4) Revision surgeries were significantly more frequent in the intermediate CSQ-8 group compared to the high CSQ-8 group.

Table 1. Demographic distribution of the patients

Parameter	**Mean±SD**	**Median**	**Min-Max**
Age	10.16±1.957	9	8-15
Baumann angle	72.87±5.654	73	59-85
Capitellohumeral angle	44.87±5.987	44	35-63
Healing time (weeks)	3.75±0.568	4	3-5
Follow-up duration (months)	34.74±8.780	32	24-60
VAS score	0.28±0.628	0	0-3
PedsQL score	94.54±8.289	98	71-100
CSQ-8 score	26.48±4.641	28	12-32
CSQ-8 mean score	2.59±0.567	3	1-3
Variable			**n (%)**
Gartland classification	Type 2		50 (47.6%)
	Type 3		55 (52.4%)
Surgical approach	Open		17 (16.1%)
	Closed		88 (83.9%)
Associated injuries	Vascular injury		0
	Nerve injury		7 (6.6%)
Gender	Male		81 (77.1%)
	Female		24 (22.9%)
Affected side	Right		57 (54.3%)
	Left		48 (45.7%)
Dominant side	Right		101 (96.2%)
	Left		4 (3.8%)
Revision surgery	Yes		6 (5.7%)
	No		99 (94.3%)

SD: Standard deviation, VAS: Visual Analog Scale, PedsQL: Pediatric Quality of Life Inventory, CSQ-8: Client Satisfaction Questionnaire-8

Table 2. Clinical and functional outcomes based on Gartland classification and reduction type

	Gartland classification			Reduction type		
	Type 2	Type 3		Open	Close	
	Mean±SD	Mean±SD	p*	Mean±SD	Mean±SD	p*
Age	10.06±1.8	10.2±2.1	0.6	9.8±1.7	10.2±1.9	0.43
Baumann angle	73.6±4.9	72.1±6.1	0.19	72.1±6.6	73.1±5.5	0.55
Capitellohumeral angle	44.8±5.08	44.9±6.7	0.9	45.9±6.7	44.6±5.8	0.42
Healing time (weeks)	3.6±0.5	3.8±0.6	0.05	4±0.7	3.7±0.5	0.04
Follow-up duration (months)	34.5±9	34.9±8.6	0.8	34±9.5	34.8±8.6	0.7
VAS score	0.2±0.5	0.3±0.7	0.4	0.18±0.5	0.3±0.6	0.46
PedsQL score	95.3±8.1	93.8±8.4	0.3	92.4±9.1	94.9±8.1	0.26
CSQ-8 score	26.7±4.7	26.2±4.6	0.6	25.2±5.7	26.7±4.3	0.23

*Independent sample t-test, SD: Standard deviation, VAS: Visual Analog Scale, PedsQL: Pediatric Quality of Life Inventory, CSQ-8: Client Satisfaction Questionnaire-8

Table 3. Comparison of Flynn cosmetic and functional outcomes

		Flynn cosmetic				p*	Flynn functional			p*
		1	2	3	4		1	2	3	
Gartland classification	2	22	25	1	2	0.11	28	21	1	0.36
	3	19	27	8	1		19	30	6	
Surgical approach	Open	9	5	3	0	0.15	7	8	2	0.65
	Close	32	47	6	3		40	43	5	
Gender	Female	9	11	3	1	0.83	10	12	2	0.9
	Male	32	41	6	2		37	39	5	

*Chi-square test

DISCUSSION

The most significant findings of this study are that Gartland type 2 fractures and cases treated with closed reduction exhibited shorter recovery times and higher PedsQL scores. Furthermore, patient and parent satisfaction, as measured by CSQ-8 scores, was significantly influenced by the presence of fracture-related injuries and the need for revision surgery.

In our study, Gartland type 2 fractures exhibited shorter recovery times compared to type 3 fractures. This finding is consistent with previous studies reporting similar recovery durations for less severe fractures (2,10,11). Additionally, our study indicates that closed reduction results in shorter recovery times compared to open reduction, thereby corroborating previous studies suggesting that closed techniques reduce complication rates and accelerate recovery (4,12,13). Prior studies have also recommended closed reduction whenever possible, as it minimizes surgical trauma and postoperative complications (13,14).

Although PedsQL scores in our study were higher in type 2 fractures and in cases treated with closed reduction, the differences were not statistically significant. This finding aligns with another study in the literature that did not observe a significant difference in long-term satisfaction between different treatment methods for SCH fractures (10). Another possible explanation for the lack of statistical significance is the overall high level of recovery observed across all groups. This suggests that despite the observed differences, appropriate surgical treatment of SCH fractures remains highly effective across different fracture types.

Parental expectations play a crucial role in managing a child's recovery, as they significantly influence their perception of the outcome and, consequently, their level of satisfaction (7,10,15,16). In our study, we observed that parental satisfaction, as measured by CSQ-8 scores, was negatively affected in patients with associated injuries and those requiring revision surgery. Notably, in the moderate CSQ-8 satisfaction group, the need for revision surgery was more prevalent. Previous studies have highlighted the impact of surgical complications and secondary interventions on satisfaction, reporting a direct correlation between surgical complications and parental dissatisfaction (6,15,16). Patients who sustained additional injuries during the initial

trauma were more likely to report lower satisfaction, likely due to prolonged treatment durations and increased postoperative challenges. These findings underscore the importance of comprehensive preoperative counseling and expectation management for patients and their families, particularly in cases involving complex fractures or potential complications.

In terms of functional and cosmetic outcomes, the Flynn criteria used in our study did not reveal significant differences between groups. This finding is consistent with the results of Rapp et al. (6), who reported that fracture type and surgical approach do not substantially affect the cosmetic outcomes of pediatric clavicle fractures. Similarly, the lack of significant differences in Baumann and capitellohumeral angles, between treatment methods, aligns with previous research indicating that radiological outcomes do not always correlate with perceived functional recovery or satisfaction (17). Another study in the literature suggests that, despite varying treatment approaches for different fracture types, satisfactory cosmetic and functional outcomes can be achieved with the appropriate surgical technique (18). This underscores the importance of radiological outcomes while also highlighting their limitations in fully capturing patient and parental experiences, emphasizing the need for more comprehensive satisfaction assessments.

Our study also identified a significant association between revision surgeries and lower satisfaction scores. This finding is consistent with the study by Keppler et al. (5), which reported that patients requiring multiple surgical procedures generally express lower levels of satisfaction. This highlights the importance of achieving optimal surgical outcomes to minimize the need for additional procedures. Furthermore, the challenges associated with managing fractures with concomitant injuries can significantly impact patient satisfaction (14).

One of the strengths of this study is its integration of both objective clinical measurements and subjective patient-reported outcomes, providing a comprehensive assessment of treatment success. By evaluating both radiological parameters and patient satisfaction measures, we were able to identify key factors influencing postoperative recovery from multiple perspectives. Additionally, the inclusion of parental satisfaction as an outcome measure adds a new dimension to the evaluation of pediatric fracture management (7).

Study Limitations

This study's retrospective design and single-center nature may limit its generalizability. While follow-up periods were adequate for short- and mid-term assessments, longer-term evaluations could provide further insights into treatment durability. Additionally, satisfaction surveys are subjective and may be influenced by psychological and familial factors.

CONCLUSION

Surgical treatment of SCH fractures in children yields favorable outcomes, with shorter healing times in Gartland type 2 fractures and when using closed reductions. While functional and cosmetic results remain similar across treatment groups, patient and parental satisfaction is influenced by concomitant injuries and the need for revision surgery.

ETHICS

Ethics Committee Approval: Ethical approval for the study was obtained from the Clinical Research Ethics Committee of the University of Health Sciences Türkiye, Bakırköy Dr. Sadi Konuk Training and Research Hospital (approval no: 2021-04-12, date: 15.02.2021).

Informed Consent: Informed consent was obtained from all individual participants and their legal guardians.

FOOTNOTES

Authorship Contributions

Surgical and Medical Practices: V.Ö., B.B.Ç., N.K.Y., M.Ç., A.B., A.D., Concept: V.Ö., A.B., A.D., Design: V.Ö., A.B., A.D., Data Collection or Processing: B.B.Ç., N.K.Y., M.Ç., Analysis or Interpretation: N.K.Y., A.B., A.D., Literature Search: N.K.Y., M.Ç., Writing: V.Ö.

Conflict of Interest: No conflict of interest was declared by the authors.

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

REFERENCES

1. Çelik M, Kantarcı M, Sapmaz İE, Öztürk V, Baca E, Duramaz A. Characterization and mapping of upper extremity fractures in children by a tertiary hospital quarantine and post-quarantine period. Jt Dis Relat Surg. 2024;35:692-8.
2. Tuomilehto N, Sommarhem A, Nietosvaara AY. 9 years' follow-up of 168 pin-fixed supracondylar humerus fractures in children. Acta Orthop. 2018;89:351-6.
3. Baumann AN, Anaspure O, Patel S, Shams K, Yoder RG, Mistovich RJ. Examining outcomes and complications for operative versus nonoperative treatment of pediatric type II supracondylar humerus fractures: a systematic review of comparative studies. J Pediatr Orthop. 2025;45:e1-e9.
4. Vajda M, Lőrincz A, Szakó L, Szabó L, Kassai T, Zahár Á, et al. The first analysis of a multicentre paediatric supracondylar humerus fracture (SCHF) registry by fracture type. Arch Orthop Trauma Surg. 2024;145:39.

5. Keppler L, Navarre F, Keppler AM, Ihle C, Becker J, Fürmetz J, et al. Fulfillment of expectations and patient satisfaction following surgical treatment of complex proximal tibial fractures. Arch Orthop Trauma Surg. 2024;145:87.
6. Rapp M, Prinz K, Kaiser MM. Elastic stable intramedullary nailing for displaced pediatric clavicle midshaft fractures: a prospective study of the results and patient satisfaction in 24 children and adolescents aged 10 to 15 years. J Pediatr Orthop. 2013;33:608-13.
7. Gill M, Drendel AL, Weisman SJ. Parent satisfaction with acute pediatric pain treatment at home. Clin J Pain. 2013;29:64-9.
8. Çelik M, Çanlıoğlu A, Tıngır M, Öztürk V, Kural C. Frequencies and lines of pediatric distal humerus fractures: where and how are the fractures? Med J Bakirkoy. 2025;21:14-8.
9. Çelik M, Bayrak A, Öztürk V, Kizilkaya C, Avkan MC. Functional and radiological comparison and evaluation of gustilo-anderson type 3 open tibia fractures treated with an ilizarov external fixator. Med J Bakirkoy. 2024;20:308-14.
10. Young S, Fevang JM, Gullaksen G, Nilsen PT, Engesæter LB. Parent and patient satisfaction after treatment for supracondylar humerus fractures in 139 children: no difference between skeletal traction and crossed pin fixation at long-term followup. Adv Orthop. 2012;2012:1-5.
11. Samaila EM, Auregli L, Pezzè L, Colò G, Magnan B. Medium-term clinical results in the treatment of supracondylar humeral fractures in children: does the surgical approach impact outcomes? J Orthop Traumatol. 2024;25:43.
12. Sinikumpu JJ, Victorzon S, Pokka T, Lindholm EL, Peljo T, Serlo W. The long-term outcome of childhood supracondylar humeral fractures: a population-based follow up study with a minimum follow up of ten years and normal matched comparisons. Bone Joint J. 2016;98-B:1410-7.
13. Barik S, Garg V, Sinha SK, Chaudhary S, Kandwal P, Singh V. A meta-analysis on comparison of open vs closed reduction of gartland type 3 supracondylar humerus fractures in children. Acta Chir Orthop Traumatol Cech. 2023;90:198-205.
14. Williams KG, Smith G, Luhmann SJ, Mao J, Gunn JD, Luhmann JD. A randomized controlled trial of cast versus splint for distal radial buckle fracture: an evaluation of satisfaction, convenience, and preference. Pediatr Emerg Care. 2013;29:555-9.
15. Basaran SH, Ercin E, Bayrak A, Bilgili MG, Kizilkaya C, Dasar U, et al. The outcome and parents-based cosmetic satisfaction following fixation of paediatric supracondylar humerus fractures treated by closed method with or without small medial incision. Springerplus. 2016;5:1-5.
16. Young S, Fevang JM, Gullaksen G, Nilsen PT, Engesæter LB. Parent and patient satisfaction after treatment for supracondylar humerus fractures in 139 children: no difference between skeletal traction and crossed pin fixation at long-term followup. Adv Orthop. 2012;2012:1-5.
17. Silva M, Sadlik G, Avoian T, Ebramzadeh E. A removable long-arm soft cast to treat nondisplaced pediatric elbow fractures: a randomized, controlled trial. J Pediatr Orthop. 2018;38:223-9.
18. Poonai N, Myslik F, Joubert G, Fan J, Misir A, Istasy V, et al. Point-of-care Ultrasound for nonangulated distal forearm fractures in children: test performance characteristics and patient-centered outcomes. Acad Emerg Med. 2017;24:607-16.

Medical Journal of Bakırköy

Med J Bakirkoy 2025;21(3):259-266
DOI: 10.4274/BMJ.galenos.2025.2024.12-2

Research

Computed Tomography Texture Analysis for Monitoring Treatment Response in Childhood Lymphomas: A Preliminary Study

Çocukluk Çağı Lenfomalarında Tedavi Yanıtının İzlenmesi için Bilgisayarlı Tomografi Doku Analizi: Bir Ön Çalışma

İbrahim Altındaş[1], Abdi Gürhan[1], Buket Kara[2], Zuhal Bayramoğlu[3], Mehmet Öztürk[1], Yavuz Köksal[2]

[1]Selçuk University Faculty of Medicine, Department of Radiology, Division of Pediatric Radiology, Konya, Türkiye

[2]Selçuk University Faculty of Medicine, Department of Pediatric Hematology and Oncology, Konya, Türkiye

[3]İstanbul University, İstanbul Faculty of Medicine, Department of Radiology, Division of Pediatric Radiology, İstanbul, Türkiye

ABSTRACT

Objective: This preliminary study aims to evaluate the effectiveness of computed tomography (CT) texture analysis parameters in monitoring treatment response in childhood lymphomas.

Methods: A retrospective study was conducted on 28 pediatric patients (15 with Hodgkin lymphoma and 13 with non-Hodgkin lymphoma), aged 3-16 years. CT images were acquired using a 64-detector spiral CT scanner and analyzed with OLEA research software (Olea Medical, La Ciotat, France). Regions of interest were placed on non-necrotic tumor areas in venous phase sequences.

Results: No significant age distribution difference was found between the genders, in both lymphoma groups (p=0.6). For Hodgkin lymphoma, significant texture parameters included median and gray-level non-uniformity in gray-level run-length matrix, gray-level size-zone matrix, and gray-level dependence matrix. For non-Hodgkin lymphoma, significant parameters were joint energy and sum entropy in gray-level co-occurrence matrices, zone entropy in gray-level size zone matrices, and dependence entropy in gray-level dependence matrices. Receiver operating characteristic analysis for Hodgkin lymphoma showed p-values of 0.0002-0.04, area under the curve (AUC) values of 65-78, sensitivity rates of 56-75%, and specificity rates of 58-75%. For non-Hodgkin lymphoma, significant parameters had p-values of 0.006-0.027, AUC values of 0.71-0.73, sensitivity of 47-87%, and specificity of 60-87%.

Conclusion: The study highlights the value of CT texture analysis in differentiating between Hodgkin and non-Hodgkin lymphoma in pediatric patients. By identifying significant texture parameters, this analysis method can improve diagnostic accuracy and treatment planning in childhood lymphomas.

Keywords: Texture, lymphoma, treatment

ÖZ

Amaç: Bu ön çalışma, çocukluk çağı lenfomalarında tedavi yanıtının izlenmesinde bilgisayarlı tomografi (BT) doku analizi parametrelerinin etkinliğini değerlendirmeyi amaçlamaktadır.

Gereç ve Yöntem: Üç ile on altı yaş arası 28 pediatrik hasta (15 Hodgkin lenfoma ve 13 Hodgkin olmayan lenfoma) üzerinde retrospektif bir çalışma yapıldı. BT görüntüleri 64 dedektörlü spiral BT tarayıcısı kullanılarak elde edildi ve OLEA (Olea Medical, La Ciotat, France) araştırma yazılımı ile analiz edildi. İlgi alanları, venöz faz sekanslarında nekrotik olmayan tümör alanlarına yerleştirildi.

Bulgular: Her iki lenfoma grubunda cinsiyetler arasında anlamlı yaş dağılımı farkı bulunmadı (p=0,6). Hodgkin lenfoma için önemli doku parametreleri arasında gri seviye çalışma uzunluğu matrisi, gri seviye boyut bölge matrisi ve gri seviye bağımlılık matrisinde medyan ve gri seviye homojen olmama durumu yer aldı. Hodgkin olmayan lenfoma için önemli parametreler, gri seviye eş oluşum matrisinde ortak enerji ve toplam entropi, gri seviye boyut bölge matrisinde bölge entropisi ve gri seviye bağımlılık matrisinde bağımlılık entropisi idi. Hodgkin lenfoma hastaları için alıcı işlem özellikleri eğrisi analizi, 0,0002 ile 0,04 arasında değişen p-değerleri, 65-78 eğri altında kalan alan değerleri, %56-75 duyarlılık oranları ve %58-75 özgüllük oranları gösterdi. Hodgkin olmayan lenfoma hastaları için önemli parametreler 0,006 ile 0,027 arasında p-değerleri, 0,71-0,73 eğri altında kalan alan değerleri, %47-87 duyarlılık değerleri ve %60-87 özgüllük değerleri gösterdi.

Address for Correspondence: Prof. Mehmet Öztürk, Selçuk University Faculty of Medicine, Department of Radiology, Division of Pediatric Radiology, Konya, Türkiye
E-mail: drmehmet2121@gmail.com **ORCID ID:** orcid.org/0000-0002-4316-5038

Received: 10.12.2024
Accepted: 14.02.2025
Publication Date: 03.09.2025

Cite as: Altındaş İ, Gürhan A, Kara B, Bayramoğlu Z, Öztürk M, Köksal Y. Computed tomography texture analysis for monitoring treatment response in childhood lymphomas: a preliminary study. Med J Bakirkoy. 2025;21(3):259-266

ÖZ

Sonuç: Çalışma, pediatrik hastalarda Hodgkin ve Hodgkin olmayan lenfoma ayrımında BT doku analizinin değerini vurgulamaktadır. Önemli doku parametrelerini belirleyerek, bu analiz yöntemi çocukluk çağı lenfomalarında tanı doğruluğunu ve tedavi planlamasını iyileştirebilir.

Anahtar Kelimeler: Doku, lenfoma, tedavi

INTRODUCTION

Lymphomas rank as the third most common cancer in children, following leukemia and brain tumors (1). These malignant tumors arise from genetic abnormalities in immune system cells or their precursors and encompass a range of pathological subtypes. Lymphomas can originate from various immune system components, including B cells, T cells, and natural killer cells. Genetic alterations in these cells lead to uncontrolled cell growth and tumor formation. Lymphomas are primarily classified as Hodgkin's lymphoma (HL) and non-HL (NHL) (2). Treatment follow-up depends on factors such as the disease type, stage, and overall health of the patient. Treatments may include chemotherapy, radiotherapy, immunotherapy, and stem cell transplantation. Due to their diverse subtypes and complex structure, lymphomas pose challenges for treatment, necessitating ongoing research and clinical trials (3).

Computed tomography (CT) is a frequently used imaging technique for diagnosing and managing pediatric lymphomas. It plays a crucial role in diagnosis, staging, and follow-up by providing detailed anatomical images to evaluate the size and structure of abnormal lymph nodes. However, CT alone has limitations, such as potentially inadequate evaluation, if lymph nodes do not significantly shrink after treatment (4). Combining CT with fluoro-2-deoxy-glucose (FDG)-positron emission tomography (PET) enhances the assessment by evaluating the metabolic activity of the lymph nodes. FDG-PET is highly sensitive in detecting malignant cells through glucose metabolism measurement, which is particularly useful for diagnosing metabolically active tumors like lymphomas and monitoring treatment response. FDG-PET/CT enables early evaluation of treatment effectiveness and can identify residual disease (5). However, FDG-PET has limitations, such as the potential for false positives due to FDG retention in inflammatory and infectious processes, which can lead to diagnostic errors. Hence, FDG-PET results should be interpreted cautiously and within a clinical context (6).

CT texture analysis (CTTA) is an advanced technique that assesses tissue characteristics, including structure, microarchitecture, symmetry, and the degree of homogeneity or heterogeneity (7). This quantitative, mathematical method analyzes spatial heterogeneity in medical imaging and pixel density interrelationships. CTTA helps differentiate between benign and malignant lesions and identifies more biologically aggressive tumors. It also evaluates tissue microenvironment heterogeneity, tumor grade, cellular processes like hypoxia and angiogenesis, and genetic features such as mutation status and treatment response. Additionally, CTTA can measure fibrosis (8). Recently, CTTA has been applied to distinguish malignant from benign lymphadenopathies in children, assess metastatic versus non-metastatic lung nodules, and diagnose and grade immature teratomas (9-11). This technique improves understanding of tissue microstructure and tumor behavior, offering significant benefits in accurate diagnosis and treatment planning when used with traditional imaging methods. It also provides valuable information for evaluating treatment response and disease prognosis (12).

This study uses texture analysis parameters to assess contrast-enhanced CT scans of patients with Hodgkin and NHL before and after treatment. It will also explore the correlation between these parameters and patient treatment responses. By investigating structural and textural changes in lymph nodes and surrounding tissues, the study seeks to offer deeper insights into treatment effectiveness and identify potential predictive markers for treatment outcomes.

METHODS

Study Design and Patient Selection

This retrospective study was meticulously conducted in a single-center university hospital after obtaining approval from the Selçuk University Local Ethics Committee (approval number: 2021/175, date: 07.04.2021). The research complied with the ethical guidelines of the Declaration of Helsinki. Over twelve years, from March 2017 to March 2021, CT images of 35 patients with a pathological diagnosis of either HL or NHL were reviewed retrospectively. However, 7 patients were excluded due to artifacts, and issues with contrast-enhanced timing. Consequently, the final analysis included 28 patients, with 15 diagnosed with HL and 13 with NHL. A total of 28 mediastinal lymphadenopathy measurements were made. The CT images used in the study were sourced from the picture archiving and communication system.

Computed Tomography Examination Protocol and Imaging Analysis

Images were acquired using a 64-detector spiral CT scanner (Aera, Siemens, Erlangen, Germany) with low kilovolt peak (kVp) settings. The CT protocol included the injection of non-ionic iodinated contrast material calculated according to the patient's weight (370 mg/mL iopromide; Ultravist, Bayer Vital, Leverkusen, Germany) into an antecubital vein at a rate of 2.5 mL/sec, followed by a 50 mL saline flush at the same rate using an auto-injector (Stellant, Medtron, Saarbruecken, Germany). Thoracic images were captured 30 seconds after the contrast injection, and abdominal/pelvic images were taken 70 seconds post-contrast. Scanning parameters included a 0.6 mm collimation, a 0.5-second rotation time, and a 0.6 mm increment. Images were obtained at a tube voltage of 120 kVp, with a tube current of 180 mAs for the neck, 100 mAs for the chest, and 200-250 mAs for the abdomen. The matrix size was 512x512, with a window center of 50 and 300 and a window width of 300 and 1,500, using a soft tissue reconstruction kernel (B30f).

Computed Tomography Texture Analysis

CT images were obtained using a Toshiba CT scanner (Toshiba Medical Systems, Tokyo, Japan), and CTTA was performed using OLEA research software (Olea Medical, La Ciotat, France) by placing freehand regions of interest on non-necrotic areas of the tumor in arterial phase sequences of lymphadenopathies (Figures 1 and 2). The analysis was conducted by a researcher with 2.5 years of general radiology experience and another with 10 years of pediatric radiology experience. All data were processed in Microsoft Office Excel, and statistical analysis was conducted using SPSS (Version 21.0; IBM Corp., Armonk, NY, USA). The distribution of the data was assessed using the Kolmogorov-Smirnov test. Descriptive statistics were expressed as minimum, maximum, mean, standard deviation, and median, with variability measured by the interquartile range. The texture analysis utilized eight first-order statistical parameters (entropy, mean, median, standard deviation, skewness, kurtosis, variance, and uniformity) and eleven second-order statistical parameters. The second-order parameters included difference variance, joint energy, sum average, and sum entropy within the gray-level co-occurrence matrix (GLCOM); gray-level non-uniformity and run entropy within the gray-level run length matrix (GLRLM); zone entropy within the gray-level size zone matrix (GLSZM); contrast within the gray tone difference matrix (GTDM); and gray-level non-uniformity and dependence entropy within the gray-level dependence matrix (GLDM).

Figure 1. Axial CT image; in a patient with HL, the pre-treatment tissue analysis program monitors the results of tissue analysis of the ROI area (yellow area) and statistical parameters drawn for lymphadenopathies in the mediastinum

CT: Computed tomography, HL: Hodgkin's lymphoma, ROI: Regions of interest

Statistical Analysis

All data were processed in Microsoft Office Excel, and statistical analysis was performed using SPSS (version 21.0, IBM Corp.). The Kolmogorov-Smirnov test was used to evaluate the data distribution. Descriptive statistics were presented as minimum, maximum, mean, standard deviation, and median with the interquartile range. In texture analysis, eight first-order statistical parameters (entropy, mean, median, standard deviation, skewness, kurtosis, variance, and uniformity) and eleven second-order statistical parameters (difference variance, joint energy, sum average, sum entropy within the GLCOM, gray-level non-uniformity and run entropy within the GLRLM, gray-level non-uniformity and zone entropy within the GLSZM, contrast within the GTDM, gray-level non-uniformity and dependence entropy within the GLDM) were utilized. Wilcoxon's signed-rank test, was used to compare texture parameters for evaluating the treatment response of pathological lymph nodes in HL and NHL patients, with a p-value considered significant at <0.05.

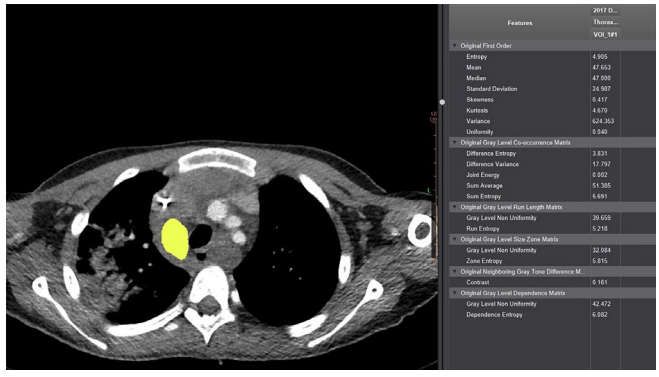

Figure 2. Axial CT image; 1 in Figure 1. In the tissue analysis program after cure chemotherapy, the results of tissue analysis of the ROI area (yellow area) and statistical parameters drawn for lymphadenopathies with shrinkage in the mediastinum are monitored

CT: Computed tomography, ROI: Regions of interest

A three-step feature selection process was used to obtain the best tissue metrics for each CT cross-section. Receiver operating characteristic (ROC) curve analysis and Pearson correlation tests were performed.

RESULTS

The study analyzed 28 patients at our center, with 15 (53%) diagnosed with HL and 13 (47%) with NHL. In the HL group, there were 8 females and 7 males, while the NHL group comprised 7 females and 6 males, with all patients aged between 3 and 16 years. The average age of female HL patients was 10, and for males, it was 11.13. In the NHL group, the average age for females was 10.38; for males, it was 13.7. Statistical analysis indicated no significant age distribution difference between males and females in HL and NHL groups (p=0.6).

Texture analysis using 19 parameters revealed statistically significant results for four parameters in differential diagnosis. For HL, the median values were gray-level non-uniformity in GLRLM, GLSZM, and GLDM. For NHL, the significant parameters were joint energy and sum entropy in GLCOM, zone entropy in GLSZM, and dependence entropy in GLDM. The median is a first-order texture feature, while gray-level non-uniformity, joint energy, sum entropy, zone entropy, and dependence entropy are second-order texture features (Tables 1 and 2).

In the ROC analysis for HL patients, the statistically significant parameters were the median and gray-level non-uniformity in GLRLM, GLSZM, and GLDM, with p-values of 0.04, 0.0002, 0.004, and 0.01, respectively. The area under the curve (AUC) for these parameters was 65, 78, 78, and 77, respectively. The sensitivity rates were 56%, 75%, 75%, and 68%, and the specificity rates were 58%, 75%, 75%, and 75%, respectively. The diagnostic accuracy values were 62%, 78%, 78%, and 77%, respectively (Table 3).

For NHL patients, the ROC analysis showed that the statistically significant parameters were joint energy and sum entropy in GLCOM, zone entropy in GLSZM, and

Table 1. Comparison of pre-and post-treatment values of CTTA parameters of HL

Hodgkin lymphoma (n=18)	**Before treatment**	**Post-treatment**	**p-value**
First-order statistical parameters	**Median (IQR)**	**Median (IQR)**	
Entropy	5.18 (5.01-5.34)	5.2 (5.09-5.41)	0.7
Mean	56.9 (52.6-70.2)	46.2 (33.6-67.4)	0.056
Median	57.5 (52.6-70.7)	48 (32.5-67)	0.04
SD	14.33 (8.73-25.7)	12.9 (10-20.4)	0.3
Skewness	-0.04 (-0.13-0.03)	0.09 (-0.04-0.5)	0.056
Kurtosis	2.96 (2.79-3.27)	2.92 (2.6-3.4)	0.57
Variance	205 (77-665)	167.6 (99.6-418)	0.37
Uniformity	0.03 (0.03-0.04)	0.03 (0.03-0.04)	0.48
Second-order statistical parameters			
Gray-level co-occurence matrix (GLCOM)			
Difference variance	35.5 (18.5-49)	18.84 (12.6-44.6)	0.44
Joint energy	0 (0-0)	0 (0-0.01)	0.058
Sum average	63.4 (60-68.3)	66.9 (53.6-74.2)	0.9
Sum entropy	5.99 (5.86-6.11)	5.84 (5.5-6.1)	0.15
Gray-level run length matrixv (GLRLM)			
Gray-level non-uniformity	24.6 (15.8-37.4)	10.56 (5.12-21.22)	0.002
Run entropy	5.43 (5.29-5.55)	5.48 (5.15-5.63)	0.68
Gray-level size zone matrix (GLSZM)			
Gray-level non-uniformity	21.2 (12.1-28.2)	8.39 (4.93-16.8)	0.004
Zone entropy	5.93 (5.74-6.1)	5.92 (5.32-6.08)	0.2
Contrast	0.4 (0.21-0.6)	0.45 (0.27-0.99)	0.1
Gray-level dependence matrix (GLDM)			
Gray-level non-uniformity	28.9 (14.5-40.6)	9.83 (5.12-22.26)	0.01
Dependence entropy	6.15 (5.96-6.34)	6.14 (5.47-6.3)	0.15

IQR: Interquartile range, HL: Hodgkin's lymphoma, CTTA: Computed tomography texture analysis, SD: Standard deviation

Table 2. Results of ROC analysis of CTTA parameters differing before and after treatment in HL

CTTA parameters	AUC	p	Cut-off value	Sensitivity	Specificity	95% CI	
						Lower limit	Upper limit
First-order statistical parameters							
Median	65	0.04	56	56	58	46	85
Gray-level run length matrixv (GLRLM)							
Gray-level non-uniformity	78	0.002	16	75	75	61	94
Gray-level size zone matrix (GLSZM)							
Gray-level non-uniformity	78	0.004	14	75	75	62	93
Gray-level dependence matrix (GLDM)							
Gray-level non-uniformity	77	0.01	20	68	75	60	94

CI: Confidence interval, AUC: Area under the curve, CTTA: Computed tomography texture analysis, ROC: Receiver operating characteristic, HL: Hodgkin's lymphoma

Table 3. Comparison of pre- and post-treatment values of CTTA parameters of NHL

Non-Hodgkin lymphoma (n=17)	Before treatment	Post-treatment	p-value
CTTA parameters	**Median (IQR)**	**Median (IQR)**	
First-order statistical parameters			
Entropy	5.27 (4.87-5.31)	5.05 (4.72-5.37)	0.33
Mean	55.9 (44.5-67.5)	53.35 (44.88-62.36)	0.6
Median	56 (41-67)	53 (44.9-62.36)	0.6
SD	13.04 (9.81-24.2)	15.4 (12.6-26.4)	0.3
Skewness	0.03 (-0.23-0.2)	0.02 (-0.2-0.28)	0.65
Kurtosis	3.1 (2.83-3.66)	2.9 (2.75-3.53)	0.9
Variance	170.1 (59.4-584.5)	236.71 (158.37-698.5)	0.28
Uniformity	0.03 (0.03-0.04)	0.03 (0.03-0.04)	0.48
Second order statistical parameters			
Gray-level co-occurence matrix (GLCOM)			
Difference variance	29.5 (22-39.7)	27 (17.8-89.2)	0.08
Joint energy	0 (0-0)	0 (0-0.1)	0.006
Sum average	62.3 (59.4-73.9)	62.67 (57.69-67.4)	0.23
Sum entropy	6 (5.8-6.11)	5.7 (5.33-5.9)	0.02
Gray-level run length matrix (GLRLM)			
Gray-level non-uniformity	14.6 (9.33-57.36)	11.8 (2.34-22.2)	0.21
Run entropy	5.51 (5.3-5.56)	5.28 (4.68-5.57)	0.07
Gray-level size zone matrix (GLSZM)			
Gray-level non-uniformity	12.54 (7.82-40.5)	10.5 (2.07-17.08)	0.25
Zone entropy	5.98 (5.76-6.08)	5.82 (5.18-5.86)	0.027
Contrast	0.37 (0.23-0.66)	0.69 (0.28-2.17)	0.11
Gray-level dependence matrix (GLDM)			
Gray-level non-uniformity	15.29 (10.5-60.48)	12.35 (2.44-24.2)	0.19
Dependence entropy	6.25 (5.99-6.34)	5.98 (5.28-6.08)	0.023

CTTA: Computed tomography texture analysis, NHL: Non-Hodgkin's lymphoma, SD: Standard deviation, IQR: Interquartile range

dependence entropy in GLDM, with p-values of 0.006, 0.02, 0.027, and 0.023, respectively. The AUC for these parameters was 0.72, 0.71, 0.73, and 0.72, respectively. The sensitivity values were 47%, 87%, 67%, and 80%, and the specificity values were 87%, 60%, 87%, and 60%, respectively. The diagnostic accuracy values were 72%, 71%, 73%, and 72%, respectively (Figures 3, 4 and Table 3b). No significant differences were found in other parameters (Table 4).

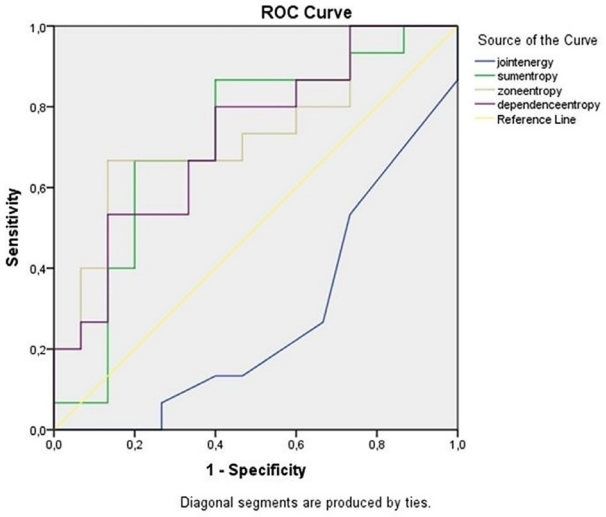

Figure 3. ROC curve of CTTA parameters differing before and after treatment in HL
CTTA: Computed tomography texture analysis, ROC: Receiver operating characteristic, HL: Hodgkin's lymphoma

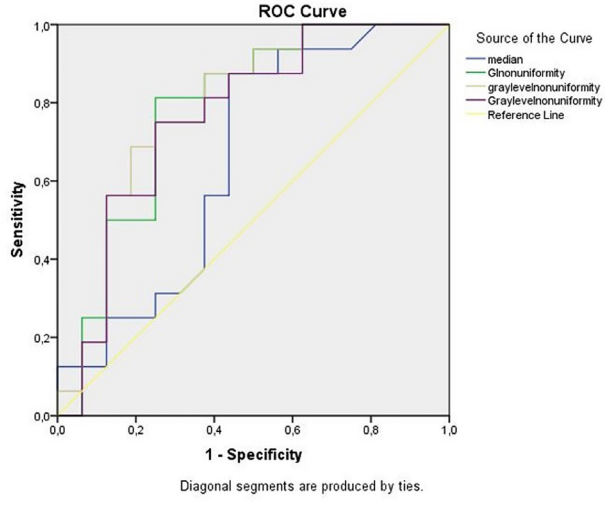

Figure 4. ROC curve of CTTA parameters differing before and after treatment in NHL
CTTA: Computed tomography texture analysis, ROC: Receiver operating characteristic, NHL: Non-Hodgkin's lymphoma

DISCUSSION

This preliminary study aimed to evaluate the efficacy of CTTA parameters in monitoring treatment response in childhood lymphomas. The results demonstrated that CTTA could effectively differentiate between HL and NHL in pediatric patients by identifying significant texture parameters. The texture analysis revealed that certain parameters significantly differed between HL and NHL, suggesting their potential role in differential diagnosis. For HL, significant parameters included the median and gray-level non-uniformity in GLRLM, GLSZM, and GLDM. These parameters reflect the heterogeneity within the tumor, which may correlate with different biological behaviors of HL. The median value in texture analysis often represents the central tendency of pixel intensity, providing insights into the overall density and structure of the tumor. Gray-level non-uniformity measures the variability of gray-levels within the tumor, indicating the degree of heterogeneity which is often higher in malignant tissues due to irregular cell growth and necrosis (13). Similarly, for NHL, significant parameters was joint energy and sum entropy in GLCOM, zone entropy in GLSZM, and dependence entropy in GLDM. Joint energy in GLCOM reflects the uniformity of pixel pairs, with lower values indicating more disorder and heterogeneity, which are characteristic of aggressive tumors. Sum entropy measures the complexity and randomness within the tumor texture, providing insights into the tumor's metabolic activity and potential aggressiveness. Zone entropy in GLSZM and dependence entropy in GLDM further quantify the spatial distribution and dependency of pixel intensities, crucial for understanding the tumor's microenvironment and potential response to treatment (14).

The ROC analysis provided further insights into the diagnostic accuracy of these texture parameters. For HL, the AUC values ranged from 65 to 78, indicating moderate to high diagnostic accuracy. The sensitivity and specificity rates were also promising, suggesting that CTTA could be a reliable tool for monitoring treatment response in HL patients. For instance, a sensitivity rate of 75% indicates that CTTA correctly identified 75% of true positive cases, while a specificity rate of 75% indicates that it detected 75% of true negative cases. These metrics are crucial for ensuring that patients receive accurate diagnoses and appropriate treatments (15). For NHL, the AUC values were slightly higher, ranging from 0.71 to 0.73, indicating a strong potential for CTTA in differentiating NHL from HL, supported by sensitivity and specificity rates. The higher AUC values for NHL suggest that the identified texture parameters are

Table 4. Results of ROC analysis of CTTA parameters differing before and after treatment in NHL

CTTA parameters	AUC	p	Cut-off value	Sensitivity	Specificity	95% CI	
						Lower limit	Upper limit
Gray-level co-occurence matrix (GLCOM)							
Joint energy	0.72	0.006	0.0	47	87	0.54	0.90
Sum entropy	0.71	0.02	5.71	87	60	0.09	0.48
Gray-level size zone matrix (GLSZM)							
Zone entropy	0.73	0.027	5.87	67	87	0.08	0.45
Gray-level dependence matrix (GLDM)							
Dependence entropy	0.72	0.023	6	80	60	0.09	0.45

CI: Confidence interval, AUC: Area under the curve, CTTA: Computed tomography texture analysis, ROC: Receiver operating characteristic, NHL: Non-Hodgkin's lymphoma

more robust in distinguishing NHL from other lymphomas, which is essential for tailoring specific treatment protocols for these patients (16).

Our findings align with previous studies that highlighted the utility of texture analysis in cancer diagnosis and treatment monitoring. For instance, Ganeshan et al. (8) demonstrated that texture analysis could differentiate between benign and malignant lung nodules by assessing the heterogeneity within the tumor. Similarly, Lubner et al. (17) showed that texture analysis could assess hepatic fibrosis by quantifying the structural changes in liver tissue. These studies support the notion that texture analysis can provide valuable information about tumor microenvironment and treatment response, reinforcing the potential of CTTA in pediatric lymphomas.

Accurately differentiating between HL and NHL using CTTA has significant clinical implications. First, it can aid in early diagnosis and appropriate treatment planning, which is crucial for improving patient outcomes. Early and accurate diagnosis allows for timely intervention, potentially reducing the disease burden and improving survival rates (12). Secondly, CTTA can provide a non-invasive method to monitor treatment response, reducing the need for repeated biopsies and invasive procedures. This is particularly important in pediatric patients, for whom minimizing radiation exposure and invasive interventions is a priority. By providing a detailed assessment of tumor heterogeneity and response to treatment, CTTA can help clinicians make informed decisions about modifying or continuing treatment regimens (11). Furthermore, CTTA's ability to evaluate tissue microenvironment heterogeneity, tumor grade, cellular processes like hypoxia and angiogenesis, and genetic features such as mutation status and treatment response, adds another layer of diagnostic precision. This comprehensive evaluation can lead to more personalized treatment plans, potentially improving the efficacy of therapeutic interventions (18,19).

Study Limitations

Despite the promising results, this study has several limitations. The sample size was relatively small, and the study was conducted at a single-center, which may limit the generalizability of the findings. Additionally, the retrospective nature of the study may introduce selection bias. Future studies with larger, multi-center cohorts are needed to validate these findings and explore the potential of CTTA in other pediatric cancers. Moreover, prospective studies could provide more robust evidence on the utility of CTTA in real-time clinical settings (3). Moreover, integrating CTTA with other imaging modalities, such as FDG-PET, could enhance the accuracy of treatment response assessment. FDG-PET provides metabolic information, while CTTA offers insights into tissue heterogeneity; combining these techniques may provide a comprehensive evaluation of the tumor. Studies have shown that FDG-PET/CT is highly sensitive in detecting malignant cells through glucose metabolism measurement, which is particularly useful for diagnosing metabolically active tumors like lymphomas and monitoring treatment response (5). However, FDG-PET has limitations, such as the potential for false positives due to FDG retention in inflammatory and infectious processes (6). Hence, combining FDG-PET with CTTA could mitigate these limitations and provide a more accurate assessment of treatment response. Additionally, advancements in artificial intelligence and machine learning could further enhance the capabilities of CTTA. Algorithms designed to analyze texture patterns could be trained to recognize subtle changes in tumor characteristics, potentially leading to earlier treatment response or resistance detection. This integration could revolutionize how we monitor and treat pediatric lymphomas, making the process more efficient and precise (9).

CONCLUSION

In conclusion, this preliminary study demonstrated that CTTA is valuable in differentiating between HL and NHL in pediatric patients. Significant texture parameters can aid in differential diagnosis and treatment response assessment. These findings suggest that CTTA can enhance diagnostic accuracy and treatment planning, providing deeper insights into treatment effectiveness and potential predictive markers for treatment outcomes. Future research should focus on validating these results in larger cohorts and integrating CTTA with other imaging modalities for a more comprehensive evaluation. Additionally, exploring the potential of CTTA in other pediatric cancers could further expand its clinical utility and improve patient care.

ETHICS

Ethics Committee Approval: This retrospective study was meticulously conducted in a single-center university hospital after obtaining approval from the Selçuk University Local Ethics Committee (approval number: 2021/175, date: 07.04.2021).

Informed Consent: Retrospective study.

FOOTNOTES

Authorship Contributions

Surgical and Medical Practices: Y.K., Concept: İ.A., Z.B., Design: İ.A., B.K., Z.B., Data Collection or Processing: A.G., B.K., Z.B., M.Ö., Analysis or Interpretation: A.G., B.K., M.Ö., Literature Search: A.G., M.Ö., Writing: M.Ö.

Conflict of Interest: No conflict of interest was declared by the authors.

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

REFERENCES

1. Swerdlow SH, Campo E, Pileri SA, Harris NL, Stein H, Siebert R, et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. Blood. 2016;127:2375-90.
2. Hunger SP, Mullighan CG. Acute lymphoblastic leukemia in children. N Engl J Med. 2015;373:1541-52.
3. Sehn LH, Salles G. Diffuse large B-cell lymphoma. N Engl J Med. 2021;384:842-58.
4. Johnson SA, Kumar A, Matasar MJ, Schöder H, Rademaker J. Imaging for staging and response assessment in lymphoma. Radiology. 2015;276:323-38.
5. Cheson BD, Fisher RI, Barrington SF, Cavalli F, Schwartz LH, Zucca E, et al. Recommendations for initial evaluation, staging, and response assessment of Hodgkin and non-Hodgkin lymphoma: the Lugano classification. J Clin Oncol. 2014;32:3059-68.
6. Adams HJ, Kwee TC, de Keizer B, Fijnheer R, de Klerk JM, Nievelstein RA. FDG PET/CT for the detection of bone marrow involvement in diffuse large B-cell lymphoma: systematic review and meta-analysis. Eur J Nucl Med Mol Imaging. 2014;41:565-74.
7. Lubner MG, Smith AD, Sandrasegaran K, Sahani DV, Pickhardt PJ. CT texture analysis: definitions, applications, biologic correlates, and challenges. Radiographics. 2017;37:1483-503.
8. Ganeshan B, Panayiotou E, Burnand K, Dizdarevic S, Miles K. Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: a potential marker of survival. Eur Radiol. 2012;22:796-802.
9. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 20143;5:4006.
10. Mayerhoefer ME, Materka A, Langs G, Häggström I, Szczypiński P, Gibbs P, et al. Introduction to radiomics. J Nucl Med. 2020;61:488-95.
11. Orlhac F, Boughdad S, Philippe C, Stalla-Bourdillon H, Nioche C, Champion L, et al. A postreconstruction harmonization method for multicenter radiomic studies in PET. J Nucl Med. 2018;59:1321-28.
12. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology. 2016;278:563-77.
13. Ganeshan B, Miles KA, Young RC, Chatwin CR. Texture analysis in non-contrast enhanced CT: impact of malignancy on texture in apparently disease-free areas of the liver. Eur J Radiol. 2009;70:101-10.
14. Castellino SM, Geiger AM, Mertens AC, Leisenring WM, Tooze JA, Goodman P, et al. Morbidity and mortality in long-term survivors of Hodgkin lymphoma: a report from the childhood cancer survivor study. Blood. 2011;117:1806-16.
15. Barrington SF, Kluge R. FDG PET for therapy monitoring in Hodgkin and non-Hodgkin lymphomas. Eur J Nucl Med Mol Imaging. 2017;44:97-110.
16. Boellaard R, Delgado-Bolton R, Oyen WJ, Giammarile F, Tatsch K, Eschner W, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. Eur J Nucl Med Mol Imaging. 2015;42:328-54.
17. Lubner MG, Jones D, Kloke J, Said A, Pickhardt PJ. CT texture analysis of the liver for assessing hepatic fibrosis in patients with hepatitis C virus. Br J Radiol. 2019;92:20180153.
18. Chalkidou A, O'Doherty MJ, Marsden PK. False discovery rates in PET and CT studies with texture features: a systematic review. PLoS One. 2015;10:e0124165.
19. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer. 2012;48:441-6.

Medical Journal of Bakırköy

Med J Bakirkoy 2025;21(3):267-273
DOI: 10.4274/BMJ.galenos.2025.2025.1-25

Research

Assessment of Liver Damage in an Experimental Model of Streptozotocin-Induced Diabetes in Sprague Dawley Rats: Focusing on Morphological and Oxidative Stress Status

Sprague Dawley Sıçanlarında Streptozotosin ile İndüklenmiş Deneysel Diyabet Modelinde Karaciğer Hasarının Değerlendirilmesi: Morfolojik ve Oksidatif Stres Durumuna Odaklanma

Serbay Özkan[1], Meral Koyutürk[2]

[1]İzmir Katip Çelebi University Faculty of Medicine, Department of Histology and Embryology, İzmir, Türkiye

[2]İstanbul University-Cerrahpaşa, Cerrahpaşa Faculty of Medicine, Department of Histology and Embryology, İstanbul, Türkiye

ABSTRACT

Objective: Diabetes mellitus (DM) is a widespread metabolic disorder that causes severe complications, including damage to various organs. The liver, a pivotal organ involved in metabolic regulation, is particularly vulnerable to diabetic injury. This study aimed to investigate the effects of diabetes on the liver, using streptozotocin (STZ)-induced diabetic model in Sprague Dawley rats.

Methods: Twelve male rats were divided into control and diabetic groups. Diabetes was induced via intraperitoneal injection of a single dose of STZ (55 mg/kg), and after eight weeks, the rats were sacrificed, and the liver tissues were collected for analysis of histological changes, apoptosis, and oxidative stress markers.

Results: Histological evaluations demonstrated significant alterations in diabetic liver tissue, characterized by hepatocyte degeneration, fibrosis, and cellular swelling. Immunohistochemical analysis revealed elevated active caspase-3 expression, indicating increased apoptosis in diabetic livers. Additionally, the analysis of oxidative stress markers revealed a significant increase in H_2O_2 levels, while superoxide dismutase activity remained unchanged, suggesting an oxidative imbalance.

Conclusion: These findings confirm that diabetes leads to liver damage through mechanisms including apoptosis, fibrosis, and oxidative stress, emphasizing the significance of therapeutic approaches targeting these pathways. In this respect, future research should focus on these mechanisms in other DM models and include additional liver injury biomarkers to better understand the progression of diabetic liver disease.

Keywords: Diabetes, liver damage, morphologic evaluation, apoptosis, oxidative stress

ÖZ

Amaç: Diabetes mellitus (DM), çeşitli organ hasarları da dahil ciddi komplikasyonlara neden olan yaygın bir metabolik hastalıktır. Metabolik aktivitelerin düzenlenmesindeki rolüyle önemli bir organ olan karaciğer, diyabetik hasara karşı oldukça hassastır. Bu çalışmada, Sprague Dawley sıçanlarında streptozotosin (STZ) ile indüklenen diyabet modeli kullanılarak diyabetin karaciğer üzerindeki etkilerinin araştırılması amaçlandı.

Gereç ve Yöntem: On iki erkek sıçan kontrol ve diyabetik gruplara ayrıldı. İntraperitoneal tek doz STZ (55 mg/kg) enjeksiyonu ile diyabet oluşturuldu ve sekiz hafta sonunda hayvanlar sakrifiye edildi ve sonrasında karaciğer dokuları histolojik değişiklikler, apoptoz ve oksidatif stres belirteçleri açısından değerlendirildi.

Bulgular: Histolojik değerlendirmeler diyabetik karaciğer dokusunda hepatosit dejenerasyonu, fibrozis ve hücresel şişme ile karakterize önemli değişikliklerin olduğunu göstermiştir. İmmünohistokimyasal analiz, aktif kaspaz-3 ekspresyonunun dolayısıyla diyabetik karaciğerlerde apoptozun arttığını ortaya koymuştur. Ek olarak, oksidatif stres belirteçlerinin analizi, H_2O_2 düzeylerinde önemli bir artış olduğunu diğer taraftan süperoksit dismutaz aktivitesinin değişmediğini göstermiştir. Bu durum oksidatif stres durumunda bir dengesizlik olduğunu düşündürtmüştür.

Sonuç: Bu sonuçlar, diyabetin apoptoz, fibrozis ve oksidatif stres gibi mekanizmalar yoluyla karaciğer hasarına yol açtığını ve bu yolları hedef alan terapötik yaklaşımların önemini vurgulamaktadır. Bu bağlamda, gelecekteki araştırmalar diğer DM modellerinde bu mekanizmalara odaklanarak diyabette karaciğer patolojilerine bağlı hastalıkların ilerleyişini daha iyi anlamak ve takip etmek için yeni hepatik hasar biyobelirteçleri içermelidir.

Anahtar Kelimeler: Diyabet, karaciğer hasarı, morfolojik değerlendirme, apoptoz, oksidatif stres

Address for Correspondence: Lec, Serbay Özkan, PhD, İzmir Katip Çelebi University Faculty of Medicine, Department of Histology and Embryology, İzmir, Türkiye
E-mail: serbay.ozkan@istanbul.edu.tr **ORCID ID:** orcid.org/0000-0001-7854-4735

Received: 03.02.2025
Accepted: 04.03.2025
Publication Date: 03.09.2025

Cite as: Özkan S, Koyutürk M. Assessment of liver damage in an experimental model of streptozotocin-induced diabetes in Sprague Dawley rats: focusing on morphological and oxidative stress status. Med J Bakirkoy. 2025;21(3):267-273

INTRODUCTION

Diabetes mellitus is a significant global health issue, affecting millions worldwide and leading to serious complications in multiple organ systems (1). The disease is characterized by chronic hyperglycemia, which stems from either insufficient insulin production or the inability of cells to respond properly to insulin (2). One key mechanism linking diabetes to cellular and tissue damage across multiple organs is oxidative stress (3). Oxidative stress is the result of disruption of the balance between the antioxidant defence of the body and the production of reactive oxygen species (ROS). In diabetes, hyperglycemia contributes to increased ROS production through several mechanisms, such as glucose autoxidation, activation of protein kinase C, and the buildup of advanced glycation end products (3). These processes bring about mitochondrial dysfunction and an overproduction of free radicals, overwhelming the body's natural antioxidant defenses (4). This oxidative imbalance is not merely a consequence of diabetes but is a driving force behind its complications. ROS can bring about destructive effects on DNA, proteins, and lipids, leading to cellular malfunction, apoptosis, inflammation, and morphological damage (4). These effects are evident in several organs and contribute to the progression of diabetic complications. For example, in the kidneys, oxidative stress plays a key role in diabetic nephropathy, where increased ROS causes glomerular injury, podocyte apoptosis, and tubulointerstitial fibrosis (5). Similarly, oxidative stress leads to vascular damage in diabetic patients, promoting endothelial dysfunction, smooth muscle cell proliferation, and the formation of atherosclerotic plaques, which increase the cardiovascular disease risk (6).

The liver is one of the organs most significantly affected by diabetes, playing a vital role in glucose regulation, lipid metabolism, and detoxification. It is essential to understand the extent of liver damage caused by diabetes, as it not only disrupts these vital functions but also contributes to the progression of systemic complications associated with the disease (7). In diabetic individuals, liver dysfunction is linked to several pathological processes, including insulin resistance, increased fatty acid accumulation, inflammation, and fibrosis. These alterations can further disrupt metabolic regulation, increase the risk of developing non-alcoholic fatty liver disease (NAFLD), cirrhosis, and hepatocellular carcinoma in diabetic patients (8). However, despite its pivotal role in maintaining metabolic balance, the liver is often underrepresented in research on diabetic complications. Elucidating the mechanisms underlying diabetes-induced liver damage, including histological changes, apoptosis, and oxidative stress, is critical for developing therapeutic strategies to mitigate these harmful effects.

In this research, our main aim is to investigate the effects of diabetes on the liver using an experimental model of streptozotocin (STZ)-induced diabetes in Sprague Dawley rats. STZ is a compound that selectively destroys pancreatic β-cells, resulting in hyperglycemia and is commonly employed in experimental diabetes models (9). We explore the liver's response to hyperglycemia by administering a single intraperitoneal injection of 55 mg/kg STZ, focusing on histological changes, apoptosis, and oxidative stress markers. Understanding these hepatic alterations is crucial for identifying the cellular and molecular pathways involved in liver damage caused by diabetes and may ultimately assist in developing targeted therapies to reduce liver-related complications in diabetic patients.

METHODS

This study was conducted using experimental animals, and no patient-related biological material was utilized. Therefore, patient approval was not required.

Animal Model of Diabetes Mellitus

Animal experiments were carried out after obtaining the necessary ethical approval from the Bezmialem Vakıf University Animal Experiments Local Ethics Committee (approval no: 1294-1, date: 21.10.2024). Twelve male Sprague Dawley rats, aged 10-12 weeks and weighing between 350-450 grams, were kept in a controlled environment with a temperature of 20±2 °C, humidity levels of 45-55%, and a 12-hour light-dark cycle. They had free access to standard laboratory chow and water. Six randomly selected rats were given intraperitoneal injections of 55 mg/kg STZ (STZ; Sigma S0130) dissolved in sodium citrate buffer (SCB) (SCB; pH 4.5, 0.1 M) to induce type 1 diabetes (T1D), while the other six rats received SCB alone and served as the control group. Blood glucose levels (BGLs) were regularly monitored with a glucometer, and the rats with BGL exceeding 250 mg/dL were considered diabetic. At the end of the 3rd day following the STZ injection, all the rats were hyperglycemic (BGL >250 mg/dL). After 8 weeks, all rats were euthanized under general anesthesia induced by ketamine and xylazine.

Histological Analysis

After the animals were sacrificed, liver samples were collected, cut into small pieces, and prepared for light microscopic evaluations. The tissues were placed in 10% neutral buffered formalin for fixation and after 24 hours, they were dehydrated through the increasing alcohol series (70%, 90%, 96%, and 100%) before being cleared in toluene.

The samples were then embedded in paraffin, and 4 µm-thick sections were cut. These sections were deparaffinized, rehydrated through a descending alcohol series (100%, 96%, 90%, and 70%), and placed in distilled water. The slides were stained with hematoxylin and eosin (H&E) and Masson's trichrome stains to assess morphological changes and fibrosis in the tissue, respectively. The preparations were evaluated with the Olympus BX43 microscope, and photographed with the Olympus SC100 camera system (10).

Immunohistochemical Analysis

Following rehydration, antigen retrieval was executed with citrate buffer (6.0 pH) in a microwave for 20 min. The sections were washed with phosphate buffered saline (PBS) and the sections were treated with 3% hydrogen peroxide for 10 min at room temperature. After another round of PBS wash, a blocking agent (Thermo Scientific TP-125-HL) was applied for 6 min. Subsequently, the sections were incubated overnight at 4 °C with 1 µg/mL active caspase-3 (Thermo Fisher, MS-1123). After washing with PBS, biotinylated secondary antibody solution (Thermo Scientific TP-125-HL) was applied to the preparations for 10 minutes. After washing again with PBS, they were incubated with streptavidin peroxidase solution (Thermo Scientific TP-125-HL) for 10 minutes. Finally, the immune reaction was made visible with 3-amino-9-ethylcarbazole chromogen application (Thermo Scientific TA-125-HA). The preparations were investigated with the Olympus BX43 microscope and photographed with the Olympus SC100 camera system. To quantify the immune reaction, a 4-point scoring system was used, where 0 indicated no staining, 1 represented positive staining in less than 30% of cells per high-power field (40X), 2 indicated positive staining in 30% to 70% of cells, and 3 represented positive staining in more than 70% of cells per high-power field (11).

Evaluation of Oxidative Stress Status

The liver samples weighing approximately 20 mg were rinsed with cold PBS and homogenized in 180 uL PBS with a tissue lyser at 4 °C for 2.5 minutes at 50 oscillations. Then the samples were centrifuged at 4000 revolutions per minute for 10 min at 4 °C to remove insoluble material. The supernatant was collected and kept on ice for detection. At this point, the protein concentrations of the supernatant were assessed with a bicinchoninic acid assay (Thermo Scientific, 23225 and 23227). For analysis of total-superoxide dismutase (T-SOD) activity, 1 mL of working buffer solution was mixed with 0.07 mL of sample, and double distilled water, to set sample and control tubes, respectively. Then, 0.1 mL of nitrosogenic agent, 0.1 mL of substrate solution, and 0.1 mL of enzyme stock working solution were added into each tube, fully mixed with a vortex mixer, and incubated at 37 °C for 40 min. In the following stage, 2 mL of chromogenic agent was added into each tube and fully mixed, and kept for 10 minutes at room temperature. Optical density (OD) values of each tube were measured at 550 nm in a quartz cuvette with a 1 cm optical path, following setting zero with double distilled water. SOD concentration was calculated using the following formula: T-SOD activity (U/mL)=i/50% x V1/V2 x f; where i is the inhibition ratio calculated by the formula (ODcontrol-ODsample)/ODcontrol x 100%; V1 is the total volume of reaction solution (mL); V2 is the volume of sample added (mL); f is the dilution factor of concentration of protein (Cpr) is the concentration of protein in the sample (g prot/L).

For determination of the concentration of H_2O_2, 1 mL of buffer solution was added to 5 mL Eppendorf tubes and incubated at 37 °C for 10 min. In the following: 0.1 mL of double distilled water, 0.1 mL of 60 mmol/L H_2O_2, and 0.1 mL of sample were separately added to the tubes to prepare the blank, standard and sample tubes, respectively. 1 mL of ammonium molybdate reagent was added to each tube prepared in the previous step, and the contents were mixed thoroughly. The spectrophotometer was set to zero with double-distilled water, and then the OD value of each tube was measured at 405 nm with a 1 cm optical path quartz cuvette. H_2O_2 content of the sample was calculated by using the following formula: H_2O_2 content (mmol/L)=ΔA1/ΔA2 x C x f ÷ Cpr. Where ΔA1: ODsample-ODblank; ΔA2: ODstandard-ODblank; C: The concentration of H_2O_2 standard, 60 mmol/L; f: The dilution factor of the sample before the test; Cpr: the concentration of protein in the sample (g prot/L).

Statistical Analysis

Statistical analyses were executed using SPSS version 20.0 (IBM Corp., Armonk, NY, USA). Data were presented as mean ± standard error. The Shapiro-Wilk test was performed to assess normality. For data that are normally distributed, one-way ANOVA and post hoc Tukey tests were performed to compare group means. For non-normally distributed data, the Kruskal-Wallis test was used. A p-value of less than 0.05 ($p<0.05$) was considered statistically significant.

RESULTS

Histological Evaluations

H&E staining of liver sections from control rats showed that the liver has a normal lobular architecture, characterized by a central vein with hepatic cords extending from it, separated by well-defined blood sinusoids. The portal tracts, located

at the periphery of the hepatic lobules, were composed of the typical triad: branches of the portal vein, hepatic artery, and bile duct. Hepatocytes appeared large and polyhedral, with slightly eosinophilic granular cytoplasm. They had rounded euchromatic nuclei with prominent nucleoli, and some cells were observed to be binucleated (Figure 1a). In contrast, liver sections from diabetic rats revealed significant histopathological alterations. The hepatic parenchyma was disorganized, with evident dilation and congestion of the central veins and hepatic sinusoids. Hepatocytes exhibited degenerative changes including cellular disorganization and vacuolization. In some sections, there were focal areas of fibrosis accompanied by cellular proliferation. Furthermore, shrunken pyknotic nuclei were noted in some of the hepatocytes along with intense eosinophilic cytoplasm in diabetic conditions, all of which indicate apoptosis (Figure 1b). On the other hand, Masson's trichrome staining, which highlights collagen deposition in light blue, demonstrated that liver sections from diabetic rats displayed notable fibrosis, particularly prominent around the central veins and at the peripheries of the portal triads (Figure 1c and 1d).

The liver samples of two diabetic rats displayed cystic dilations of ductules which are characterized by biliary-like epithelium, surrounded by increased dense fibrotic tissue (Figure 2).

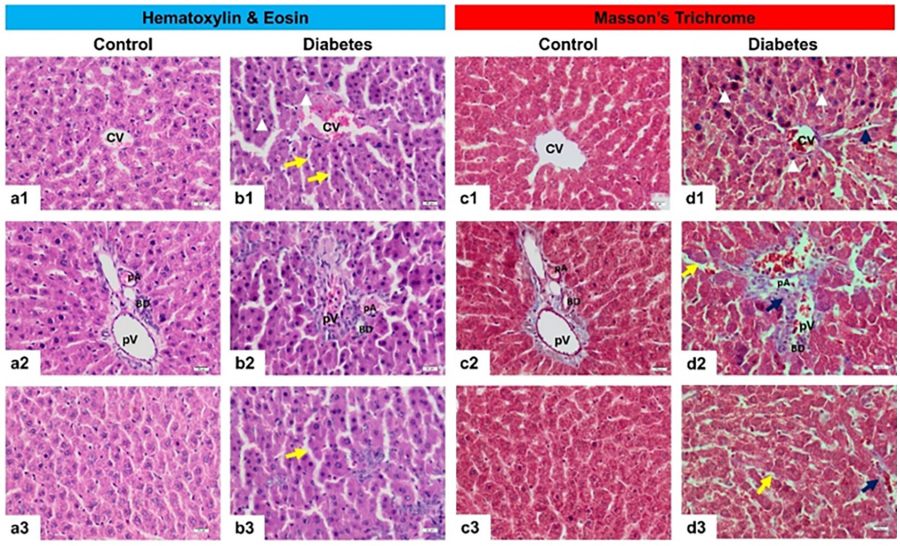

Figure 1. **(a,b)** Representative figures for hematoxylin & eosin and **(c,d)** Masson's trichrome staining. Central vein (CV), portal artery (pA), portal vein (pV), bile duct (BD), dilated sinusoidal capillaries (yellow arrow), pyknotic nuclei (white arrowhead) and fibrosis (purple arrow)

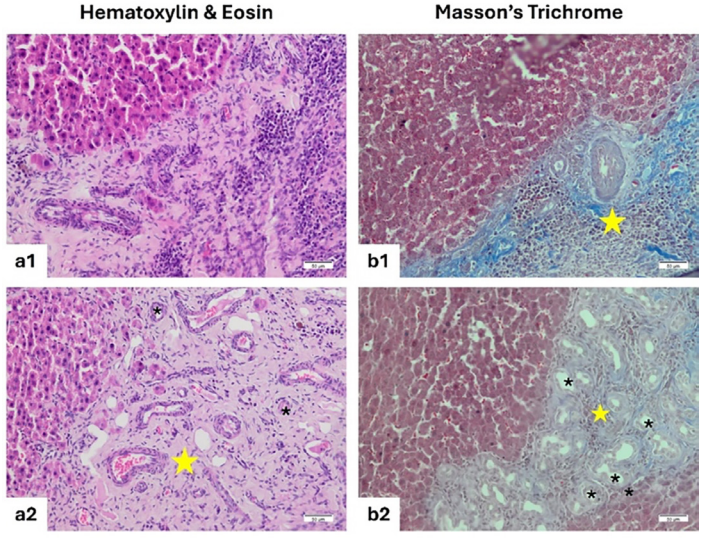

Figure 2. **(a)** Representative figures for the cystic dilations of ductules stained with hematoxylin & eosin and **(b)** Masson's trichrome. Cystic dilations of ductules (asterisk) were surrounded by increased dense fibrotic tissue (yellow star)

Immunohistochemical Analysis

Active caspase-3 expression was determined with immunohistochemical analysis. Semi-quantitative scoring of active caspase-3 staining showed that diabetes significantly increased the apoptotic death of hepatocytes in comparison to the control group (Figure 3).

Evaluation of Oxidative Stress Status

The oxidative stress status of the obtained liver tissue samples was determined by the spectrophotometric analysis of SOD activity and H_2O_2 concentration. It was found that there was no statistical difference for the SOD activity between the control and diabetes groups, while there was a statistically significant increase noted for the H_2O_2 level of the diabetes group, compared to that of the control group (Figure 4).

DISCUSSION

The findings of this study contribute to an increasing body of evidence on the role of diabetes in liver damage, with a particular emphasis on oxidative stress, histopathological alterations, and apoptosis using an STZ-induced diabetes model in Sprague Dawley rats. Our results showed that diabetes led to significant histopathological alterations in the liver, increased apoptosis as evidenced by active caspase-3 expression, and increased oxidative stress, particularly reflected by elevated H_2O_2 levels. These results are well aligned with the previous research on diabetic hepatopathy and further reinforce the detrimental impact of hyperglycemia and oxidative stress on hepatic tissue.

Histologically, liver sections from diabetic rats exhibited notable alterations, including hepatocyte degeneration, cellular swelling, and fibrotic changes. These observations are consistent with earlier studies that have shown structural disorganization of liver tissue and fibrosis in diabetic models (7,12). Fibrosis, a recognized outcome of chronic liver injury and a key feature of diabetic liver damage, was observed in focal areas, particularly around central veins and portal triads. These results support the idea that hyperglycemia accelerates fibrotic processes via pro-inflammatory and

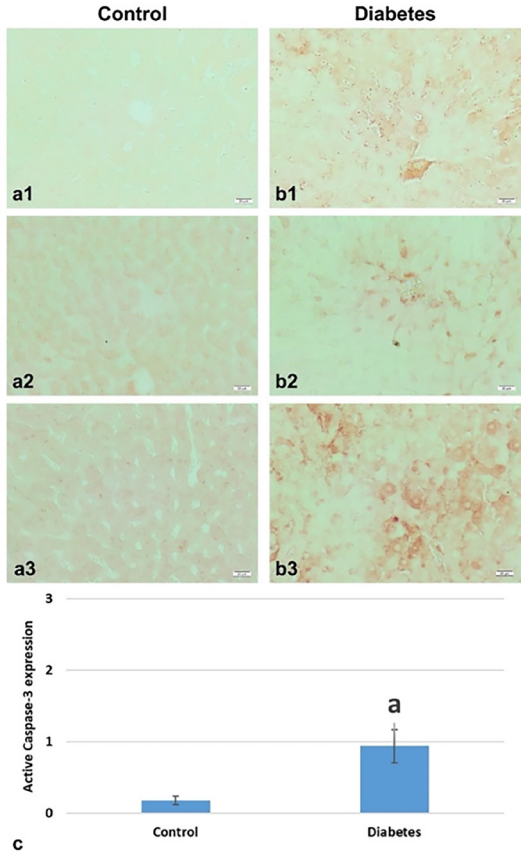

Figure 3. Representative figures for the immunohistochemical staining of active caspase-3 for control **(a)** and diabetes **(b)** groups. Graphical expression of semi-quantitative evaluation of active caspase-3 expression **(c)**. aP<0.005

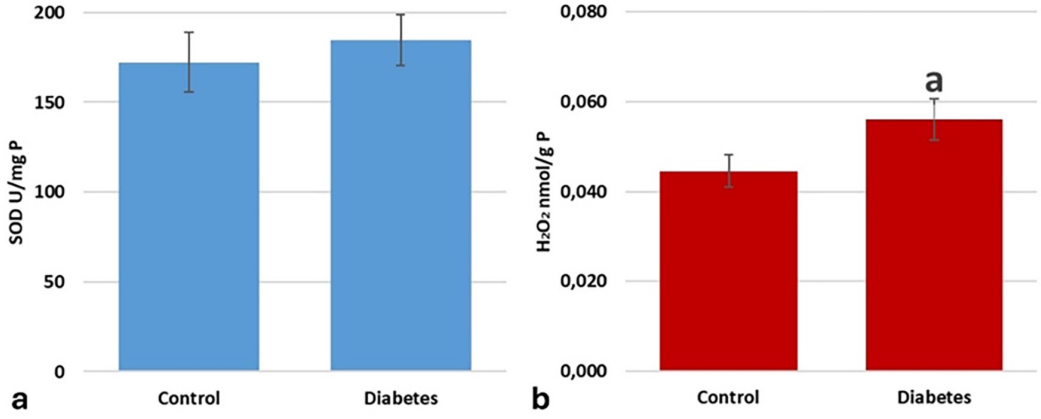

Figure 4. **(a)** Graphical representation for evaluation of superoxide dismutase activity (SOD) and **(b)** concentration of hydrogen peroxide (H_2O_2). aP<0.05

fibrogenic pathways (8,13). The presence of fibrosis in diabetic rats emphasizes the progression of liver injury, with potential implications for the occurrence of conditions such as NAFLD and cirrhosis (14). At this point, investigation of biochemical parameters including the levels of plasma alanine transaminase (ALT) and aspartate transaminase (AST) would have been complementary, where it was noted that ALT level was higher than that of AST in most of the NAFLD patients, while it was reversed in patients with alcoholic fatty livers (15). Additionally, the presence of cystic dilatations of ductules in certain diabetic liver samples, characterized by biliary-like epithelium and surrounding fibrosis, represents a novel finding that suggests early involvement of bile ducts in diabetic liver damage, warranting additional investigation.

Immunohistochemical analysis revealed increased active caspase-3 expression in diabetic liver samples, pointing to enhanced hepatocyte apoptosis. This is in line with previous research showing that oxidative stress induced by hyperglycemia triggers mitochondrial dysfunction, leading to apoptosis (6,16). In diabetic liver damage, it was shown that apoptosis has a significant role in the loss of functional hepatocytes and further impairment of liver function (17). The increased apoptotic activity noted in our study emphasizes the importance of targeting apoptosis in therapeutic strategies to alleviate liver damage in diabetic patients.

In terms of oxidative stress markers, the marked increase in the level of H_2O_2 in the diabetic group, along with unchanged SOD activity, suggests a persistent oxidative imbalance in diabetic liver tissue. As SOD is a crucial antioxidant enzyme that is responsible for converting superoxide radicals into less harmful species, however, stable SOD activity may indicate that antioxidant defences are insufficient to counteract the excessive production of ROS, including H_2O_2. This imbalance could explain the ongoing oxidative damage observed in liver tissue, resulting in apoptosis and fibrosis, as seen in this study. Previous studies have similarly reported increased oxidative stress markers in diabetic liver tissue, further supporting the idea that oxidative damage plays a central role in the pathogenesis of diabetic liver injury, including the promotion of collagen deposition and extracellular matrix remodelling (4,5).

Overall, our findings provide valuable insights into the mechanisms of liver damage in diabetes. The observed histopathological changes, fibrosis, increased apoptosis, and oxidative stress in diabetic rats underscore the complex nature of liver injury in diabetes. These results highlight the importance of targeting oxidative stress and apoptosis in therapeutic interventions to prevent or mitigate liver damage in diabetic patients. Future research should investigate the molecular pathways linking hyperglycemia, oxidative stress, and liver fibrosis, with the goal of developing targeted therapies to protect the liver from the detrimental effects of diabetes.

Study Limitations

A limitation of this study is the use of a single model of STZ-induced diabetes, which primarily mimics T1D. Future studies should include models of type 2 diabetes to assess whether similar hepatic alterations occur in the context of insulin resistance and metabolic syndrome, which are more common in the general population. Additionally, while we measured oxidative stress markers, a broader assessment of other ROS and antioxidant enzymes could provide a clearer picture of the oxidative balance in diabetic liver tissue. Furthermore, evaluation of certain liver injury biochemical markers like AST and ALT would provide a more comprehensive understanding of the pathophysiology of diabetes-associated hepatopathy.

CONCLUSION

In conclusion, this study demonstrated that diabetes leads to notable histopathological changes, increases apoptosis, and elevates oxidative stress in liver tissue, all of which contribute to fibrosis and hepatic dysfunction. These findings highlight the liver's vulnerability to diabetic injury and underscore the need for further research into therapeutic strategies to counteract oxidative stress and prevent diabetes-related liver complications.

ETHICS

Ethics Committee Approval: Animal experiments were carried out after obtaining the necessary ethical approval from the Bezmialem Vakıf University Animal Experiments Local Ethics Committee (approval no: 1294-1, date: 21.10.2024).

Informed Consent: Animal experimentation

FOOTNOTES

Authorship Contributions

Surgical and Medical Practices: S.Ö., Concept: S.Ö., Design: S.Ö., Data Collection or Processing: S.Ö., Analysis or Interpretation: S.Ö., M.K., Literature Search: S.Ö., Writing: S.Ö., M.K.

Conflict of Interest: No conflict of interest was declared by the authors.

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

REFERENCES

1. International Diabetes Federation. IDF Diabetes Atlas. 10th ed. Brussels: International Diabetes Federation; 2021.
2. Shi GJ, Shi GR, Zhou JY, Zhang WJ, Gao CY, Jiang YP, et al. Involvement of growth factors in diabetes mellitus and its complications: a general review. Biomed Pharmacother. 2018;101:510-27.
3. Ighodaro OM. Molecular pathways associated with oxidative stress in diabetes mellitus. Biomed Pharmacother. 2018;108:656-62.
4. Evans JL, Goldfine ID, Maddux BA, Grodsky GM. Oxidative stress and stress-activated signaling pathways: a unifying hypothesis of type 2 diabetes. Endocr Rev. 2002;23:599-622.
5. Jha JC, Banal C, Chow BS, Cooper ME, Jandeleit-Dahm K. Diabetes and kidney disease: role of oxidative stress. Antioxid Redox Signal. 2016;25:657-84.
6. Yuan T, Yang T, Chen H, Fu D, Hu Y, Wang J, et al. New insights into oxidative stress and inflammation during diabetes mellitus-accelerated atherosclerosis. Redox Biol. 2019;20:247-60.
7. Jiang S, Young JL, Wang K, Qian Y, Cai L. Diabetic-induced alterations in hepatic glucose and lipid metabolism: the role of type 1 and type 2 diabetes mellitus (Review). Mol Med Rep. 2020;22:603-11.
8. Mohamed J, Nazratun Nafizah AH, Zariyantey AH, Budin SB. Mechanisms of diabetes-induced liver damage: the role of oxidative stress and inflammation. Sultan Qaboos Univ Med J. 2016;16:e132-41.
9. Ghasemi A, Jeddi S. Streptozotocin as a tool for induction of rat models of diabetes: a practical guide. EXCLI J. 2023;22:274-94.
10. Mohamed DA, Mohamed NM, Abdelrahaman S. Histological and biochemical changes in adult male rat liver after spinal cord injury with evaluation of the role of granulocyte-colony stimulating factor. Ultrastruct Pathol. 2020;44:395-411.
11. Mousa AA, El-Gansh HAI, Eldaim MAA, Mohamed MAE, Morsi AH, El Sabagh HS. Protective effect of Moringa oleifera leaves ethanolic extract against thioacetamide-induced hepatotoxicity in rats via modulation of cellular antioxidant, apoptotic and inflammatory markers. Environ Sci Pollut Res Int. 2019;26:32488-504.
12. Faddladdeen KA, Ojaimi AA. Protective Effect of Pomegranate (Punica granatum) Extract against diabetic changes in adult male rat liver: histological study. J Microsc Ultrastruct. 2019;7:165-70.
13. Tuleta I, Frangogiannis NG. Diabetic fibrosis. Biochim Biophys Acta Mol Basis Dis. 2021;1867:166044.
14. Han SK, Baik SK, Kim MY. Non-alcoholic fatty liver disease: definition and subtypes. Clin Mol Hepatol. 2023;29(suppl):S5-S16.
15. Rikhi R, Singh T, Modaresi Esfeh J. Work up of fatty liver by primary care physicians, review. Ann Med Surg (Lond). 2020;50:41-8.
16. Zhang Z, Huang Q, Zhao D, Lian F, Li X, Qi W. The impact of oxidative stress-induced mitochondrial dysfunction on diabetic microvascular complications. Front Endocrinol (Lausanne). 2023;14:1112363.
17. Yuniartha R, Arfian N, Setyaningsih WAW, Kencana SMS, Sari DCR. Accelerated senescence and apoptosis in the rat liver during the progression of diabetic complications. Malays J Med Sci. 2022;29:46-59.

Medical Journal of Bakırköy

Med J Bakirkoy 2025;21(3):274-287
DOI: 10.4274/BMJ.galenos.2025.2024.10-8

Research

Predicting ICU Stay Duration for Hemodialysis Patients with Poisoning: A Study Comparing Deep Learning with Machine Learning Models

Zehirlenme Nedeniyle Hemodiyaliz Hastalarında Yoğun Bakımda Kalış Süresinin Tahmini: Makine Öğrenimi Modellerini Derin Öğrenme Modelleriyle Karşılaştıran Bir Çalışma

Khadijeh Moulaei[1], Shahin Shadnia[2], Mohammad Reza Afrash[3], Babak Mostafazadeh[2], Heliya Rafsanjani[4], Peyman Erfan Talab Evini[2], Mohanna Sharifi[5], Mohadeseh Sarbaz Bardsiri[2,6], Hadi Kazemi-Arpanahi[7], Babak Sabet[8], Mohammad Parvin[9], Vahideh Montazeri[3], Amir Hossein Daeechini[10], Mitra Rahimi[2]

[1]Iran University of Medical Sciences, Health Management and Economics Research Center, Health Management Research Institute, Tehran, Iran

[2]Shahid Beheshti University of Medical Sciences, Loghman Hakim Hospital, Toxicological Research Center, Excellence Center of Clinical Toxicology, Department of Clinical Toxicology, Tehran, Iran

[3]Smart University of Medical Sciences, Artificial Intelligence in Medical Sciences Research Center, Tehran, Iran

[4]Network of Interdisciplinarity in Neonates and Infants (NINI), Universal Scientific Education and Research Network (USERN), Tehran, Iran

[5]Tor Vergata University, Department of Pharmacy, Rome, Italy

[6]Iran University of Medical Sciences, Firouzgar Hospital, Department of Clinical Toxicology, Tehran, Iran

[7]Abadan University of Medical Sciences, Department of Health Information Technology, Abadan, Iran

[8]Shahid Beheshti University of Medical Sciences Faculty of Medicine, Department of Surgery, Tehran, Iran

[9]Auburn University, Department of Industrial and Systems Engineering, Auburn, AL, USA

[10]Shahid Beheshti University of Medical Sciences, School of Allied Medical Sciences, Student Research Committee, Department of Health Information Technology and Management, Tehran, Iran

ABSTRACT

Objective: This study compares the effectiveness of machine learning and deep learning models in predicting hemodialysis patients' length of stay in the intensive care unit (ICU).

Methods: This retrospective cohort study used data from 980 poisoned patients undergoing hemodialysis. A variety of eight well-known machine learning [support vector machine, extreme gradient boosting, random forest (RF), decision tree] and deep learning (deep neural network, feedforward neural network, long short-term memory, convolutional neural network) models were employed.

Results: Feature importance analyses using Shapley Additive exPlanation and local interpretable model-agnostic explanation methodologies identified Glasgow Coma Scale (GCS <8), intubation, acute kidney injury, PO_2, blood urea nitrogen, metabolic acidosis, and number of hemodialysis sessions as key predictors of ICU stay duration in poisoned hemodialysis patients, with intubation score, GCS score, and ICU admission type being the most significant predictors. Overall, the RF model displayed exceptional performance across various metrics.

Conclusion: Our findings emphasize the importance of neurological status, respiratory function, and renal injury in predicting ICU duration, offering valuable insights for clinical decision-making and resource allocation in this high-risk population.

Keywords: Machine learning, deep learning, length of stay, intensive care, poisoning

Address for Correspondence: Mitra Rahimi MD, Shahid Beheshti University of Medical Sciences, School of Allied Medical Sciences, Student Research Committee, Department of Health Information Technology and Management, Tehran, Iran

E-mail: mrahimi744@gmail.com **ORCID ID:** orcid.org/0000-0002-8785-8207

Cite as: Moulaei K, Shadnia S, Afrash MR, Mostafazadeh B, Rafsanjani H, Evini PET, et al. Predicting ICU stay duration for hemodialysis patients with poisoning: a study comparing deep learning with machine learning models. Med J Bakirkoy. 2025;21(3):274-287

Received: 11.11.2024
Accepted: 24.03.2025
Publication Date: 03.09.2025

ÖZ

Amaç: Bu çalışma, hemodiyaliz hastalarının yoğun bakım ünitesinde (YBÜ) kalış sürelerini tahmin etmede makine öğrenimi ve derin öğrenme modellerinin etkinliğini karşılaştırır.

Gereç ve Yöntem: Bu retrospektif kohort çalışmasında, hemodiyalize giren 980 zehirlenmiş hastadan alınan veriler kullanılmıştır. Çeşitli sekiz iyi bilinen makine öğrenmesi (destek vektör makinesi, aşırı gradyan artırma, rastgele orman, karar ağacı) ve derin öğrenme (derin sinir ağı, ileri beslemeli sinir ağı, uzun kısa süreli bellek, evrişimli sinir ağı) modeli kullanılmıştır.

Bulgular: Shapley Katkı Açıklaması ve yerel yorumlanabilir modelden bağımsız açıklama metodolojileri kullanılarak yapılan özellik önem analizleri, zehirlenmiş hemodiyaliz hastalarında YBÜ'de kalış süresinin temel belirleyicileri olarak Glasgow Koma Skalası (GCS <8), entübasyon, akut böbrek hasarı, PO_2, kan üre azotu, metabolik asidoz ve hemodiyaliz seansı sayısını belirlemiştir. En önemlileri entübasyon puanı, GCS puanı ve YBÜ'ye kabul türü olmuştur. Genel olarak, rastgele orman modeli çeşitli ölçütlerde olağanüstü bir performans göstermiştir.

Sonuç: Bulgularımız, YBÜ'de kalış süresini tahmin etmede nörolojik durum, solunum fonksiyonu ve böbrek hasarının önemini vurgulamakta ve bu yüksek riskli popülasyonda klinik karar alma ve kaynak dağıtımı için değerli bilgiler sunmaktadır.

Anahtar Kelimeler: Makine öğrenimi, derin öğrenme, kalış süresi, yoğun bakım, zehirlenme

INTRODUCTION

Poisoning, whether intentional or accidental, is a significant health issue worldwide, imposing substantial financial, physical, and mental burdens on patients, families, and society (1,2). According to the World Health Organization (WHO), over 3 million people are poisoned annually, resulting in 220,000 deaths, mostly in developing countries due to easy access to toxic substances, lack of awareness, and limited hospital resources (3,4). Poisoning accounts for over 2.4% of emergency department visits and 3-6% of intensive care unit (ICU) admissions (5,6). While most poisoned patients recover with supportive treatment, critically ill patients with severe symptoms are admitted to ICUs for intensive care and monitoring. Studies show that 4.1-30.8% of life-threatening poisoning cases require ICU admission. In developing countries, including Iran, poisoning incidents, especially from pesticides and rice tablets, have doubled in recent decades despite improved ICU facilities (7,8).

Hemodialysis patients are frequently admitted to the ICU due to poisoning, with about 2% of chronic dialysis patients requiring ICU care annually. They are at increased risk of infections due to factors like immune deficiency from uremia, impaired phagocytic function, older age, and comorbidities such as diabetes. Frequent use of vascular access for hemodialysis heightens the risk of bloodstream infections (9). Managing poisoned hemodialysis patients in the ICU is challenging due to their altered pharmacokinetics and pharmacodynamics, which affect toxin clearance. Renal failure complicates toxin elimination, often necessitating hemodialysis. Additionally, these patients are prone to fluid and electrolyte imbalances, requiring close monitoring. Infections, especially bloodstream infections from vascular access, are a major concern for immunocompromised patients (10). Comprehensive ICU management for these patients involves toxin removal, careful renal replacement therapy, fluid balance, electrolyte management, and infection control.

Moreover, healthcare systems face constant pressure to improve patient outcomes and reduce costs. ICUs, which provide critical care, are expensive and resource-intensive (11,12). The increasing number of poisoned patients has heightened the demand for ICU beds. The WHO highlights the importance of monitoring the length of stay (LOS) as a measure of care quality and resource use (13,14). Patients with prolonged LOS consume a significant portion of resources, so reducing LOS can enhance bed turnover, optimize resource allocation, improve patient safety, and lower costs. Identifying patients with long LOS, particularly in overwhelmed hospitals, can alleviate pressure and boost ICU efficiency. Policymakers are adopting evidence-based solutions to optimize ICU resources like beds, staff, and mechanical ventilation (15-17).

Recently, artificial intelligence-based solutions such as machine learning (ML) and deep learning (DL) have gained further attention for their ability to predict the outcome of interest based on a great amount of data, especially when the relationships between variables are complex and non-linear (18-20). DL and ML models can be used for predicting the patients' ICU LOS (21,22). These models leverage large datasets comprising patient demographics, clinical variables, and possibly real-time monitoring data such as vital signs and laboratory results (21). By analyzing patterns and correlations within these data, DL and ML algorithms can generate accurate predictions of how long a patient is likely to stay in the ICU (22,23). This predictive capability not only helps in optimizing resource allocation and bed management but also assists healthcare providers in identifying patients at risk of prolonged ICU stays early on, enabling proactive interventions and potentially improving patient outcomes (21,24).

To the best of our knowledge, no study has compared ML models with DL models for predicting ICU stay duration for hemodialysis patients with poisoning. Therefore, the purpose of this study is to develop and evaluate the effectiveness of both ML and DL models in predicting the ICU LOS for hemodialysis patients suffering from poisoning. By comparing the performance of these models, we aim to identify the most accurate and reliable approach, ultimately improving resource allocation and patient outcomes in critical care settings.

METHODS

Study Design and Setting

This retrospective cohort study was performed among all poisoned patients who were admitted to the ICU at the Loghman Hakim Hospital (LHH) between January 1, 2016, and December 31, 2020. The hospital, which is known as the Iran's largest poison center, is crucial for managing patients who need specialized care as a result of various poisonings and toxic exposures. The ICU of LHH is essential to the management of critical situations, providing cutting-edge medical techniques and specialized care to guarantee the best outcomes for patients. In this study, the LOS of patients who were admitted to the ICU was analyzed. Patients were categorized into two groups based on the LOS: short (lasting 4 days or less), and long (exceeding 4 days). We utilized several data-driven ML and DL models to develop an accurate prediction model the ICU LOS in poisoned patients undergoing hemodialysis. The key steps taken were as follows (Figure 1):

- establishing the study roadmap and experiment environment,
- preprocessing the data,
- using feature selection algorithms,
- selecting appropriate classification algorithms and their hyperparameters,
- splitting the data into training and testing sets, and
- evaluating model performance.

Data Collection and Preprocessing

The data set was collected by reviewing the electronic medical records (EMRs) of patients undergoing hemodialysis who were poisoned admitted to the ICU at LHH between January 2016 and December 2020. Relevant variables were extracted from poisoned patient records and entered into a database. Age, sex, kind of poisoning, history of underlying conditions, medication usage and habits, laboratory test results, and vital signs, number of extracorporeal technique uses, type of extracorporeal method, and patient outcome were among the variables under investigation (Table 1).

Statistical Analysis

In this study, before further analysis and feeding data into ML methods, the rows of datasets collected from EMRs laboratory tests, underwent a series of preprocessing steps.

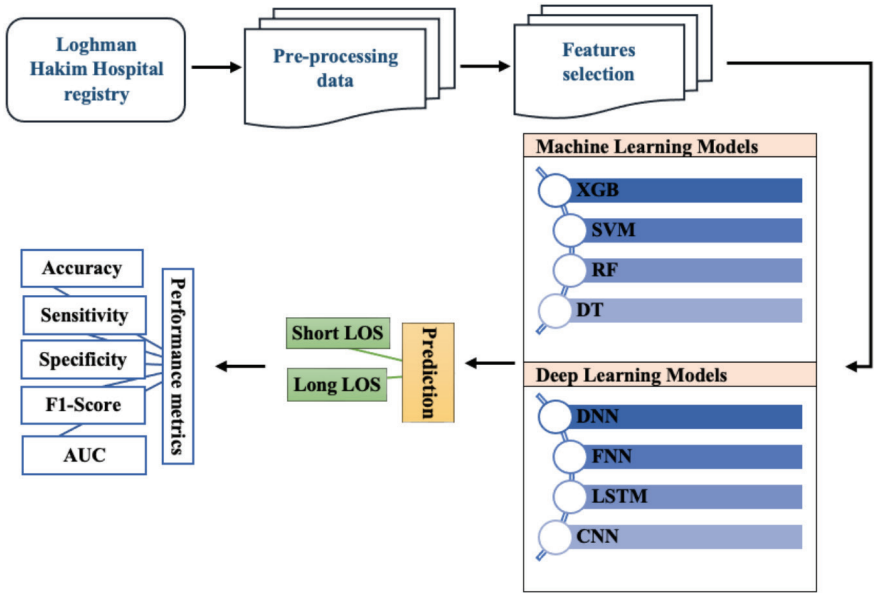

Figure 1. The road map of the proposed system for the prognosis of poisoned patients who are candidates for dialysis
AUC: Area under the curve, LOS: Length of stay, XGB: Extreme gradient boosting, SVM: Support vector machine, RF: Random forest, DT: Decision tree, DNN: Deep neural network, FNN: Feedforward neural network, CNN: Convolutional neural network

Table 1. The descriptive statistics of variables

Demographic data			
Variables		**Frequency (%)**	**Values**
Gender	Female	187 (19.1%)	Female/male
	Male	793 (80.9%)	
Age (year)	17-20	100 (10/4%)	Numerical
	21-40	604 (61.6%)	
	41-60	211 (21.5%)	
	>61	65 (6.6%)	
Co-ingestion		62 (6.3%)	Yes/no
Smoking		46 (4.7%)	Yes/no
Alcohol consumption		627 (64%)	Yes/no
Opium abuse		75 (7.7%)	Yes/no
Stimulants abuse		13 (1.3%)	Yes/no
History of the previous disease		150 (15.3%)	Yes/no
History of taking medication		77 (7.9%)	Yes/no
Hemoperfusion		9 (0/9%)	Yes/no
Intubation		203 (20.7%)	Yes/no
Antidote therapy		904 (92.2%)	Yes/no
Duration of hospitalization (day) median (min-max)		8 (1-116)	Numerical
Laboratory and clinical data			
13≤GCS<15		659 (67.2%)	Numerical
8≤GCS<13		62 (6.3%)	Numerical
Coma (GCS <8)		108 (11%)	Yes/no
Bradypnea		46 (4.7%)	Numerical
Temperature (mean±SD)		36.9±0.56	Numerical
Bradycardia		14 (1.4%)	Numerical
Tachycardia		145 (14.8%)	Numerical
Hypotension		38 (3.9%)	Numerical
Hypertension		293 (29.9%)	Numerical
Metabolic acidosis		823 (84%)	Numerical
Acute kidney injury (AKI)		536 (54.7%)	Numerical
Renal disease		641 (65%)	Yes/no
Serum HCO_3 (mean±SD)		14.4±23.8	Numerical
BUN (meq/L) (mean±SD)		36.7±27.2	Numerical
Creatinine (meq/L) (mean±SD)		1.7±4.1	Numerical
Blood glucose (mg/dL) (mean±SD)		132.5±66.6	Numerical
Blood pH (mean±SD)		7.2±0.4	Numerical
Sodium (meq/L) (mean±SD)		137.8±10.8	Numerical
Potassium (meq/L) (mean±SD)		4.7±3.4	Numerical
Partial pressure of carbon dioxide (PCO_2), partial pressure of oxygen (PO_2), bicarbonate (HCO_3), base excess (BE), blood sugar (Bs), white blood cell count (WBC), hemoglobin (Hb), hematocrit (Hct), platelet count (PLT), aspartate aminotransferase (AST), alanine aminotransferase (ALT), alkaline phosphatase (ALP), lactate dehydrogenase (LDH), creatine phosphokinase (CPK), creatine phosphokinase-MB (CPK-MB), prothrombin time (PT), partial thromboplastin time (PTT), international normalized ratio (INR).			Numerical
Outcome variable: ICU LOS	Short LOS ≤4 days	640	Yes/no
	Long LOS >4 days	340	Yes/no

ICU: Intensive care unit, LOS: Length of stay, SD: Standard deviation, GCS: Glasgow Coma Scale

These steps include; any rows with over 70% missing values were removed. Patient data were excluded if they were missing critical demographic or clinical information (e.g., comorbidities) necessary for analysis. Additionally, duplicate records and entries with inconsistent or implausible values (e.g., negative age or erroneous laboratory results) were excluded to ensure data quality. Missing values for remaining variables were handled using imputation techniques where feasible, and rows with irreparable missingness for key variables were removed. Next, minimum-maximum scaling was applied to normalize all values between 0 and 1. Standard scaler scaling was then used to standardize the data distribution. Data validation checked the integrity and accuracy of the dataset. To handle class imbalance, under-sampling, balanced the classes by keeping all samples from the minority class.

After preprocessing, the final dataset contained 980 patients. This dataset was randomly split so that 70% of the data (686 patients) was assigned to the training set, and the remaining 30% (294 patients) was assigned to the test set. The training set was used to develop the feature selection and ML models, while the test set was held out for model evaluation. The descriptive statistics of the variables in the dataset are shown in Table 1. This includes variable names, the frequency of each variables, and their values.

Model Development and Evaluations

The LOS in the ICU for hemodialysis patients with poisoning was predicted using eight well-known models from the domains of DL and ML. Convolutional neural networks (CNN), feedforward neural networks, long short-term memory (LSTM), and deep neural networks (DNN) were among the DL models. The ML models included random forest (RF), decision tree (DT), support vector machine, and extreme gradient boosting (XGB).

Cross-validation and Tweaking of Hyperparameters

We trained all suggested models using 10-fold cross-validation to reduce overfitting. The dataset is divided into ten equal segments using this method. The model is trained on nine segments and validated on the remaining segments in each iteration. To ensure complete validation, this procedure is carried out ten times. A robust assessment of the model's overall performance is provided by the final performance metric, which is produced by averaging the outcomes from each iteration (25).

In addition to 10-fold cross-validation, we implemented other regularization techniques such as L2 regularization and dropout to further reduce the risk of overfitting. L2 regularization adds a penalty to the loss function based on the magnitude of the model's coefficients, discouraging overly complex models. Furthermore, we monitored training and validation loss during the training process to detect signs of overfitting early. Early stopping was employed to halt training if the validation loss did not improve for a pre-defined number of epochs, thus preventing the model from overfitting to the training data.

We also performed hyperparameter tuning to improve the performance of each method. Using a grid search approach, we methodically assessed a large number of hyperparameter values. Finding the parameter configurations that maximize each model's accuracy and efficiency was the goal. We successfully adjusted the models because of this exhaustive and repetitive analysis of the hyperparameter space. Our models' capacity to assess and forecast outcomes using the provided dataset is greatly improved by this meticulous calibration (26).

Justification and Explanation of the Machine Learning and Deep Learning Models' Output

Because of their intricate and opaque internal workings, ML and DL techniques are commonly referred to as "black box" models (27,28). This intricacy frequently leads to a lack of interpretability, which can be especially troublesome in crucial domains, like healthcare, where comprehending the reasoning behind forecasts is essential. Researchers have been creating methods to improve the interpretability of these models in order to address this problem. Shapley Additive exPlanations (SHAP), first presented by Lundberg and Lee (29), is a well-known technique that has gained popularity recently. By utilizing the idea of Shapley values from cooperative game theory, SHAP seeks to clarify the predictions of ML models. Because of its ability to yield insightful information on model predictions, this method has become widely accepted and used in a variety of fields, including clinical research (30,31).

In our study, we incorporated SHAP to interpret the outputs of our ML and DL models. By applying SHAP, we were able to break down the predictions into contributions from each feature, offering a clear and detailed understanding of how each variable influenced the model's decisions. This transparency is particularly valuable in healthcare applications, as it allows clinicians to trust and verify the predictions made by the models. Furthermore, we complemented SHAP with other interpretability techniques, such as local interpretable model-agnostic explanations (LIME) and feature importance analysis, to provide a multifaceted view of model behavior. These methods together helped us ensure that our models were not only accurate but also interpretable and trustworthy.

By integrating these interpretability methods, we aimed to bridge the gap between model complexity and usability, ensuring that our ML and DL models can be effectively deployed in real-world healthcare settings, where understanding and trust are paramount. This comprehensive approach to interpretability underscores our commitment to developing reliable and transparent predictive models that can aid in critical decision-making processes.

Model Performance Evaluation

Performance measures obtained from the confusion matrix were used to thoroughly assess the efficacy of both ML and DL models, as shown below. Key performance indicators such as accuracy, specificity, sensitivity, F1-score, and the receiver operating characteristic (ROC) curve were used in a thorough evaluation of the predictive models.

$$1)\ \text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} * 100$$

$$2)\ \text{Sensitivity} = \frac{Tp}{TP+FN} * 100$$

$$3)\ \text{Sensitivity} = \frac{TN}{TN+FP} * 100$$

$$5)\ \text{f - measure} = 2\frac{\text{precision} * \text{sensitivity}}{\text{precision} + \text{sensitivity}}$$

Ethical Considerations

For this study, ethical approval was obtained from the ethics committee at Shahid Beheshti University of Medical Sciences (approval no: IR.SBMU.RETECH.REC.1401.767, date: 12.02.2023). Due to the study's non-invasive methodology and strict adherence to patient anonymity and data confidentiality, the ethics committee granted a waiver for written informed consent. This waiver ensured the collection of confidential data without any identifying information. Access to the data was limited to the research team, thereby mitigating any potential risk to patients in accordance with the study's protocols.

RESULTS

Patients' Characteristics

This retrospective observational study examined data from 68,181 patients hospitalized for poisoning over a specified time. After applying exclusion criteria, 980 (1.4%) underwent hemodialysis. The cohort included 793 (80.9%) males and 187 (19.1%) females with a mean age of 36.5±14 years. The age distribution significantly differed, with most patients (604, 61.6%) aged 21-40 years, ($p<0.001$). Intentional poisoning accounted for 117 (11.9%) cases. Methanol was the most common poisoning agent (858, 87.6%), followed by multidrug ingestions. The majority (830, 84.7%) had no prior history of kidney disease, and 903 (92.1%) had no prior drug use. However, 627 (64%) admitted alcohol use. Hemodialysis was the most widely used extracorporeal method (980 cases, 99.1%). Hemoperfusion was additionally used to treat 9 patients poisoned by methanol, multidrug ingestions, or methadone. This study characterized the demographics, toxins, and extracorporeal treatment approaches for a large cohort of poisoned patients requiring hemodialysis. Figure 2 shows the cause of intoxication in the studied patients.

Hyperparameters Tuning

Table 2 presents the tuned hyperparameters of four ML algorithms.

Performance Evaluation of Selected Models

Deep Learning Models

Table 3 presents the performance assessment of selected models. According to the findings in this table, the best DL model identified is the CNN with a sensitivity of 82%, specificity of 83%, accuracy of 95%, and F1-score of 82%. The best ROC value was associated with the DNN model.

Machine Learning Models

Overall, the performance of the RF model was superior to all other models, with a sensitivity of 92%, an accuracy of 98%, an F1-score of 98%, and an ROC score of 98%. However, the XGB model also achieved a sensitivity of 92%, an accuracy of 98%, and an ROC score of 98%. Additionally, the DT model achieved a sensitivity of 92% and an accuracy of 98%.

Deep Learning vs. Machine Learning Models

In terms of sensitivity, DL models performed from 80.0% to 82.0%, whereas ML models achieved sensitivities between

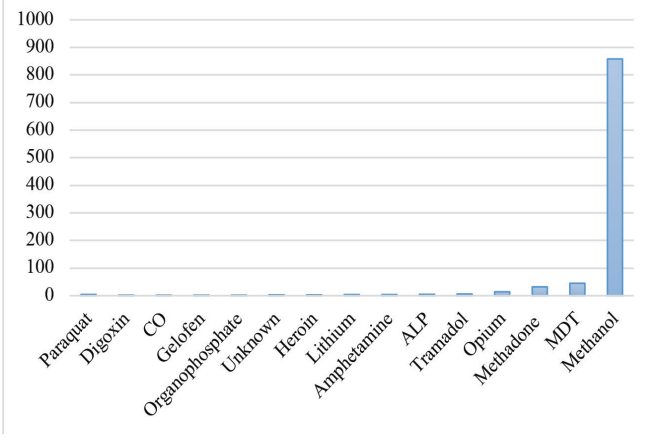

Figure 2. Cause of intoxication in the studied patients
MDT: Multiple drug toxicity

80.0% and 92.0%. The specificity of DL models ranged from 78.0% to 83.0%, whereas ML models achieved specificities between 84.0% and 94.0%. DL models had an accuracy ranging from 94.0% to 95.0%, compared to ML models, which ranged from 95.0% to 98.0%. The F1-scores for DL models were between 79.0% and 82.0%, while ML models ranged from 82.0% to 95.0%. Additionally, DL models achieved ROC scores between 97.0% and 98.0%, whereas ML models had ROC scores ranging from 95.0% to 98.0%.

Overall, while the RF model demonstrated exceptional performance across various metrics, other ML models showed superior performance compared to DL models across most measures.

Figure 3 shows the performance of various ML and DL models in predicting ICU stay duration for hemodialysis patients with poisoning, while Figure 4 compares the ROC curves of these models.

Table 2. Hyperparameters selected for ML and DL models

ML and DP models	Hyper parameters
CNN	Conv1D: filters: 64, kernel size: 3, activation: 'relu', MaxPooling1D: pool size: 2 dense (layer 1): units: 64 activation: 'relu', dense (output layer): units: 1, activation: 'sigmoid', optimizer: 'adam', loss function: 'binary_crossentropy', epochs: 10, batch size: 32
DT	Criterion: 'gini', max depth: none, min samples split: 2, min samples leaf: 1, max features: none, splitter: 'best', random state: 42
DNN	Dense (layer 1): units: 64, activation: 'relu', input shape: (X.shape[1],) dense (layer) units: 32, activation: 'relu', dense (output layer): units: 1, activation: 'sigmoid' optimizer: 'adam', loss function: 'binary_crossentropy', epochs: 10, batch size: 32
FNN	Hidden layer sizes: (100,) max iterations: 1000, random state: 42, n_splits: 10, shuffle: true, random_state: 42
LSTM	Units: 100, input_shape: (X_train.shape[1], X_train.shape[2]) optimizer: 'adam' loss: 'binary_crossentropy' metrics: ['accuracy'] Training hyperparameters epochs: 100, batch_size: 32, verbose: 0 StratifiedKFold parameters, n_splits: 10, huffle: true, random_state: 42
RF	n_estimators: 100, random_state: 42
SVM	Kernel: 'linear', robability: true, random_state: 42
XGB	Base_score: 0.5, booster: gbtree, colsample_bylevel: 1, colsample_bynode: 1 colsample_bytree: 1, gamma: 0, gpu_id: -1, importance_type: gain, learning_rate: 0.300000012, max_delta_step: 0, max_depth: 6, min_child_weight: 1, missing: nan, monotone_constraints: () n_estimators: 100, random_state: 42, validate_parameters: 1

XGB: Extreme gradient boosting, SVM: Support vector machine, RF: Random forest, DT: Decision tree, DNN: Deep neural network, FNN: Feedforward neural network, CNN: Convolutional neural network, LSTM: Long short-term memory, ML: Machine learning, DL: Deep learning

Table 3. Performance evaluation of selected models

Models	**Performance of each deep learning models**				
	Sensitivity (%)	**Specificity (%)**	**Accuracy (%)**	**F1-score (%)**	**ROC (%)**
CNN	**82.0**	**83.0**	**95.0**	**82.0**	**97.0**
LSTM	**82.0**	80.0	94.0	81.0	97.0
DNN	80.0	82.0	94.0	81.0	**98.0**
FNN	80.0	78.0	94.0	79.0	97.0
Models	**Performance of each machine learning models**				
	Sensitivity (%)	**Specificity (%)**	**Accuracy (%)**	**F1-score (%)**	**ROC (%)**
RF	**92.0**	92.0	**98.0**	**95.0**	**98.0**
DT	**92.0**	92.0	98.0	92.0	95.0
SVM	80.0	84.0	95.0	82.0	96.0
XGB	90.0	**94.0**	**98.0**	90.0	**98.0**

XGB: Extreme gradient boosting, SVM: Support vector machine, RF: Random forest, DT: Decision tree, DNN: Deep neural network, FNN: Feedforward neural network, CNN: Convolutional neural network, LSTM: Long short-term memory, ROC: Receiver operating characteristic

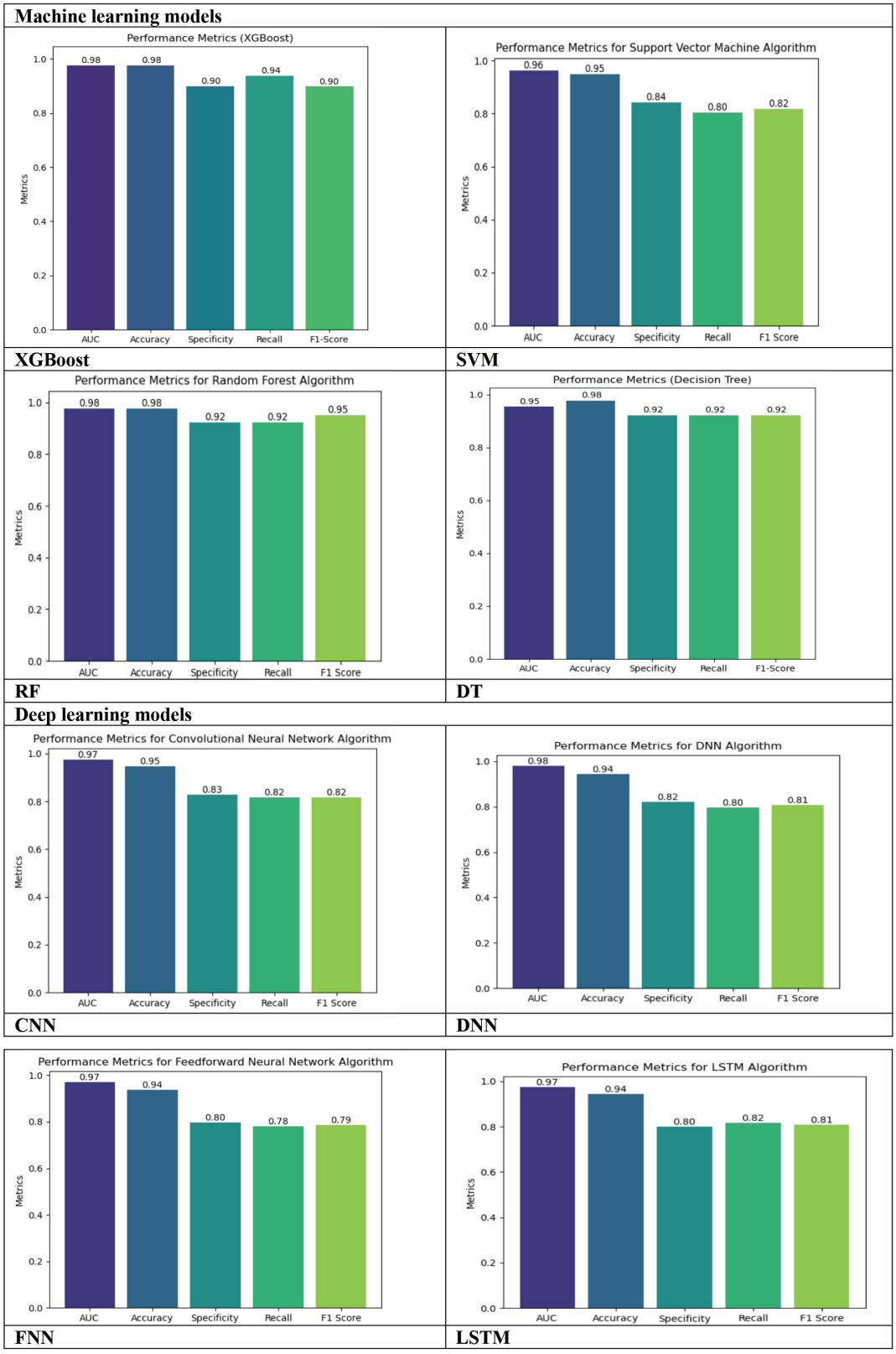

Figure 3. Performance evaluation of ML and DL models in predicting ICU stay duration for hemodialysis patients with poisoning

XGB: Extreme gradient boosting, SVM: Support vector machine, RF: Random forest, DT: Decision tree, DNN: Deep neural network, CNN: Convolutional neural network, FNN: Feedforward neural network, LSTM: Long short-term memory, ML: Machine learning, DL: Deep learning, ICU: Intensive care unit, AUC: Area under the curve

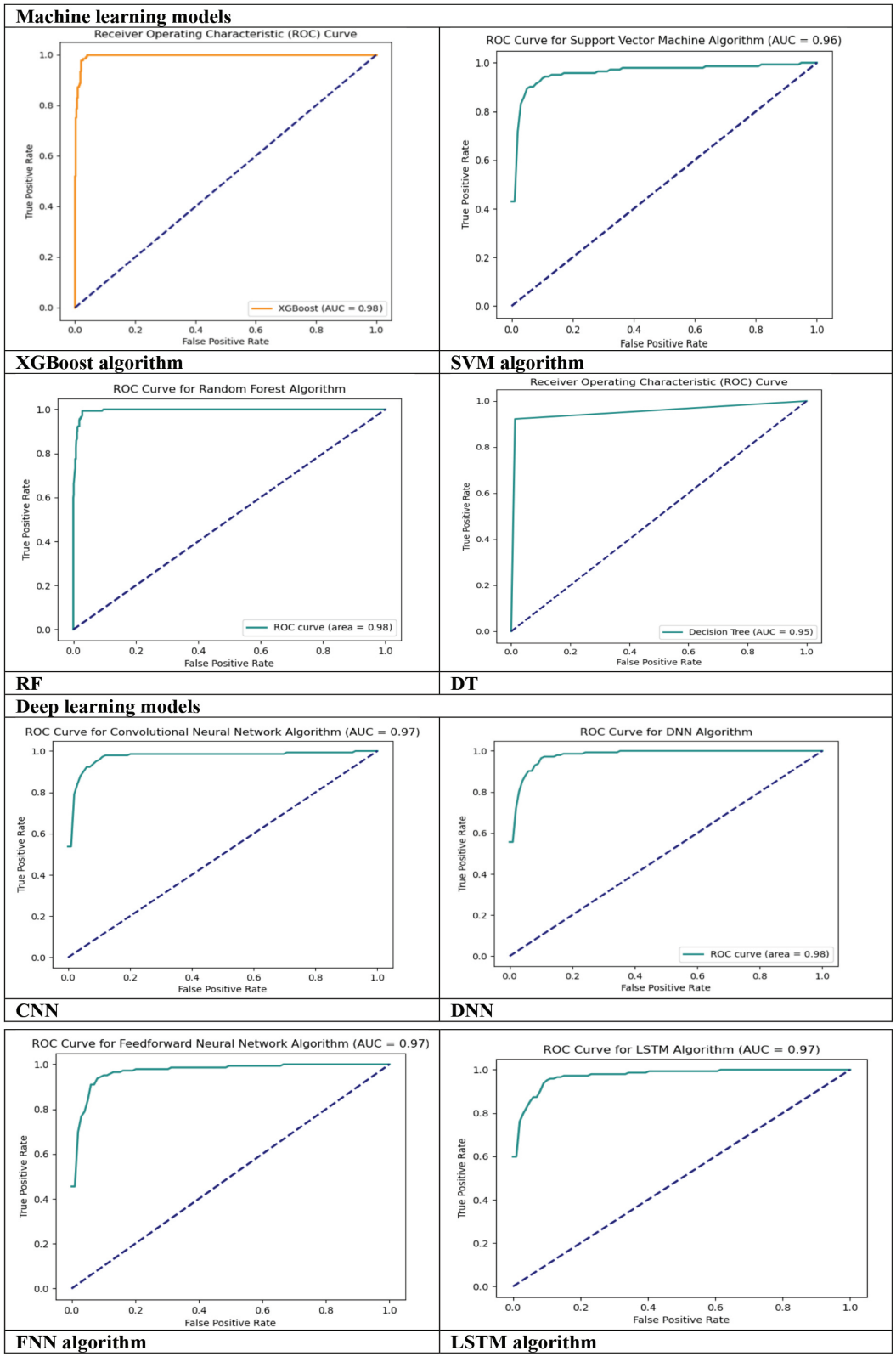

Figure 4. ROC curves comparing ML and DL models in predicting ICU stay duration for hemodialysis patients with poisoning
XGB: Extreme gradient boosting, SVM: Support vector machine, RF: Random forest, DT: Decision tree, DNN: Deep neural network, CNN: Convolutional neural network, FNN: Feedforward neural network, LSTM: Long short-term memory, ROC: Receiver operating characteristic, ML: Machine learning, DL: Deep learning, ICU: Intensive care unit

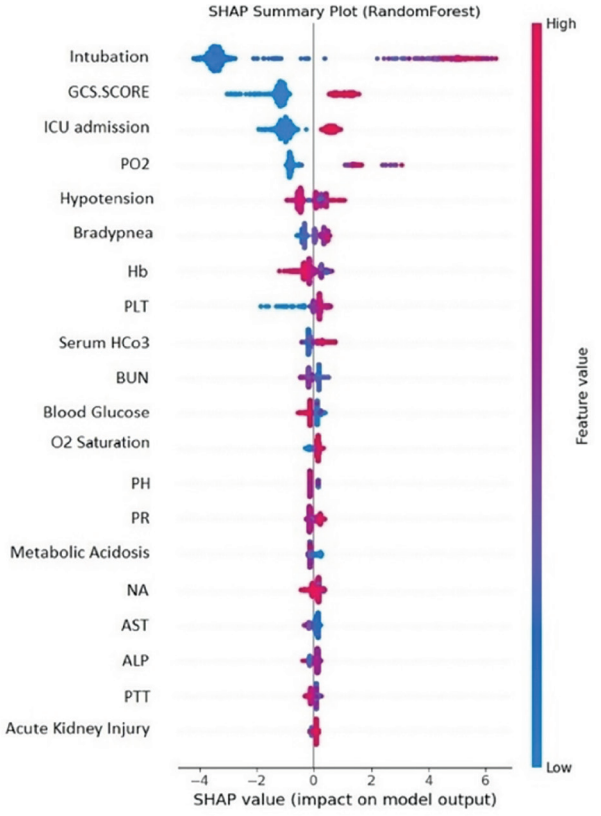

Figure 5. SHAP summary plot
SHAP: Shapley Additive exPlanations, GCS: Glasgow Coma Scale, ICU: Intensive care unit

Explanation and Justification the Output of Machine Learning and Deep Learning Models

Shapley Additive exPlanations (SHAP)

The SHAP summary plot shows the impact of each feature on the model's output, with higher values indicating a greater positive impact and lower values indicating a greater negative impact (Figure 5). For example, the feature "intubation" has a positive impact on the model's output, while the feature "blood glucose" has a negative impact. The most important features for predicting the duration of ICU stay among hemodialysis patients suffering from poisoning were intubation score, GCS score, and ICU admission type.

Local Interpretable Model-agnostic Explanations (LIME)

Figure 6 illustrates LIME plots depicting the cumulative influence of essential features on the model's prediction of ICU stay duration for hemodialysis patients with poisoning. The most influential features were GCS <8, intubation, acute kidney injury (AKI), PO_2, blood urea nitrogen (BUN), metabolic acidosis, and number of hemodialysis sessions.

DISCUSSION

In this study, we aimed to predict the ICU LOS for hemodialysis patients with poisoning by comparing the performance of various ML and DL models. The importance of timely prediction and intervention in poisoning cases cannot be overstated, as delayed treatment can lead

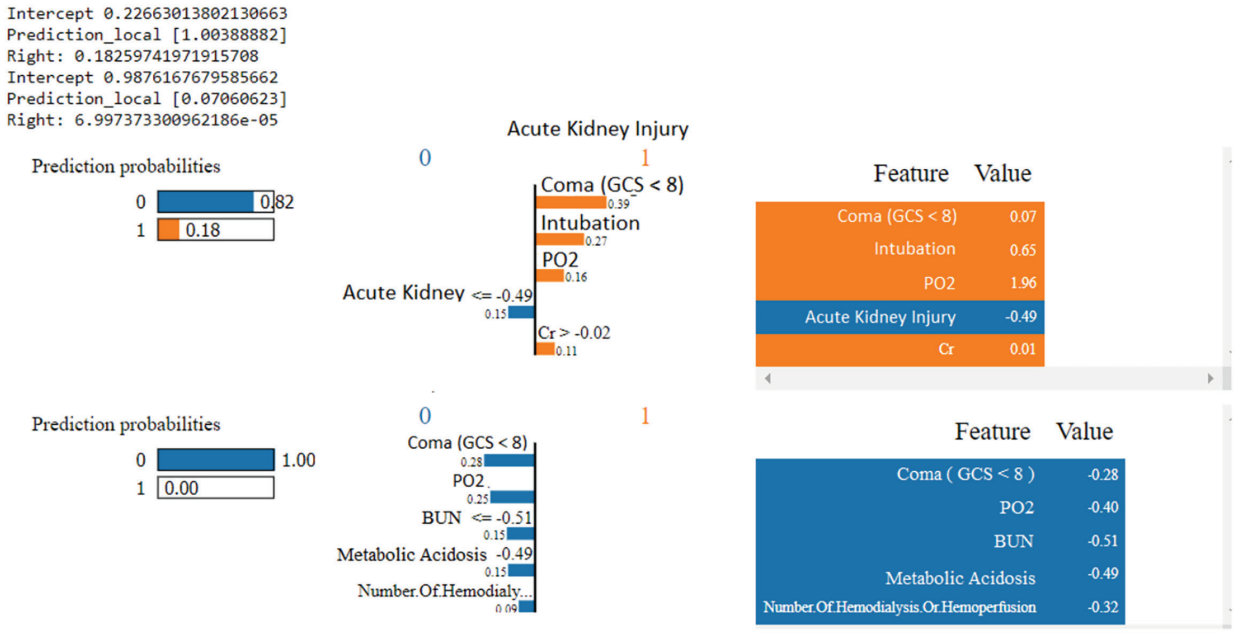

Figure 6. LIME plot
LIME: Local interpretable model-agnostic explanations, GCS: Glasgow Coma Scale

to severe complications or even death. Our findings demonstrated that the RF model outperformed all other models, achieving sensitivity, accuracy, F1-score, and an ROC score. In comparison, the best-performing DL model, the CNN, achieved a sensitivity, specificity, accuracy, and an F1-score. Overall, ML models exhibited higher sensitivities, specificities, accuracies, and F1-scores than DL models, indicating that, especially the RF model, they are more effective for predicting ICU stay duration for this patient population.

As mentioned above, our results showed that the RF model surpassed all ML and DL models, achieving superior sensitivity, accuracy, F1-score, and ROC score. In contrast, the best-performing DL model, the CNN, attained notable sensitivity, specificity, accuracy, and F1-score. The RF model and the CNN are pivotal in our study due to their exemplary performance and complementary strengths, which highlight the potential of both ML and DL in medical predictions. The RF model's superior sensitivity, accuracy, F1-score, and ROC score underscore its robustness and reliability in predicting ICU stay duration for hemodialysis patients with poisoning. This high performance can be attributed to the RF model's ability to handle diverse data inputs and reduce overfitting through its ensemble learning approach, making it a powerful tool for clinical decision-making. On the other hand, the CNN's impressive sensitivity, specificity, accuracy, and F1-score demonstrate the advanced capabilities of DL in capturing complex patterns within data. The CNN's performance highlights its potential for applications where intricate data structures and high-dimensional features are present. Together, these models illustrate the importance of leveraging both traditional ML techniques and advanced DL methods to achieve optimal predictive accuracy and clinical relevance, ultimately improving patient outcomes through precise and timely interventions. Huang et al. (32) study revealed that out of all the algorithms examined, the RF and ensemble methods exhibited superior predictive performance. The study indicated that RF is particularly effective for predictive modeling of blood pressure during hemodialysis (32). Other studies (33,34) have also shown that CNN can be effectively used to predict hemodialysis.

Our study findings indicate that ML models exhibited higher sensitivities, specificities, accuracies, and F1-scores than DL models, suggesting that particularly the RF model is more effective for predicting ICU stay duration for this patient population. These results underscore the potential of ML techniques in clinical decision-making processes for poisoned hemodialysis patients. The importance of comparing ML and DL models in our study lies in their distinct strengths and applications within predictive modeling for medical outcomes. ML models have shown superior sensitivities, specificities, accuracies, and F1-scores compared to DL models, highlighting their effectiveness in handling structured data and generating precise predictions. This superiority is crucial in clinical settings where accurate predictions can significantly impact patient care and outcomes. ML models, such as RF, excel in interpreting relationships between input variables and outcomes, making them valuable tools for predicting ICU stay durations and guiding timely interventions for poisoned hemodialysis patients (32). Conversely, DL models like CNN and LSTM offer advantages in capturing intricate patterns from complex, unstructured data, although in our study, their performance metrics were comparatively lower. By understanding and leveraging the strengths of both ML and DL approaches, healthcare practitioners can enhance their predictive capabilities and ultimately improve patient care and treatment outcomes in critical medical scenarios.

Jordan and Mitchell (35) mentioned that ML addresses these limitations by enhancing a computer program's performance through experience with specific tasks and performance measures. Essentially, ML aims to automate analytical model building for cognitive tasks like object detection or natural language translation. This is accomplished by applying algorithms that iteratively learn from training data, enabling computers to uncover hidden insights and complex patterns without explicit programming (36). By learning from past computations and identifying regularities in large datasets, ML can facilitate reliable and repeatable decision-making. Consequently, ML algorithms have been successfully applied in numerous domains, including fraud detection, credit scoring, next-best offer analysis, speech and image recognition, and natural language processing (36).

In our study, we have identified several crucial prognostic factors that warrant further discussion. Our findings represent a significant advancement in the field, as previous research has predominantly focused on overall hospital LOS, rather than specifically addressing ICU duration in poisoned patients requiring hemodialysis. The most influential features identified in our study were GCS <8, intubation, AKI, PO_2, BUN, metabolic acidosis, and number of hemodialysis sessions. Among these, Intubation score, GCS score, and ICU admission type emerged as the paramount predictors of ICU stay duration. These results align with and expand upon the findings of Rahimi et al. (37), who reported that intubation, GCS, and ICU admission were significant prognostic factors in poisoned patients undergoing hemodialysis. Furthermore, the study by Brenner et al. (38) on arteriovenous access failure highlights

the impact of complications on hospitalization duration and costs. This emphasizes the need for meticulous vascular access management in poisoned hemodialysis patients to potentially reduce the duration of ICU stays. Additionally, the study by Yan et al. (39) on racial and ethnic disparities in hospitalization rates among hemodialysis patients introduces an important dimension worthy of exploration in future research on ICU stay duration. While our current study did not specifically address these demographic factors, their potential influence on ICU outcomes in poisoned hemodialysis patients merits further investigation. Overall, this section of our study represents a significant advancement in understanding the determinants of ICU stay duration for poisoned patients requiring hemodialysis. By focusing on specific ICU-related outcomes, we provide valuable insights that can inform clinical practices and improve patient management strategies in this critical patient population. Future research should aim to incorporate demographic factors and explore their potential impact on ICU outcomes to build a more comprehensive understanding of the determinants influencing ICU stay duration in poisoned hemodialysis patients.

Study Limitations

The study has several limitations, including the single-center data from Loghman Hakim Hospital, which may affect generalizability, and the sample size of 980 patients, which may not capture all variability in ICU stay duration for hemodialysis patients with poisoning. Future research should aim to include multicenter data to enhance generalizability and increase sample size to capture a broader range of variability.

Additionally, this study acknowledges that the models were developed based on static data, which may limit their applicability to real-time clinical decision-making. Future research should focus on testing and validating these models in dynamic, real-time clinical environments to assess their practical utility and performance in such settings.

Moreover, this study acknowledges potential biases introduced by the class imbalance in the dataset, particularly the higher prevalence of short LOS cases compared to long LOS cases. Although under-sampling techniques were applied to address this issue, such methods might not fully eliminate the inherent bias, potentially impacting the generalizability and robustness of the model's predictions for underrepresented classes. To further address class imbalance, methods such as SMOTE, class weighting, hybrid sampling, or ensemble approaches could be employed, alongside emphasizing evaluation metrics like F1-score and precision-recall-area under the curve.

Lastly, this study acknowledges that temporal trends, such as changes in hospital practices or treatments between 2016 and 2020, were not explicitly accounted for in the analysis. Such trends could impact the generalizability of the findings, as variations in clinical protocols, resource availability, or treatment methods over time might influence the outcomes. Future studies should consider incorporating temporal stratification or modeling to assess the potential impact of these trends and minimize their confounding effects on the results.

CONCLUSION

This study aimed to predict the ICU LOS for hemodialysis patients suffering from poisoning, by comparing ML models with DL models. The results showed that among DL models, the CNN performed best with sensitivity, specificity, accuracy, and F1-score values, while the DNN achieved the best ROC value. However, the RF model, an ML model, outperformed all other DL and ML models, achieving higher scores across sensitivity, accuracy, F1-score, and ROC metrics.

In general, ML models demonstrated superior performance compared to DL models, with higher sensitivities, specificities, accuracies, and F1-scores. These findings suggest that ML models, particularly the RF model, are more effective for predicting ICU stay duration for hemodialysis patients with poisoning. Therefore, incorporating ML models into clinical practice could enhance the prediction and management of ICU stay durations in this patient population, potentially improving outcomes and resource allocation.

ETHICS

Ethics Committee Approval: Ethical approval was obtained from the ethics committee at Shahid Beheshti University of Medical Sciences (approval no: IR.SBMU.RETECH.REC.1401.767, date: 12.02.2023).

Informed Consent: Retrospective study.

Acknowledgment

We thank the Research Deputy of the Shahid Beheshti University of Medical Sciences for financially supporting this project. We also would like to extend our gratitude to all the medical experts at Loghman Hakim Hospital who participated in this study.

FOOTNOTES

Authorship Contributions

Surgical and Medical Practices: M.R., Concept: K.M., M.R.A., H.R., P.E.T.E., M.S., H.K.A., Design: K.M., S.S., M.R.A., B.M., Data Collection or Processing: K.M., S.S., H.R., B.S., M.P., M.S.B., Analysis or Interpretation: K.M., M.R., Literature Search: K.M., B.M., P.E.T.E., H.K.A., V.M., A.H.D., M.R., M.S.B., Writing: K.M., B.M., V.M., M.R.

Conflict of Interest: No conflict of interest was declared by the authors.

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

REFERENCES

1. Yaylacı S, Genc A, Demir M, Cinemre H, Tamer A. Retrospective evaluation of patients at follow-up with acute poisoning in intensive care unit. Niger J Clin Pract. 2016;19:223-6.
2. Kabiri M, Hosseini SM, Veisi F, Mortezaei S, Ahmadi A, Agashteh Z, et al. Investigation of poisoning prevalence and its related factors in patients referred to Farsan's Hospital during 2018-2019, Iran. Epidemiol Health Syst J. 2022;9:75-9.
3. Boedeker W, Watts M, Clausing P, Marquez E. The global distribution of acute unintentional pesticide poisoning: estimations based on a systematic review. BMC public health. 2020;20:1-19.
4. Elgazzar FM, Afifi AM, Shama MAE, Askary AE, El-Sarnagawy GN. Development of a risk prediction nomogram for disposition of acute toxic exposure patients to intensive care unit. Basic Clin Pharmacol Toxicol. 2021;129:256-67.
5. Lindqvist E, Edman G, Hollenberg J, Nordberg P, Forsberg S. Long-term mortality and cause of death for patients treated in intensive care units due to poisoning. Acta Anaesthesiol Scand. 2019;63:500-5.
6. Lindqvist E, Edman G, Hollenberg J, Nordberg P, Ösby U, Forsberg S. Intensive care admissions due to poisoning. Acta Anaesthesiol Scand. 2017;61:1296-304. Erratum in: Acta Anaesthesiol Scand. 2020;64:139.
7. Resiere D, Kallel H, Oxybel O, Chabartier C, Florentin J, Brouste Y, et al. Clinical and epidemiological characteristics of severe acute adult poisoning cases in Martinique: implicated toxic exposures and their outcomes. Toxics. 2020;8:28.
8. Ng F. Ten-year profile of acute poisoning patients presenting to an accident and emergency department requiring intensive care in a regional hospital of Hong Kong. Hong Kong J Emerg Med. 2019;26:3-14.
9. Chan M, Ostermann M. Outcomes of chronic hemodialysis patients in the intensive care unit. Crit Care Res Pract. 2013;2013:715807.
10. D'Agata EM, Mount DB, Thayer V, Schaffner W. Hospital-acquired infections among chronic hemodialysis patients. Am J Kidney Dis. 2000;35:1083-8.
11. Peres IT, Hamacher S, Oliveira FLC, Thomé AMT, Bozza FA. What factors predict length of stay in the intensive care unit? Systematic review and meta-analysis. J Crit Care. 2020;60:183-94.
12. Almashrafi A, Elmontsri M, Aylin P. Systematic review of factors influencing length of stay in ICU after adult cardiac surgery. BMC Health Serv Res. 2016;16:1-12.
13. Shaw C. How can hospital performance be measured and monitored? Copenhagen: World Health Organization, Regional Office for Europe; 2003.
14. Abd-Elrazek MA, Eltahawi AA, Abd Elaziz MH, Abd-Elwhab MN. Predicting length of stay in hospitals intensive care unit using general admission features. Ain Shams Eng J. 2021;12:3691-702.
15. Li C, Chen L, Feng J, Wu D, Wang Z, Liu J, et al. Prediction of length of stay on the intensive care unit based on least absolute shrinkage and selection operator. IEEE Access. 2019;7:110710-21.
16. Daghistani TA, Elshawi R, Sakr S, Ahmed AM, Al-Thwayee A, Al-Mallah MH. Predictors of in-hospital length of stay among cardiac patients: a machine learning approach. Int J Cardiol. 2019;288:140-7.
17. Orooji A, Shanbehzadeh M, Mirbagheri E, Kazemi-Arpanahi H. Comparing artificial neural network training algorithms to predict length of stay in hospitalized patients with COVID-19. BMC Infect Dis. 2022;22:1-13.
18. Wang K, Yan LZ, Li WZ, Jiang C, Wang NN, Zheng Q, et al. Comparison of four machine learning techniques for prediction of intensive care unit length of stay in heart transplantation patients. Front Cardiovasc Med. 2022;9:863642.
19. Afrash MR, Kazemi-Arpanahi H, Nopour R, Tabatabaei ES, Shanbehzadeh M. Proposing an intelligent monitoring system for early prediction of need for intubation among COVID-19 hospitalized patients. J Environ Health Sustain Dev. 2022.
20. Hosseini SM, Rahimi M, Afrash MR, Ziaeefar P, Yousefzadeh P, Pashapour S, et al. Prediction of acute organophosphate poisoning severity using machine learning techniques. Toxicology. 2023;486:153431.
21. Wu J, Lin Y, Li P, Hu Y, Zhang L, Kong G. Predicting prolonged length of ICU stay through machine learning. Diagnostics (Basel). 2021;11:2242.
22. Kaji DA, Zech JR, Kim JS, Cho SK, Dangayach NS, Costa AB, et al. An attention based deep learning model of clinical events in the intensive care unit. PloS One. 2019;14:e0211057.
23. Syed M, Syed S, Sexton K, Syeda HB, Garza M, Zozus M, et al. Application of machine learning in intensive care unit (icu) settings using MIMIC dataset: systematic review. Informatics (MDPI). 2021;8:16.
24. Ayouni S, Maddeh M, Al-Otaibi S, Alazzam MB, Alturki NM, Hajjej F. Development of a smart hospital bed based on deep learning to monitor patient conditions. J Disabil Res. 2023;2:25-36.
25. Malakouti SM, Menhaj MB, Suratgar AA. The usage of 10-fold cross-validation and grid search to enhance ML methods performance in solar farm power generation prediction. Cleaner Eng Technol. 2023;15:100664.
26. Bardenet R, Brendel M, Kégl B, Sebag M. Collaborative hyperparameter tuning. In: Proceedings of the International Conference on Machine Learning (ICML); 2013 May 13; Atlanta, USA. PMLR; 2013:199-207.
27. Papernot N, McDaniel P, Goodfellow I, Jha S, Celik ZB, Swami A. Practical black-box attacks against machine learning. In: Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security; 2017:506-19.
28. Sudmann A. On the media-political dimension of artificial intelligence: deep learning as a black box and OpenAI. Digital Cult Soc. 2018;4:181-200.
29. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;30.
30. Sun J, Sun CK, Tang YX, Liu TC, Lu CJ. Application of SHAP for explainable machine learning on age-based subgrouping mammography questionnaire data for positive mammography prediction and risk factor identification. Healthcare (Basel). 2023;11:2000.
31. Wang F, Wang D, Xu Y, Jiang H, Liu Y, Zhang J. Potential of the non-contrast-enhanced chest CT radiomics to distinguish molecular

subtypes of breast cancer: a retrospective study. Front Oncol. 2022;12:848726.

32. Huang JC, Tsai YC, Wu PY, Lien YH, Chien CY, Kuo CF, et al. Predictive modeling of blood pressure during hemodialysis: a comparison of linear model, random forest, support vector regression, XGBoost, LASSO regression and ensemble method. Comput Methods Programs Biomed. 2020;195:105536.
33. Wang F, Wang Y, Tian Y, Zhang P, Chen J, Li J. Pattern recognition and prognostic analysis of longitudinal blood pressure records in hemodialysis treatment based on a convolutional neural network. J Biomed Inform. 2019;98:103271.
34. Ota K, Nishiura Y, Ishihara S, Adachi H, Yamamoto T, Hamano T. Evaluation of hemodialysis arteriovenous bruit by deep learning. Sensors (Basel). 2020;20:4852.
35. Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. Science. 2015;349:255-60.
36. Janiesch C, Zschech P, Heinrich K. Machine learning and deep learning. Electron Mark. 2021;31:685-95.
37. Rahimi M, Afrash MR, Shadnia S, Mostafazadeh B, Evini PET, Bardsiri MS, et al. Prediction the prognosis of the poisoned patients undergoing hemodialysis using machine learning algorithms. BMC Med Inform Decis Mak. 2024;24:38.
38. Brenner L, Singh AK, Campbell D, Frei F, Winkelmayer WC. Associations between demographic factors and provider structures on cost and length of stay for hemodialysis patients with vascular access failure. Clin J Am Soc Nephrol. 2006;1:455-61.
39. Yan G, Norris KC, Greene T, Alison JY, Ma JZ, Yu W, Cheung AK. Race/ethnicity, age, and risk of hospital admission and length of stay during the first year of maintenance hemodialysis. Clin J Am Soc Nephrol. 2014;9:1402-9.

Medical Journal of Bakırköy

Med J Bakirkoy 2025;21(3):288-295
DOI: 10.4274/BMJ.galenos.2025.2025.2-1

Research

Investigation of an Open AI Model in the Analysis of STN Microelectrode Recordings: Consistency With Clinicians and Potential for DBS Targeting

STN Analizinde Açık Bir Yapay Zeka Modelinin Araştırılması Mikroelektrot Kayıtları: Klinisyenlerle Tutarlılık ve DBS Hedefleme Potansiyeli

Ozan Haşimoğlu[1], Ayça Altınkaya[2], Taha Hanoğlu[1], Tuba Özge Karaçoban[1], Nur Bahar Geylan[1], Fırat Demir[1], Bekir Tuğcu[1]

[1]University of Health Sciences Türkiye, Başakşehir Çam and Sakura City Hospital, Clinic of Neurosurgery, İstanbul, Türkiye

[2]University of Health Sciences Türkiye, Başakşehir Çam and Sakura City Hospital, Clinic of Neurology, İstanbul, Türkiye

ABSTRACT

Objective: Deep brain stimulation (DBS) of the subthalamic nucleus (STN) requires precise electrode placement, often assisted by microelectrode recording (MER). However, the interpretation of MER remains highly subjective, varying among clinicians based on experience. This study evaluates the ability of an artificial intelligence (AI) model (ChatGPT 4.0) to classify STN MER recordings and evaluate its consistency with experienced and less experienced clinicians.

Methods: A total of 32 STN MER recordings were independently evaluated by two experienced clinicians, two less experienced clinicians, and the AI model. Classifications were assigned to artifact, thalamus, silent, STN, suspicious STN, substantia nigra, and N/A (no recording), categories. Fleiss' Kappa was used to assess inter-rater consistency, while Cohen's Kappa measured agreement between generative pre-trained transformer (GPT) and each clinician. Additionally, precision and recall were calculated for each category.

Results: The overall Fleiss' Kappa among all evaluators was 0.544, with higher agreement among experienced clinicians (0.738) compared to less experienced ones (0.631). GPT showed low agreement with both groups, with Cohen's Kappa values ranging from 0.341 to 0.375. GPT demonstrated the highest accuracy in detecting STN (73.47%), but its performance was significantly lower for other categories. Within-category consistency (14.28%) indicated variability in transition zones, with a misclassification rate of 45.87% compared to the majority opinion of clinicians.

Conclusion: While GPT exhibited partial consistency with clinicians in identifying the STN, its reliability in classifying transition zones and adjacent structures was low. For AI to serve as a reliable tool in STN targeting, further refinement of its algorithms and expanded training datasets is necessary. Although GPT is not yet suitable for clinical decision making, its potential for future DBS applications is promising.

Keywords: Deep brain stimulation, subthalamic nucleus, microelectrode recordings, artificial intelligence, machine learning, ChatGPT

ÖZ

Amaç: Subtalamik nükleusun (STN) derin beyin stimülasyonu (DBS), genellikle mikroelektrot kayıtları (MER) ile desteklenen hassas elektrot yerleşimi gerektirir. Ancak, MER'nin yorumlanması oldukça özneldir ve klinisyenler arasında deneyime bağlı olarak değişkenlik gösterebilir. Bu çalışma, bir yapay zeka (YZ) modelinin (ChatGPT 4.0) STN MER kayıtlarını sınıflandırma yetisini değerlendirerek, deneyimli ve daha az deneyimli klinisyenlerle tutarlılığını analiz etmektedir.

Gereç ve Yöntem: Toplam 32 STN MER kaydı, iki deneyimli klinisyen, iki daha az deneyimli klinisyen ve YZ modeli tarafından bağımsız olarak değerlendirildi. Kayıtlar artefakt, talamus, sessiz, STN, şüpheli STN, substantia nigra ve N/A (kayıt yok) kategorilerine ayrıldı. Değerlendiriciler arasındaki tutarlılık Fleiss' Kappa, üretici önceden eğitilmiş dönüştürücü (GPT) ile her bir klinisyen arasındaki uyum ise Cohen's Kappa ile ölçüldü. Ayrıca, her kategori için kesinlik ve duyarlılık hesaplandı.

Bulgular: Tüm değerlendiriciler arasında genel Fleiss' Kappa 0,544 olarak bulundu; deneyimli klinisyenler arasında tutarlılık 0,738, daha az deneyimli klinisyenler arasında ise 0,631 idi. GPT'nin her iki grup ile uyumu düşük olup, Cohen's Kappa 0,341 ila 0,375 arasında değişti. GPT,

Address for Correspondence: Ozan Haşimoğlu, MD, University of Health Sciences Türkiye, Başakşehir Çam and Sakura City Hospital, Clinic of Neurosurgery, İstanbul, Türkiye
E-mail: ozanhasim@hotmail.com **ORCID ID:** orcid.org/0000-0003-1394-5188

Cite as: Haşimoğlu O, Altınkaya A, Hanoğlu T, Karaçoban TÖ, Geylan NB, Demir F, et al. Investigation of an Open AI model in the analysis of STN microelectrode recordings: consistency with clinicians and potential for DBS sting. Med J Bakirkoy. 2025;21(3):288-295

Received: 05.02.2025
Accepted: 08.04.2025
Publication Date: 03.09.2025

STN tespitinde en yüksek doğruluğu (%73,47) gösterse de, diğer kategorilerde performansı belirgin şekilde daha düşüktü. Kategori içi tutarlılığı %14,28 olarak belirlenen modelin, geçiş bölgelerindeki değişkenliği yüksek olup, klinisyenlerin çoğunluk görüşüne göre yanlış sınıflandırma oranı %45,87idi.

Sonuç: GPT, STN'yi tanımlamada klinisyenlerle kısmi bir tutarlılık gösterse de, geçiş bölgeleri ve komşu yapıları sınıflandırmadaki güvenilirliği düşüktü. YZ'nin STN hedeflemede güvenilir bir araç olabilmesi için algoritmalarının iyileştirilmesi ve eğitim veri setlerinin genişletilmesi gereklidir. GPT henüz klinik karar vermeye uygun olmasa da, gelecekteki DBS uygulamaları için umut vadetmektedir.

Anahtar Kelimeler: Derin beyin stimülasyonu, subtalamik nükleus, mikroelektrot kaydı, yapay zeka, makine öğrenmesi, ChatGPT

INTRODUCTION

The subthalamic nucleus (STN) is one of the most frequently targeted structures in deep brain stimulation (DBS) for Parkinson's disease (1). Given that the accuracy of electrode placement in DBS surgery directly impacts clinical outcomes, microelectrode recordings (MER) obtained intraoperatively serve as a crucial tool for delineating the boundaries of the STN (2-4). The interpretation of MER data relies on clinical expertise to distinguish the STN from surrounding structures (5-7). However, precisely defining the exact boundaries of the STN and ensuring consistency among different evaluators remain challenging (8). The interpretation of MER recordings is inherently subjective, often yielding varying results when assessed by different clinicians. These discrepancies can arise due to differences in evaluator experience and anatomical variations in the STN among individuals (9). In particular, identifying the onset and termination points of the STN, as well as accurately interpreting transition zones involving silent areas or structures such as the thalamus and substantia nigra, presents a significant challenge (10). Therefore, it is crucial to develop methods that enhance consistency among clinicians and make the decision-making process more objective.

In recent years, artificial intelligence (AI) and machine learning-based models have been increasingly utilized as supportive tools in the analysis of neurophysiological recordings. The automated classification of MER recordings can both accelerate the surgical process by saving time and provide a more objective analytical approach by minimizing human-induced variations in interpretation (11,12). However, the reliability of AI models and their consistency with clinicians remain subjects of ongoing debate (2,8,9).

The aim of this study is to compare how four clinicians (two experienced and two less experienced) and the OpenAI ChatGPT 4.0 model evaluate STN MER recordings, and to statistically analyze the consistency of their interpretations. The study examines the agreement among clinicians with different levels of experience, the consistency between clinicians and AI, and the distribution of errors across specific depth levels. In particular, it investigates the extent to which uncertainties in the boundary regions of the STN hinder the establishment of a shared interpretation, and it assesses the consistency of evaluations across different classification categories. By selecting ChatGPT as the AI model, the study aims to assess the reliability of a widely accessible software that clinicians can use. These analyses are expected to provide valuable insights into improving clinical decision-making in STN targeting, evaluating the potential of AI models, and developing new strategies for achieving a more objective interpretation of STN MER recordings.

METHODS

Participant Selection

This study was conducted with the approval of the Scientific Research Ethics Committee of the University of Health Sciences Türkiye, Başakşehir Çam and Sakura City Hospital (approval no: 47, date: 04.12.2024). Between 2021 and 2024, a total of thirty-two MERs from patients diagnosed with Parkinson's disease who underwent STN-DBS performed by the same surgical team were included in this study. The surgical decision was made by a multidisciplinary movement disorder board, including a neurosurgeon, movement disorder neurologist, psychiatrist, neuropsychologist, speech therapist, and physiotherapist. Patients were evaluated based on objective criteria established by the board, and those deemed suitable for DBS surgery.

All patients underwent comprehensive preoperative clinical evaluations, including the Unified Parkinson's Disease Rating Scale, Parkinson's Disease Questionnaire, and an extensive neuropsychological test battery. Patients included in the study underwent these assessments both preoperatively and postoperatively. Only those who demonstrated significant clinical benefit from DBS, had confirmed STN stimulation in all postoperative evaluations, and continued to live with an active stimulator system, were considered for analysis.

Patients with incomplete, inaccurate, or unreliable electrophysiological recordings, those who did not experience the expected benefit from DBS, or those whose stimulation systems were deactivated were excluded from

the study. Written informed consent was obtained from all of the participants.

Surgical Planning and Electrophysiological Recording

All patients underwent high-resolution 1.5 Tesla or 3 Tesla magnetic resonance imaging (MRI) 1 to 3 days prior to surgery. The MRI protocol included T1-weighted images, T2-weighted images, and contrast-enhanced T1 images, as well as diffusion tensor imaging (DTI). The DTI sequences were used to map white matter tracts. Targeting was performed using both direct and indirect methods. The dorsolateral region of the STN was identified as the optimal target for controlling motor symptoms in Parkinson's disease. The initial coordinates were determined based on the anterior commissure-posterior commissure (PC) line, with a starting reference of X: ±12, Y: -2, Z: -4 mm. Adjustments were made using direct MRI visualization to precisely target the dorsolateral STN. All surgical planning was conducted using Stealth and BrainLab Elements™ surgical navigation software. All patients remained awake during surgery, and a stereotactic frame was placed on the morning of the procedure. A 1-mm slice computed tomography (CT) scan was obtained with the frame in place and fused with preoperative MRI to determine the stereotactic coordinates. MER and macrostimulation were performed in all patients. Electrophysiological recordings were obtained intraoperatively using Alpha Omega, AlphaRS, and FHC Guideline 5 systems. The MER technique had been detailed in a previous study (13). MER was initiated 10 mm above the target coordinate, advancing in 1-mm increments until reaching 5 mm below, and then in 0.5 mm increments until the final depth was reached. Recordings were terminated when the STN electrophysiological activity ended, when a substantia nigra pars reticulata recording was obtained, or, at +4.5 mm. Macrostimulation was performed in the orientation where the longest STN recording was observed. MER recordings were stored in PDF format for further analysis (Figure 1). During macrostimulation, motor responses to stimulation were observed and evaluated. The placement of the DBS electrodes was guided by MER and macrostimulation findings, following a Ben-Gun orientation. The final positioning of the electrodes ensured that the middle contact points were aligned with the intended target, while the tip of the electrode was placed in contact with the substantia nigra. The decision to use directional or non-directional electrodes was made intraoperatively based on electrophysiological findings and the patient's motor response. DBS hardware included Boston Scientific Gevia, Genus, and Medtronic Activa RC/PC models. The stimulator implantation was performed during a single surgical session, with the device placed in the midclavicular region. Postoperatively, 1 mm slice CT scans were fused with preoperative MRI to assess electrode placement. Revisions were performed for patients with a deviation greater than 2 mm from the planned target. Deep brain stimulators were typically activated within 3 to 7 days postoperatively. Prior to activation, patients underwent a 12-hour medication withdrawal. Neurology and neurosurgery specialists performed stimulator activation, and motor effects were evaluated to determine stimulation parameters. After optimizing individualized medication and stimulation settings, patients were discharged.

Artificial Intelligence Analysis Process

In this study, 32 STN MERs were independently evaluated by two experienced and two less experienced MER clinicians. Each clinician was instructed to classify each depth level into one of the following categories: artifact, thalamus,

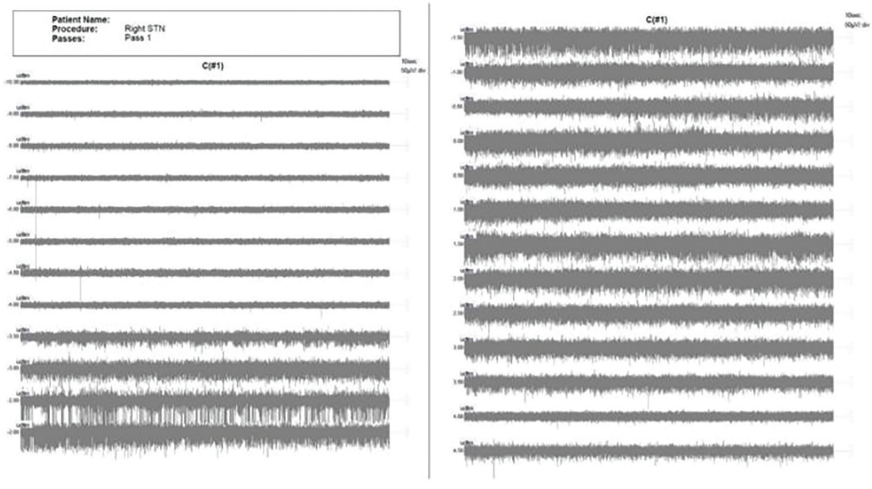

Figure 1. Sample microelectrode recording (MER) output. This figure shows a sample MER output, displaying depth annotations on the left side and 10-second recordings obtained at each depth level

silent, STN, suspicious STN, substantia nigra, and N/A (no recording). The same recordings were also analyzed by the ChatGPT 4.0 model, and the results were recorded. A specially designed prompt was used to enable ChatGPT to analyze MER recordings. Through this prompt, the model independently assessed wave patterns at each depth level and assigned them to the predetermined categories. The prompt structure used for generative pre-trained transformers (GPTs) interpretation of MER recordings was designed as follows (8,9,14).

Task Description: This prompt was used to analyze 10-second MER recordings in PDF format, categorizing wave patterns based on depth levels. Each depth level was independently evaluated.

Depth Values: Recordings began 10 mm above the target and proceeded down to -4.5 mm. The scale on the left side of the recordings indicated depth levels, and the analysis was conducted separately for each level.

Categories: Evaluations were assigned to one of the following categories based on wave patterns:

• **Silent:** Regions with little to no electrical activity. The wave pattern is nearly silent. This activity is typically observed outside high-activity regions such as the STN, the thalamus, or within the zona incerta.

• **Thalamus:** Contains low-frequency, regular patterns. Typical firing frequency ranges from 10 to 30 Hz, with action potential amplitude between 50 and 100 mV. This region is involved in motor and sensory transmission and is typically observed before the STN.

• **Suspicious STN:** Regions where baseline broadening begins or diminishes are included in this category. Changes in frequency and amplitude occur before or after bursts intensify. This phenomenon is generally observed after the thalamus and in transition zones between the STN and substantia nigra.

• **STN:** High-frequency, high-amplitude regions containing intense burst activities are classified as STN. The firing frequency ranges from 15 to 30 Hz, with action potential amplitude between 60 and 80 mV. It plays a critical role in motor control and is observed at depths where baseline broadening becomes prominent.

• **Substantia Nigra:** Refers to specific regions characterized by low-frequency, regular activity. Dopaminergic neurons in the pars compacta fire at 1-8 Hz, while GABAergic neurons in the pars reticulata fire at 20-40 Hz. Action potential amplitude ranges from 40 to 80 mV.

• **Artifact:** Disturbances or unwanted noise arising during electrical recording are classified as artifacts. These may result from electrode movement or environmental factors.

• **N/A:** Depths where no recordings were obtained or where the data were unprocessable were classified as N/A. Cases where measurements were not performed during electrode transitions were also included in this category.

Analysis Principles: Each depth level was analyzed independently. All assessments were performed objectively based on predefined wave patterns.

Output Format: All analyses were presented in a table format, systematically recording the assigned categories for each depth level.

Throughout the study, manual corrections were made only in cases of technical errors in GPT's decision-making processes. However, no external intervention was applied to the category decisions made by the model.

Statistical Analysis

To assess the overall consistency between the classifications made by clinicians and GPT, Fleiss' Kappa coefficient was calculated. Fleiss' Kappa is a multi-rater agreement coefficient used to measure the consistency among multiple evaluators. Additionally, the consistency of each clinician with GPT was calculated separately using Cohen's Kappa coefficient. Cohen's Kappa is a statistical method that quantifies the level of agreement between two raters, adjusting for chance agreement to determine the actual level of consistency. Values below 0 indicate no agreement between evaluators, suggesting a level of disagreement worse than randomness. Values between 0.00 and 0.20 indicate very weak agreement, meaning no meaningful association between evaluators' decisions. Values between 0.21 and 0.40 indicate weak agreement, where the evaluators' decisions are only slightly better than random chance. Values between 0.41 and 0.60 indicate moderate agreement, suggesting that evaluators partially share a common perspective. Values between 0.61 and 0.80 indicate strong agreement, indicating evaluators mostly make the same decisions. Values between 0.81 and 1.00 indicate almost perfect agreement, meaning evaluators reach nearly identical classifications. The consistency among experienced clinicians and less experienced clinicians were analyzed separately. Additionally, the agreement between clinician groups and GPT was compared. Furthermore, GPT's reliability for each classification category was assessed using precision and recall calculations to evaluate the accuracy of its assigned classifications. For depth-based analysis, classifications from GPT and clinicians were

compared across different depth levels, with a particular focus on the transition zones at the STN's entry and exit points. The stability of GPT's classifications was measured by evaluating its consistency with previous and subsequent predictions, providing insights into category transition consistency. This analysis aimed to determine how reliably GPT identified specific categories and to what extent its predictions fluctuated in transition zones. To assess misclassification rates, cases where GPT's classification did not align with the majority opinion were identified. The majority opinion was defined as the most frequently chosen category among the four clinicians, and GPT's consistency with this majority classification was analyzed. Statistical analyses were conducted using SPSS version 22 (IBM Corp., Armonk, NY, USA), and data visualization was performed with Python Matplotlib (Matplotlib Development Team, Python Software Foundation, USA).

RESULTS

In this study, 32 STN MERs were independently evaluated by two experienced clinicians, two less experienced clinicians, and an AI model (ChatGPT 4.0), and the results were analyzed statistically. The overall consistency among all evaluators, calculated using Fleiss' Kappa coefficient, was found to be 0.544. The Fleiss' Kappa coefficient between the experienced clinicians was 0.738, while the Fleiss' Kappa coefficient between the less experienced clinicians was 0.631 (Figure 2). The agreement between GPT and clinicians was analyzed using Cohen's Kappa coefficient. Cohen's Kappa coefficient between experienced clinician 1 and GPT was 0.356, while the value between experienced clinician 2 and GPT was 0.375. Among the less experienced clinicians, Cohen's Kappa coefficient between beginner clinician 1 and GPT was 0.341, and between beginner clinician 2 and GPT was 0.375. These findings indicate that GPT exhibited low consistency with both experienced and less experienced clinicians (Figure 3). In category-based analysis, the accuracy of GPT in identifying the STN category was 73.47%; the silent category, it was 66.41%; the artifact category, it was 42.85%; the thalamus category, it was 7.69%; and the substantia nigra category, it was 3.33%. GPT demonstrated the highest accuracy in identifying the STN category, suggesting significant agreement with clinicians in recognizing the STN region. The 66.41% accuracy for the silent category indicates that the model was relatively successful in identifying low-activity regions. For the artifact category, the accuracy was 42.85%, indicating that the model struggled to distinguish artifacts caused by electrode movement or environmental factors. The accuracy rates for the thalamus and substantia nigra categories were 7.69% and 3.33%, respectively, indicating that GPT was unreliable in classifying these regions. The precision for STN was 0.73, and recall for STN was 0.65. For the silent category, precision was 0.66, and recall was 0.59. For artifact, precision was 0.42, and recall was 0.37. For thalamus, precision was 0.08, and recall was 0.12. For the substantia nigra, precision was 0.03, and recall was 0.05. These findings indicate that GPT was relatively successful in identifying the STN, but it exhibited a significant error margin in transition zones (Figure 4). In depth-based analysis, GPT's category transition consistency was calculated as 51.5%, indicating that the model selected the same category for consecutive depth

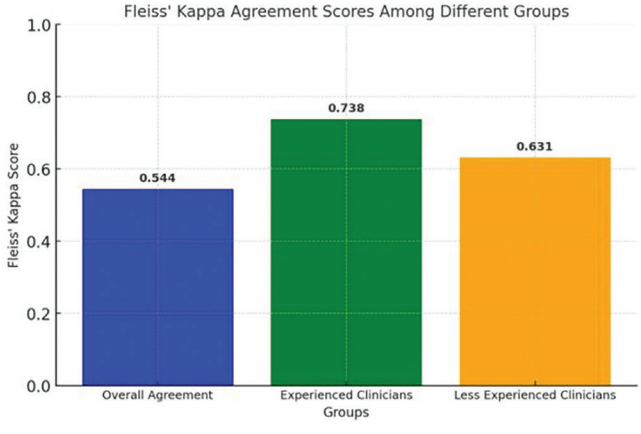

Figure 2. Fleiss' Kappa agreement scores among evaluator groups. This figure illustrates the overall Fleiss' Kappa agreement scores among all evaluators. The overall agreement was calculated as 0.544, reflecting moderate consistency among all evaluators. Among the experienced clinicians, the agreement score was 0.738, indicating strong consistency, while the agreement score among the less experienced clinicians was 0.631, showing moderate consistency

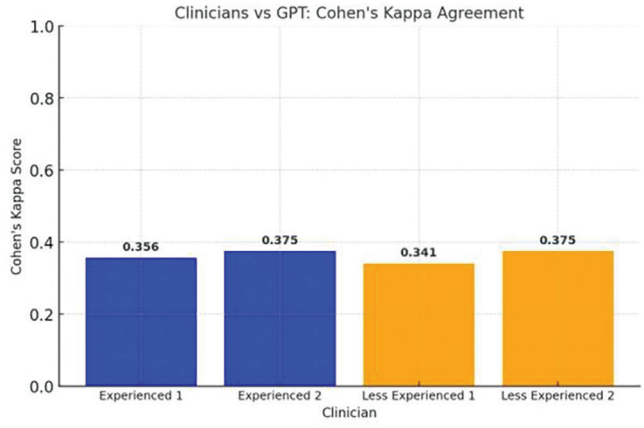

Figure 3. Cohen's Kappa agreement between GPT and clinicians. This figure shows the Cohen's Kappa agreement scores between GPT and individual clinicians

GPT: Generative pre-trained transformer

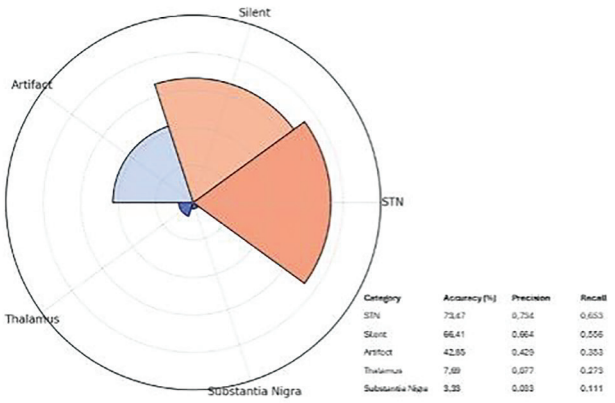

Figure 4. Performance of GPT in identifying categories: accuracy, precision, and recall. This figure presents the accuracy, precision, and recall values for each category evaluated by GPT. The STN category showed the highest performance, with an accuracy of 73.47%, precision of 0.734, and recall of 0.653, indicating relatively successful identification of the STN region

GPT: Generative pre-trained transformer, STN: Subthalamic nucleus

levels 51.5% of the time. The misclassification rate was found to be 45.87%, meaning that GPT produced different results from the majority opinion of the clinicians in 45.87% of cases. To measure the model's tendency to maintain a category assignment over multiple depths, within-category consistency was calculated as 14.28%, indicating that GPT exhibited a high degree of variability in its decision-making process.

DISCUSSION

DBS procedures are becoming increasingly widespread, and successfully performing these procedures requires a high level of expertise. In some cases, the inability to obtain preoperative imaging of the desired quality or the difficulty in identifying the STN sweet spot solely through radiological methods demonstrate that relying exclusively on imaging techniques does not always lead to optimal targeting (8,9,11,12). MER is a valuable alternative for more reliable targeting. However, MER techniques are still predominantly reliant on clinicians' subjective visual and auditory assessments during surgery, which introduces variability in outcomes. To address this, AI-based technologies have been developed to both assist novice surgical teams in mastering this highly specialized technique and to reduce subjectivity in decision-making. These advancements aim to enable real-time and automated evaluation of MER recordings, enhancing the objectivity and consistency of the process (2,8,15).

These methods have primarily been developed by first implementing preprocessing and artifact removal steps, followed by the application of techniques designed to detect differences in various signal characteristics (9,12). Some approaches focus on spike-related parameters (16,17) others analyze power changes in specific frequency bands (18-20), and yet others rely on wavelet analysis (21,22). Another group of methods is based entirely on deep learning algorithms (23-26). In these studies, the supervised learning method has been predominantly used for AI training. The primary reason for this preference is that supervised learning allows AI to be trained with smaller datasets, making the training process easier and faster. Although supervised learning enables rapid implementation, its performance ceiling is inherently limited by the expertise of the annotators who label the training data. Since AI learns both correct and incorrect classifications, its effectiveness is directly influenced by the accuracy and consistency of the human-provided labels (2,8,15). Despite these limitations, the full integration of AI-driven real-time models into DBS procedures appears to be a realistic near-future possibility. AI could serve as a guidance tool for specialists by providing feedback and fine-tuning suggestions, thereby improving decision-making in MER interpretation (15). When designing this study, our goal was to evaluate an AI model with access to open-source big data and assess its ability to interpret MER recordings independently, without relying on clinician directives or modifications to its source code. The objective was to determine whether the model could function like a human evaluator who retrospectively analyzes MER recordings and to assess its suitability for clinical use. In this regard, our study differs from previous AI-based research in this field. At this point, an unsupervised machine learning approach, which could eliminate the need for human expert input, was considered. While in theory this method could overcome the limitations of supervised learning and lead to more advanced models, the requirement for extremely large datasets in unsupervised learning remains a significant barrier. At present, no clinical setting has access to datasets of the necessary scale, making the application of unsupervised learning in this field an unrealistic goal.

Another significant issue with the existing methods is that they are either commercially available at high costs or, if open source, they require complex technical knowledge and software expertise, making them difficult for clinicians to utilize. At this point, the idea of using ChatGPT, which is easily accessible and does not require extensive technical background knowledge, emerges as a potential tool to assist in the interpretation of MER recordings. In this study, we aimed to explore this possibility by evaluating 32

STN MERs, independently assessed by two experienced clinicians, two less experienced clinicians, and an AI model (ChatGPT 4.0), with statistical analyses conducted on the results. However, the findings indicate that this AI model still exhibits low consistency with experienced and less experienced clinicians.

In this study, GPT demonstrated the highest accuracy in the STN category, indicating partial consistency with clinicians in identifying the STN region. Similar findings have been reported in AI models focused on background neural activity (17,27-29). The STN exhibits twice the background activity compared to its neighboring structures (30), a phenomenon likely associated with the high neuronal density within the STN (14). Since the AI model used in this study primarily focuses on basal activity, it performed better in identifying the STN. However, the study also revealed that the model has a significant margin of error in the STN entry and exit zones as well as in adjacent structures. In the model proposed by Rajpurohit et al. (16), the accuracy for STN entry and exit regions was reported to be between 60% and 80%. Similarly, Chaovalitwongse et al. (28) successfully identified STN and its neighboring structures with approximately 90% accuracy using a combination of seven spike-dependent and six spike-independent approaches. These findings highlight that feature-based machine learning models have demonstrated higher accuracy than the AI model in our study.

In summary, our study demonstrates that while GPT shows partial consistency with clinicians in identifying the STN, it exhibits low accuracy in other categories. The low within-category consistency (14.28%) indicates that the model exhibits significant variability in transition zones and makes unstable decisions. Although GPT has achieved a certain level of accuracy in STN detection, its reliability remains low in differentiating transition zones and low-activity structures. For the AI model to be considered a supportive tool in STN targeting, its training must be expanded, and its algorithms must be refined to enhance precision.

Study Limitations

This study has several limitations. First, the evaluation of GPT's performance was based on a limited dataset of MER recordings obtained from a single center, which may restrict the generalizability of the results. Second, the ground truth was established through the consensus of clinicians, which, although a common method, may still reflect inter-rater variability and subjective interpretation. Third, the model was not specifically trained or fine-tuned on electrophysiological data related to DBS, which may have affected its performance in distinguishing between critical anatomical regions. Lastly, technical constraints in data formatting and input length limitations may have impacted the accuracy and consistency of the model's classifications.

CONCLUSION

AI-based algorithms, which are increasingly assisting us in various fields, also hold significant potential to support DBS procedures. Soon, the impact of open AI models in clinical practice is expected to grow. However, at present, many clinicians remain unable to benefit from these advancements due to both technical and financial barriers. While ChatGPT is widely used in various fields as a practical and cost-effective tool, our findings indicate that it is not yet sufficiently reliable for DBS-related applications. Nonetheless, its potential for future development is promising. At this stage, while the model may serve as a guiding tool, its integration into clinical decision-making processes still appears to be premature. Further studies, in light of ongoing advancements and updates, may reveal whether this situation will change.

ETHICS

Ethics Committee Approval: This study was conducted with the approval of the Scientific Research Ethics Committee of the University of Health Sciences Türkiye, Başakşehir Çam and Sakura City Hospital (approval no: 47, date: 04.12.2024).

Informed Consent: Written informed consent was obtained from all of the participants.

FOOTNOTES

Authorship Contributions

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

Conflict of Interest: No conflict of interest was declared by the authors.

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

REFERENCES

1. Hariz M, Blomstedt P. Deep brain stimulation for Parkinson's disease. J Intern Med. 2022;292:764-78.
2. Wan KR, Maszczyk T, See AAQ, Dauwels J, King NKK. A review on microelectrode recording selection of features for machine learning in deep brain stimulation surgery for Parkinson's disease. Clin Neurophysiol. 2019;130:145-54.
3. Vinke RS, Geerlings M, Selvaraj AK, Georgiev D, Bloem BR, Esselink RA, et al. The role of microelectrode recording in deep brain stimulation surgery for Parkinson's disease: a systematic review and meta-analysis. J Parkinsons Dis. 2022;12:2059-69.

4. Al Awadhi A, Tyrand R, Horn A, Kibleur A, Vincentini J, Zacharia A, et al. Electrophysiological confrontation of Lead-DBS-based electrode localizations in patients with Parkinson's disease undergoing deep brain stimulation. Neuroimage Clin. 2022;34:102971.
5. Pastor J, Vega-Zelaya L. Can we put aside microelectrode recordings in deep brain stimulation surgery? Brain Sci. 2020;10:571.
6. Kinfe TM, Vesper J. The impact of multichannel microelectrode recording (MER) in deep brain stimulation of the basal ganglia. Acta Neurochir Suppl. 2013;117:27-33.
7. van den Munckhof P, Bot M, Schuurman PR. Targeting of the subthalamic nucleus in patients with Parkinson's disease undergoing deep brain stimulation surgery. Neurol Ther. 2021;10:61-73.
8. Lima J, de Carvalho M, Dias JM. Analysis and classification of microelectrode recordings in deep brain stimulation surgery. Academia. 2015:1-13.
9. Coelli S, Levi V, Del Vecchio Del Vecchio J, Mailland E, Rinaldo S, Eleopra R, et al. Characterization of microelectrode recordings for the subthalamic nucleus identification in Parkinson's disease. Annu Int Conf IEEE Eng Med Biol Soc. 2020;2020:3485-8.
10. Novak P, Daniluk S, Ellias SA, Nazzaro JM. Detection of the subthalamic nucleus in microelectrographic recordings in Parkinson disease using the high-frequency (> 500 Hz) neuronal background. J Neurosurg. 2007;106:175-9.
11. Valsky D, Blackwell KT, Tamir I, Eitan R, Bergman H, Israel Z. Real-time machine learning classification of pallidal borders during deep brain stimulation surgery. J Neural Eng. 2020;17:016021.
12. Martin T, Jannin P, Baxter JSH. Generalisation capabilities of machine-learning algorithms for the detection of the subthalamic nucleus in micro-electrode recordings. Int J Comput Assist Radiol Surg. 2024;19:2445-51.
13. Tugcu B, Hasimoglu O, Altinkaya A, Barut O, Hanoglu T. Comparison of electrophysiological and radiological subthalamic nucleus length and volume. Turk Neurosurg. 2023;33:126-33.
14. Benazzouz A, Breit S, Koudsie A, Pollak P, Krack P, Benabid AL. Intraoperative microrecordings of the subthalamic nucleus in Parkinson's disease. Mov Disord. 2002;17 Suppl 3:S145-9.
15. Inggas MAM, Coyne T, Taira T, Karsten JA, Patel U, Kataria S, et al. Machine learning for the localization of subthalamic nucleus during deep brain stimulation surgery: a systematic review and meta-analysis. Neurosurg Rev. 2024;47:774.
16. Rajpurohit V, Danish SF, Hargreaves EL, Wong S. Optimizing computational feature sets for subthalamic nucleus localization in DBS surgery with feature selection. Clin Neurophysiol. 2015;126:975-82.
17. Schiaffino L, Muñoz AR, Martínez JG, Villora JF, Gutiérrez A, Torres IM. STN area detection using K-NN classifiers for MER recordings in Parkinson patients during neurostimulator implant surgery. J Phys Conf Ser 2016;705:012050.
18. Vargas Cardona HD, Álvarez MA, Orozco ÁA. Multi-task learning for subthalamic nucleus identification in deep brain stimulation. Int J Mach Learn Cybern 2018;9:1181-92.
19. Valsky D, Marmor-Levin O, Deffains M, Eitan R, Blackwell KT, Bergman H, et al. Stop! border ahead: automatic detection of subthalamic exit during deep brain stimulation surgery. Mov Disord. 2017;32:70-9.
20. Khosravi M, Atashzar SF, Gilmore G, Jog MS, Patel RV. Intraoperative localization of STN during DBS surgery using a data-driven model. IEEE J Transl Eng Health Med. 2020;8:1-9.
21. Karthick P, Wan KR, Qi ASA, Dauwels J, King NKK. Automated detection of subthalamic nucleus in deep brain stimulation surgery for Parkinson's disease using microelectrode recordings and wavelet packet features. J Neurosci Methods. 2020;343:108826.
22. Park KH, Sun S, Lim YH, Park HR, Lee JM, Park K, et al. Clinical outcome prediction from analysis of microelectrode recordings using deep learning in subthalamic deep brain stimulation for Parkinsons disease. PloS One. 2021;16:e0244133.
23. Peralta M, Bui QA, Ackaouy A, Martin T, Gilmore G, Haegelen C, et al. SepaConvNet for localizing the subthalamic nucleus using one second micro-electrode recordings. Annu Int Conf IEEE Eng Med Biol Soc. 2020;2020:888-93.
24. Martin T, Peralta M, Gilmore G, Sauleau P, Haegelen C, Jannin P, et al. Extending convolutional neural networks for localizing the subthalamic nucleus from micro-electrode recordings in Parkinson's disease. Biomed Signal Process Control 2021;67:102529.
25. Hosny M, Zhu M, Gao W, Fu Y. Deep convolutional neural network for the automated detection of Subthalamic nucleus using MER signals. J Neurosci Methods. 2021;356:109145.
26. Xiao L, Li C, Wang Y, Si W, Lin H, Zhang D, et al. Amplitude-frequency-aware deep fusion network for optimal contact selection on STN-DBS electrodes. Sci China Inf Sci 2022;65:140404.
27. Cagnan H, Dolan K, He X, Contarino MF, Schuurman R, van den Munckhof P, Wadman WJ, Bour L, Martens HC. Automatic subthalamic nucleus detection from microelectrode recordings based on noise level and neuronal activity. J Neural Eng. 2011;8:046006.
28. Chaovalitwongse WA, Jeong YS, Jeong MK, Danish SF, Wong S. Pattern recognition approaches for identifying subcortical targets during deep brain stimulation surgery. IEEE Intell Syst 2011;26:54-63.
29. Ciecierski K, Mandat T, Rola R, Raś ZW, Przybyszewski AW. Computer aided subthalamic nucleus (STN) localization during deep brain stimulation (DBS) surgery in Parkinson's patients. Ann Acad Med Siles 2014;68:275-83.
30. Kano T, Katayama Y, Kobayashi K, Kasai M, Oshima H, Fukaya C, et al. Multiple-cell spike density and neural noise level analysis by semimicroelectrode recording for identification of the subthalamic nucleus during surgery for Parkinson's disease. Neuromodulation. 2008;11:1-7.

Medical Journal of Bakırköy

Med J Bakirkoy 2025;21(3):296-304
DOI: 10.4274/BMJ.galenos.2025.2025.4-15

Research

Comparison of Long-term Outcomes of On-pump and Off-pump Techniques in Isolated Coronary Artery Bypass Surgery: A Cohort Study

İzole Koroner Arter Bypass Cerrahisinde Pompalı ve Pompasız Tekniklerin Uzun Dönem Sonuçlarının Karşılaştırılması: Bir Kohort Çalışması

Hasan Toz, Yusuf Kuserli, Gülsüm Türkyılmaz, Ali Aycan Kavala, Saygın Türkyılmaz

University of Health Sciences Türkiye, Bakırköy Dr. Sadi Konuk Training and Research Hospital, Clinic of Cardiovascular Surgery, İstanbul, Türkiye

ABSTRACT

Objective: This study aims to analyze and compare the short- and long-term clinical outcomes of on-pump versus off-pump coronary artery bypass grafting (CABG) techniques in patients undergoing isolated CABG.

Methods: This study was designed as a retrospective, single-institution observational analysis, including 285 patients who underwent isolated CABG between 2010 and 2023. The participants were classified into two cohorts: on-pump (n=191) and off-pump (n=94). A comprehensive evaluation of perioperative and postoperative parameters was conducted, encompassing inflammatory markers, mechanical ventilation duration, hospital and intensive care unit (ICU) stay, and long-term outcomes such as mortality, myocardial infarction (MI), and revascularization rates.

Results: The on-pump group exhibited a significantly higher postoperative inflammatory response, with elevated white blood cell count (7.7±2.5 vs. 6.1±1.3 x10³/μL, p<0.001) and C-reactive protein levels (11.6±13.2 vs. 7.1±1.5 mg/L, p<0.001). The postoperative drainage volume was significantly higher in the on-pump group (565.2±146.6 vs. 263.8±46.2 mL, p<0.001), as were, mechanical ventilation duration (5.2±1.0 vs. 3.5±0.7 hours, p<0.001), ICU stay (2.6±1.3 vs. 2.1±0.3 days, p<0.001), and hospital length of stay (6.7±1.8 vs. 5.3±0.7 days, p<0.001). The incidence of postoperative atrial fibrillation was significantly higher in the on-pump group (8.4% vs. 2.1%, p=0.041), whereas the prevalence of peripheral artery disease (37.7% vs. 52.1%, p=0.020) and hypercholesterolemia (34.0% vs. 48.9%, p=0.015) was higher in the off-pump group. No significant differences were found in long-term mortality, MI, or revascularization rates between the groups (p>0.05).

Conclusion: While off-pump CABG was associated with lower postoperative inflammation, shorter ICU and hospital stays, and fewer early complications, both techniques demonstrated comparable long-term clinical outcomes.

Keywords: Coronary artery bypass grafting, on-pump CABG, off-pump CABG, long-term outcomes, postoperative complications

ÖZ

Amaç: Bu çalışmanın amacı, izole koroner arter baypas greftleme (KABG) uygulanan hastalarda pompalı ve pompasız KABG tekniklerinin kısa ve uzun dönem klinik sonuçlarını analiz etmek ve karşılaştırmaktır.

Gereç ve Yöntem: 2010-2023 yılları arasında izole KABG uygulanan 285 hastanın dahil edildiği, retrospektif ve tek merkezli gözlemsel bir çalışma yürütülmüştür. Hastalar pompa destekli (n=191) ve pompasız (n=94) olmak üzere iki gruba ayrılmıştır. Preoperatif, intraoperatif ve postoperatif parametreler analiz edilmiştir. Bu parametreler arasında enflamatuvar belirteçler, mekanik ventilasyon süresi, hastane ve yoğun bakım ünitesinde (YBÜ) kalış süresi ile uzun dönem sonuçlar (mortalite, miyokard enfarktüsü ve revaskülarizasyon oranları) yer almaktadır.

Bulgular: Pompa destekli grupta postoperatif enflamatuvar yanıt anlamlı derecede yüksek olup, beyaz kan hücresi sayısı (7,7±2,5 vs. 6,1±1,3 x10³/μL, p<0,001) ve C-reaktif protein seviyeleri (11,6±13,2 vs. 7,1±1,5 mg/L, p<0,001) daha yüksek bulunmuştur. Postoperatif drenaj hacmi pompa destekli grupta belirgin şekilde yüksek olup (565,2±146,6 vs. 263,8±46,2 mL, p<0,001), mekanik ventilasyon süresi (5,2±1,0 vs. 3,5±0,7 saat, p<0,001), YBÜ'de kalış süresi (2,6±1,3 vs. 2,1±0,3 gün, p<0,001) ve hastanede yatış süresi de (6,7±1,8 vs. 5,3±0,7 gün, p<0,001) daha uzun olarak saptanmıştır. Pompa destekli grupta postoperatif atriyal fibrilasyon insidansı daha yüksek bulunmuştur (%8,4 vs. %2,1, p=0,041). Öte yandan, pompasız grupta periferik arter hastalığı (%37,7 vs. %52,1, p=0,020) ve hiperkolesterolemi (%34,0 vs. %48,9, p=0,015) oranları daha yüksektir. Uzun dönem mortalite, miyokard enfarktüsü ve revaskülarizasyon oranları açısından iki grup arasında anlamlı fark saptanmamıştır (p>0,05).

Sonuç: Pompasız KABG, daha düşük postoperatif enflamasyon, daha kısa YBÜ ve hastane yatış süresi ile daha az erken komplikasyon ile ilişkilendirilmiştir. Ancak, uzun dönem klinik sonuçlar açısından her iki teknik benzer etkinlik göstermektedir.

Anahtar Kelimeler: Koroner arter baypas greftleme, pompa destekli CABG, pompasız CABG, uzun dönem sonuçlar, postoperatif komplikasyonlar

Address for Correspondence: Hasan Toz MD, University of Health Sciences Türkiye, Bakırköy Dr. Sadi Konuk Training and Research Hospital, Clinic of Cardiovascular Surgery, İstanbul, Türkiye
E-mail: tozhasan@hotmail.com **ORCID ID:** orcid.org/0000-0002-4228-6233

Received: 03.05.2025
Accepted: 31.05.2025
Publication Date: 03.09.2025

Cite as: Toz H, Kuserli Y, Türkyılmaz G, Kavala AA, Türkyılmaz S. Comparison of long-term outcomes of on-pump and off-pump techniques in isolated coronary artery bypass surgery: a cohort study. Med J Bakirkoy. 2025;21(3):296-304

INTRODUCTION

Globally, coronary artery disease (CAD) represents a common cardiovascular condition, substantially influencing both mortality and morbidity rates (1). Despite the pharmacological and interventional treatment options offered by modern medicine, coronary artery bypass grafting (CABG) continues to be the leading approach for enhancing myocardial blood flow and extending long-term survival, particularly in individuals with advanced CAD (2). CABG involves revascularization using alternative vascular grafts to bypass diseased coronary arteries, and represents one of the most frequently conducted surgical interventions across the globe (3).

Traditional CABG surgeries are performed using cardiopulmonary bypass (CPB) while the heart is arrested (4). This technique provides a more stable environment for the surgeon to perform anastomoses. However, CPB use has several disadvantages, including the induction of a systemic inflammatory response, an increased risk of neurological complications, and potential adverse effects on postoperative recovery (5). To minimize these issues, off-pump CABG (OPCABG) was developed. OPCABG is performed without the use of a CPB machine and is associated with lower complication rates, particularly in high-risk patients (6).

The advantages and disadvantages of on-pump and off-pump techniques have been debated for many years. It has been suggested that on-pump CABG (ONCABG) offers better long-term revascularization success due to its technical feasibility and higher graft patency rates (7). In contrast, OPCABG is reported to reduce complications associated with CPB, making it a safer option, particularly for elderly patients and those with comorbid systemic diseases (8). However, there is no clear consensus in the literature regarding the long-term outcomes of these two techniques.

This research sought to evaluate and contrast the postoperative and extended outcomes of on-pump versus off-pump approaches in individuals undergoing standalone CABG. By evaluating the 1-month, 1-year, and 3-year follow-up data of individuals who received CABG, with or without CPB, we analyzed the impact of both techniques on mortality, stroke (cerebrovascular accident), reintervention rates, and complications such as bleeding. The results of this investigation will offer essential insights to inform clinical decision-making and assess which method yields superior benefits concerning long-term survival and potential complications.

METHODS

Study Design and Patient Selection

This research was designed as a retrospective observational study conducted at a single center. It was conducted at the department of cardiovascular surgery at a tertiary center, including patients who received isolated CABG within the period of 2010 to 2023. The study exclusively included individuals who had undergone isolated CABG, while those who had additional cardiac interventions, including valve replacement surgery, aortic surgery, or repair of atrial septal defects were not considered within the study. This study was approved by the Non-Interventional Scientific Research Ethics Committee of the University of Health Sciences Türkiye, Bakırköy Dr. Sadi Konuk Training and Research Hospital (approval no: 2024-13-03, date: 27.11.2024). The study was conducted in accordance with the principles of the Declaration of Helsinki.

Patient Grouping and Follow-up

Participants were classified into two distinct groups according to the surgical technique applied. The first group included individuals who underwent ONCABG, in which CPB was utilized to temporarily suspend cardiac function during the procedure. The second group consisted of patients who underwent OPCABG, where the surgery was performed while the heart was still beating, eliminating the need for CPB. This approach is believed to reduce the systemic inflammatory response and may lead to improved postoperative recovery. The study population consisted of 285 patients who met the predefined inclusion criteria. To assess long-term clinical outcomes, postoperative follow-up evaluations were conducted at 1 month, 1 year, and 3 years. These follow-up periods allowed for the monitoring of potential complications, survival rates, and overall patient recovery trajectories.

Inclusion and Exclusion Criteria

For participation in the study, patients were required to fulfill specific eligibility criteria. The inclusion criteria specified that participants should be aged between 40 and 85 years, have undergone only isolated CABG surgery, and have been treated using an open surgical technique via sternotomy. Additionally, patients had to have available follow-up data for at least three years. Exclusion criteria included patients requiring emergency CABG, patients who underwent single-vessel revascularization, and patients with a history of previous cardiac surgery or requiring redo surgery. Moreover, individuals who needed additional cardiac surgical procedures or had a diagnosis of peripheral artery disease or advanced cerebrovascular

disease were excluded. These criteria were established to ensure the study was conducted on a homogeneous patient population and to allow for a direct comparison between the two surgical techniques.

Evaluated Parameters

Data covering preoperative, intraoperative, and postoperative periods were retrospectively obtained from medical records, surgical reports, and the hospital database. The demographic and clinical factors analyzed included age, sex, smoking history, and the presence of comorbid conditions such as hypertension, diabetes mellitus, and chronic obstructive pulmonary disease (COPD). Additionally, peripheral vascular disease, chronic kidney disease, and hypercholesterolemia were among the other medical conditions assessed. To evaluate preoperative functional capacity, the New York Heart Association (NYHA) classification was applied (9). Furthermore, the European System for Cardiac Operative Risk Evaluation II (EuroSCORE II) was utilized to calculate surgical risk scores for all patients, aiding in the assessment of potential postoperative outcomes (10).

The laboratory evaluations encompassed assessments of white blood cell (WBC) count, hemoglobin (Hb), hematocrit (HCT), C-reactive protein (CRP), and ejection fraction (EF) during both the preoperative and postoperative phases. Intraoperative parameters included the use of the internal mammary artery (right, left, or bilateral), cross-clamp time, CPB duration, and the number of bypass grafts performed.

Postoperative Outcomes

Early postoperative results were evaluated by examining complications that arose within the initial 30 days following surgery. Throughout this period, parameters such as the duration of mechanical ventilation, length of stay in the intensive care unit (ICU), total hospitalization period, mortality rates, and the occurrence of postoperative atrial fibrillation were documented. Furthermore, early postoperative complications, including chest drainage volume and incidences of bleeding, were analyzed. For long-term follow-up, outcomes were assessed at 1-month, 1-year, and 3-years post-surgery. Key factors such as myocardial infarction (MI), cerebrovascular events, mortality, developed dialysis dependency, and the necessity for repeat revascularization were closely monitored. The long-term survival rates and surgical success of both groups were statistically analyzed to identify any significant differences.

Surgical Techniques

In ONCABG procedures, standard ascending aortic cannulation and two-stage venous cannulation of the right atrium were performed to establish CPB. Aortic cross-clamping was applied to induce cardioplegic arrest with antegrade intermittent cold blood cardioplegia. After completing the anastomoses, protamine was administered to reverse heparinization. In OPCABG procedures, surgery was performed without the use of a CPB device. Stabilization was achieved using the Medtronic Octopus device, along with pericardial anchoring sutures to maintain hemodynamic stability. After completing the distal anastomoses, side clamping of the aorta was performed for proximal anastomoses if necessary. Both techniques were carried out by the same surgical team, and anesthesia and pharmacological management followed standardized protocols.

Statistical Analysis

The statistical evaluations were performed using IBM SPSS Statistics for Windows, Version 27.0 (IBM Corp., Armonk, NY, USA). The assumption of normality for continuous variables was assessed utilizing the Kolmogorov-Smirnov test. Continuous variables were expressed as mean±standard deviation, while categorical variables were presented as frequencies and percentages. For comparisons between two groups, the independent samples t-test was applied to analyze parametric data, whereas the Mann-Whitney U test was used for non-parametric distributions. Categorical variables were compared using the chi-square (χ^2) test, and Fisher's exact test was implemented when the expected frequency was below a predefined threshold. A p-value of less than 0.05 was considered statistically significant, and statistically significant p-values were highlighted in bold within the tables to facilitate interpretation.

RESULTS

There was no statistically significant variation between the groups concerning age, body mass index, gender distribution, smoking status, or hypertension ($p>0.05$). However, one of the parameters that exhibited a significant difference in this study was the EF. The preoperative EF was measured as 44.5±6.0% in the on-pump group and 42.3±6.7% in the off-pump group, with this difference reaching statistical significance ($p=0.007$) (Table 1).

The postoperative WBC count was recorded as 7.7±2.5 ($10^3/\mu L$) in the on-pump group, whereas it was 6.1±1.3 ($10^3/\mu L$) in the off-pump group. This difference was statistically significant ($p<0.001$). The postoperative Hb concentration was 8.3±0.4 g/dL in the on-pump group and 9.9±0.9 g/dL in the off-pump group, with significantly higher levels observed in the latter group ($p<0.001$). Similarly, the postoperative HCT level was 25.1±1.3% in the on-pump

group and 29.9±2.4% in the off-pump group, demonstrating a statistically significant elevation in the off-pump group (p<0.001). Additionally, postoperative CRP levels were 11.6±13.2 mg/L in the on-pump group and 7.1±1.5 mg/L in the off-pump group; values were significantly elevated in the on-pump group (p<0.001). No statistically significant difference was identified between the groups regarding preoperative Hb, HCT, and CRP levels (p>0.05) (Table 2).

The preoperative EF was measured at 44.5±6.0% in the on-pump group and 42.3±6.7% in the off-pump group, showing significantly higher values in the on-pump cohort (p=0.007). However, no statistically significant differences were identified between the groups in terms of postoperative EF values (p=0.381). Regarding postoperative fluid balance, the drainage volume was found to be 565.2±146.6 mL in the on-pump group and 263.8±46.2 mL in the off-pump group, with a significantly greater volume observed in the on-pump group (p<0.001). The mean duration of mechanical ventilation was 5.2±1.0 hours for the on-pump group and 3.5±0.7 hours for the off-Pump group, indicating a statistically significant prolongation in the on-pump group (p<0.001). Similarly, the ICU stay was 2.6±1.3 days for on-pump patients and 2.1±0.3 days for off-pump patients, demonstrating an extended ICU stay for the on-pump cohort (p<0.001). The reported total hospitalization duration was 6.7±1.8 days for the on-pump group and 5.3±0.7 days for the off-pump group, with significantly longer hospital stays observed in the on-pump cohort (p<0.001). These findings suggest that while ONCABG may be associated with prolonged postoperative recovery, further investigation is needed to determine its long-term clinical implications (Table 3).

Based on the EuroSCORE II evaluation, the proportion of patients classified as low-risk was 12.6% in the on-pump group and 3.2% in the off-pump group, demonstrating a statistically significant distinction (p=0.010). The incidence of postoperative atrial fibrillation was documented as 8.4% in the on-pump cohort and 2.1% in the off-pump group, with a considerably higher occurrence in the on-pump cohort (p=0.041). Similarly, the prevalence of peripheral artery disease was noted at 37.7% in the on-pump group and 52.1% in the off-pump group, with significantly higher rates observed in the off-pump cohort (p=0.020). The incidence of hypercholesterolemia was 34.0% among on-pump patients and 48.9% in the off-pump group, indicating a statistically significant increase in the off-pump cohort (p=0.015). Despite these differences, no statistically significant variation was found between the two groups concerning NYHA functional capacity classification, preoperative atrial fibrillation, diabetes mellitus, COPD, or chronic kidney disease (p>0.05).

Table 1. Comparison of demographic characteristics and risk factors between on-pump and off-pump groups

		On-pump (n=191)	Off-pump (n=94)	p-value
Age (years)		60.5±6.8	61.5±7.9	0.245[a]
BMI (kg/m²)		27.8±3.2	28.3±4.0	0.566[b]
Gender	Female Male	81 (42.4) 110 (57.6)	38 (40.4) 56 (59.6)	0.750
Smoking		115 (60.2)	61 (64.9)	0.444
Hypertension		63 (33.0)	36 (38.3)	0.376

[a]: Independent samples t-test, [b]: Mann-Whitney U test, BMI: Body mass index

Table 2. Comparison of hematological and inflammatory parameters between on-pump and off-pump groups

	On-pump (n=191)	Off-pump (n=94)	p-value
	Mean±SD	**Mean±SD**	
WBC (10³/µL)	6.4±1.8	4.7±0.9	**<0.001[b]**
Postop WBC (10³/µL)	7.7±2.5	6.1±1.3	**<0.001[b]**
Hemoglobin (g/dL)	12.5±0.9	12.7±0.9	0.286[a]
Postop hemoglobin (g/dL)	8.3±0.4	9.9±0.9	**<0.001[b]**
Hematocrit (%)	37.7±2.8	38.1±22.5	0.510[b]
Postop hematocrit (%)	25.1±1.3	29.9±2.4	**<0.001[b]**
CRP (mg/L)	4.5±2.3	3.7±1.1	**0.001[b]**
Postop CRP (mg/L)	11.6±13.2	7.1±1.5	**<0.001[b]**

[a]: Independent samples t-test, [b]: Mann-Whitney U test, WBC: White blood cell, CRP: C-reactive protein, SD: Standard deviation

These findings emphasize the potential impact of different surgical approaches on postoperative cardiovascular risk factors and suggest the need for further investigation into their long-term clinical implications (Table 4).

The early postoperative complication rate was 15.2% in the on-pump group and 4.3% in the off-pump group, with a significantly higher incidence in the on-pump group (p=0.003). The need for reoperation due to bleeding was 1.6% in the on-pump group and 0.0% in the off-pump group (p=0.007). The requirement for hemofiltration due to dialysis was 1.1% in the on-pump group and 0.0% in the off-pump group, with a significantly higher incidence in the on-pump group (p=0.043). No statistically significant differences were found between the groups in terms of other early postoperative complications, including MI, pulmonary complications, neurological complications, sternal wound complications, gastrointestinal complications, and reintubation/tracheostomy (p>0.05). No significant difference was observed between the groups in terms of hospital mortality (p=0.482) (Table 5).

In the postoperative 1-month evaluation, the incidence of new dialysis requirement was 1.0% in the on-pump group and 2.1% in the off-pump group; however, this difference was not statistically significant (p>0.05). In the postoperative 1-year and 3-year evaluations, no significant differences were observed between the groups in terms of mortality, MI, revascularization, or cardiovascular disease (p>0.05) (Table 6).

DISCUSSION

This study compared the long-term outcomes of on-pump and off-pump techniques in patients undergoing isolated CABG. The results suggest that the postoperative inflammatory response was more pronounced in the on-pump group, as evidenced by elevated WBC and CRP levels. Additionally, the on-pump group exhibited significantly

Table 3. Comparison of cardiac function, drainage volume, and clinical outcomes between on-pump and off-pump groups

	On-pump (n=191)	Off-pump (n=94)	p-value
	Mean±SD	Mean±SD	
Preop ejection fraction (%)	44.5±6.0	42.3±6.7	**0.007**[a]
Postop ejection fraction (%)	43.6±6.0	43.1±6.6	0.381[b]
Average chest drainage amount (mL)	565.2±146.6	263.8±46.2	**<0.001**[b]
Ventilation time (hours)	5.2±1.0	3.5±0.7	**<0.001**[b]
Total ICU stay (days)	2.6±1.3	2.1±0.3	**<0.001**[b]
Hospital length of stay (days)	6.7±1.8	5.3±0.7	**<0.001**[a]

[a]: Independent samples t-test, [b]: Mann-Whitney U test, SD: Standard deviation, ICU: Intensive care unit

Table 4. Comparison of functional capacity, risk scores, and comorbidities between on-pump and off-pump groups

		On-pump (n=191)	Off-pump (n=94)	p-value
		Count (%)	Count (%)	
Functional capacity	NYHA class 2	79 (41.4)	39 (41.5)	0.781
	NYHA class 3	111 (58.1)	55 (58.5)	
	NYHA class 4	1 (0.5)	0 (0.0)	
EuroSCORE II	Low score	24 (12.6)	3 (3.2)	**0.010**
	Moderate score	165 (86.4)	87 (92.6)	
	High score	2 (1.0)	4 (4.3)	
Diabetes mellitus		112 (58.6)	56 (59.6)	0.880
Preop atrial fibrillation		4 (2.1)	2 (2.1)	0.985
Postop atrial fibrillation		16 (8.4)	2 (2.1)	**0.041**
COPD		84 (44.0)	40 (42.6)	0.819
Peripheral artery disease		72 (37.7)	49 (52.1)	**0.020**
Chronic kidney disease		2 (1.0)	0 (0.0)	0.319
Hypercholesterolemia		65 (34.0)	46 (48.9)	**0.015**

NYHA: New York Heart Association, EuroSCORE II: European System for Cardiac Operative Risk Evaluation II, COPD: Chronic obstructive pulmonary disease

greater postoperative drainage volume, prolonged mechanical ventilation duration, and extended stays in both the ICU and hospital. Furthermore, the incidence of postoperative atrial fibrillation was notably higher among patients in the on-pump cohort. In contrast, peripheral artery disease and hypercholesterolemia were more prevalent in the off-pump group. Regarding long-term outcomes, no statistically significant differences were identified between the groups in terms of mortality, MI, or the necessity for revascularization. These findings indicate that while the on-pump technique may be associated with increased systemic inflammation and postoperative complications, both surgical approaches exhibit comparable effectiveness in long-term clinical outcomes. Further research with larger sample sizes and extended follow-up periods is warranted to better delineate the advantages and drawbacks of each technique.

Managing morbidity and mortality after coronary revascularization in high-risk patients continues to be a significant challenge (11,12). Various observational studies suggest that the OPCABG approach may provide a viable alternative in addressing this concern (13-15). Unlike the conventional on-pump method, OPCABG eliminates the necessity for CPB and cardioplegic arrest, thereby reducing systemic inflammatory responses and minimizing the effects of global hypoxia. This physiological benefit is believed to lower the incidence of organ-specific postoperative

Table 5. Comparison of artery usage, mortality, and early postoperative complications between on-pump and off-pump groups

		On-pump (n=191)	Off-pump (n=94)	p-value
		Count (%)	Count (%)	
Internal mammary artery usage	Right	4 (2.1)	0 (0.0)	0.078
	Left	181 (94.8)	94 (100.0)	
	Bilateral	6 (3.1)	0 (0.0)	
Hospital mortality		1 (0.5)	0 (0.0)	0.482
Early postop complications		29 (15.2)	4 (4.3)	**0.003**
Reoperation for bleeding		3 (1.6)	0 (0.0)	**0.007**
MI/reintervention		1 (0.5)	0 (0.0)	0.222
Pulmonary complications		1 (0.5)	0 (0.0)	0.482
Neurological complications		16 (8.4)	3 (3.2)	0.462
Sternal wound complications		8 (4.2)	0 (0.0)	0.097
Hemofiltration for dialysis		2 (1.1)	0 (0.0)	**0.043**
GIS complications		7 (3.7)	0 (0.0)	0.060
Reintubation/tracheostomy		1 (0.5)	0 (0.0)	0.482

GIS: Gastrointestinal, MI: Myocardial infarction

Table 6. Comparison of postoperative outcomes at 1-month, 1-year, and 3-years between on-pump and off-pump groups

		On-pump (n=191)	Off-pump (n=94)	p-value
		Count (%)	Count (%)	
Postop 1-month evaluation	No	181 (94.8)	87 (92.6)	0.926
	Mortality	1 (0.5)	1 (1.1)	
	MI+revascularisation	2 (1.0)	1 (1.1)	
	CVD	5 (2.6)	3 (3.2)	
	New dialysis	2 (1.0)	2 (2.1)	
Postop 1-year evaluation	No	189 (99.5)	93 (98.9)	0.610
	CVD	1 (0.5)	1 (1.1)	
Postop 3-year evaluation	No	185 (97.4)	93 (98.9)	0.522
	Mortality	1 (0.5)	0 (0.0)	
	MI+revascularisation	1 (0.5)	1 (1.1)	
	CVD	3 (1.6)	0 (0.0)	

MI: Myocardial infarction, CVD: Cardiovascular disease

complications, which are more frequently observed in high-risk populations (16,17). Additionally, OPCABG has been associated with a reduction in postoperative morbidities, including shorter ventilation times, a decreased likelihood of atrial fibrillation, reduced transfusion requirements, and a lower risk of stroke, renal impairment, and prolonged ICU stays (18). Despite these advantages, large-scale randomized trials such as the ROOBY trial (19) and the CORONARY trial (20) have not demonstrated a significant superiority of OPCABG over conventional ONCABG in the general patient population. Furthermore, the BEST bypass surgery trial (21), which focused on high-risk individuals with a EuroSCORE II of 5, found no significant difference in morbidity or mortality between OPCABG and ONCABG at the 30-day follow-up. A study by Dhurandhar et al. (22) reported that the off-pump technique was associated with lower postoperative morbidity, reduced atrial fibrillation rates, and decreased transfusion needs. However, it did not improve long-term mortality outcomes compared to those of the on-pump approach (22). Similarly, findings from our study revealed a higher incidence of postoperative atrial fibrillation and a more intense inflammatory response in the on-pump group. In contrast to Dhurandhar et al.'s (22) findings, our study identified prolonged ICU and hospital stays among on-pump patients. Regarding long-term survival, neither study detected a significant difference suggesting that despite the early postoperative benefits of off-pump surgery its long-term clinical efficacy remains comparable to that of the on-pump technique. Further research with extended follow-up periods and larger sample sizes is warranted to establish a more definitive comparison between these two surgical approaches.

Previous meta-analyses and large-scale propensity-matched observational studies have reported a significant increase in long-term mortality associated with OPCABG (23-27). This observation aligns with the higher mid-term coronary reintervention rates linked to off-pump procedures. Consequently, the rise in long-term mortality incidence appears to be a plausible outcome. However, data from the 10-year follow-up suggest a more nuanced perspective. Findings from the ROOBY trial did not reveal a statistically significant elevation in long-term mortality for patients undergoing OPCABG compared to those receiving ONCABG (28). These results contrast with the 5-year follow-up data, which demonstrated a substantial increase in long-term mortality among the OPCABG cohort (29). A study conducted by He et al. (30) suggested that OPCABG may lower the short-term risk of stroke. However, it was also associated with an increased need for revascularization and a potential rise in long-term mortality rates (30). In our study, the off-pump group exhibited a reduced postoperative inflammatory response, shorter hospital stays, and fewer early postoperative complications. Additionally, no statistically significant difference was identified between the groups regarding the necessity for revascularization necessitated by MI or repeat intervention. These results indicate that contrary to the findings of He et al. (30), off-pump surgery did not contribute to an increased requirement for long-term revascularization. Further investigations incorporating a larger sample size and extended follow-up periods are necessary to refine these conclusions and determine the optimal surgical strategy for long-term patient outcomes.

Study Limitations

This study has certain limitations. Patients were monitored at 1-month, 1-year, and 3-years postoperatively. Extending the follow-up duration could have facilitated a more comprehensive assessment, particularly concerning the necessity for revascularization and long-term mortality outcomes. Although revascularization rates were documented, angiographic follow-up was not conducted, potentially leading to gaps in data regarding graft patency and long-term myocardial perfusion. Furthermore, while all surgical procedures were performed by the same team, variations in individual surgeon and anesthesia practices may have influenced the findings. Consequently, the impact of off-pump surgery performed by highly experienced surgical teams could not be fully evaluated. This study primarily concentrated on clinical outcomes, including mortality, MI, revascularization, and renal failure. However, patient-centered factors such as quality of life, neurocognitive function, and functional recovery were not considered, which may limit the generalizability of the results. Future research incorporating these additional parameters could provide a more holistic understanding of postoperative outcomes.

CONCLUSION

In this study, the early and long-term outcomes of isolated CABG operations performed using on-pump and off-pump techniques were compared. The on-pump technique was associated with a higher postoperative inflammatory response, increased drainage volume, prolonged mechanical ventilation, and longer ICU and hospital stays. In contrast, the off-pump technique provided advantages such as lower postoperative complication rates and shorter hospital stays. Long-term follow-up revealed no significant differences between the two techniques regarding mortality, MI, and revascularization. These findings suggest, that while

the off-pump technique may offer advantages in the early postoperative period, it demonstrates similar efficacy to the on-pump technique in long-term clinical outcomes. Further validation of these findings requires larger sample sizes, multicenter studies, and prospective research.

ETHICS

Ethics Committee Approval: This study was approved by the Non-Interventional Scientific Research Ethics Committee of the University of Health Sciences Türkiye, Bakırköy Dr. Sadi Konuk Training and Research Hospital (approval no: 2024-13-03, date: 27.11.2024).

Informed Consent: Retrospective study.

FOOTNOTES

Authorship Contributions

Concept: H.T., G.T., Design: Y.K., A.A.K., Data Collection or Processing: H.T., S.T., Analysis or Interpretation: G.T., Literature Search: Y.K., S.T., Writing: H.T., A.A.K.

Conflict of Interest: No conflict of interest was declared by the authors.

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

REFERENCES

1. Stone PH, Libby P, Boden WE. Fundamental pathobiology of coronary atherosclerosis and clinical implications for chronic ischemic heart disease management-the plaque hypothesis: a narrative review. JAMA Cardiol. 2023;8:192-201.
2. Olsen DB, Pedersen PU, Noergaard MW. Prehabilitation before elective coronary artery bypass grafting surgery: a scoping review. JBI Evid Synth. 2023;21:1190-242.
3. Sef D, Raja SG. Bilateral internal thoracic artery use in coronary artery bypass grafting in the post-ART era - perspective. Int J Surg. 2021;86:1-4.
4. Alvarez JM. Off-pump coronary bypass surgery: time for the myths to face reality. Heart Lung Circ. 2023;32:284-6.
5. Shirasaka T, Arayawudhikul N, Tantraworasin A, Chartrungsan A, Sakboon B, Cheewinmethasiri J, et al. Successful introduction of off-pump coronary artery bypass grafting in Southeastern Asian countries: a single center's experience in Thailand. Surg Open Sci. 2022;8:27-32.
6. Kilic Y, Jalalzai I, Sonmez E, Erkut B. On pump beating heart coronary artery surgery in patients requiring urgent revascularization. Heart Surg Forum. 2023;26:E808-e16.
7. Shaefi S, Mittel A, Loberman D, Ramakrishna H. Off-pump versus on-pump coronary artery bypass grafting-a systematic review and analysis of clinical outcomes. J Cardiothorac Vasc Anesth. 2019;33:232-44.
8. Ren Q, Li G, Chu T, Liu Q, Huang Y, Liu K, et al. Off-pump versus on-pump coronary artery bypass grafting in diabetic patients: a meta-analysis of observational studies with a propensity-score analysis. Cardiovasc Drugs Ther. 2024.
9. Caraballo C, Desai NR, Mulder H, Alhanti B, Wilson FP, Fiuzat M, et al. Clinical implications of the new york heart association classification. J Am Heart Assoc. 2019;8:e014240.
10. Atashi A, Amini S, Tashnizi MA, Moeinipour AA, Aazami MH, Tohidnezhad F, et al. External validation of European System for Cardiac Operative Risk Evaluation II (EuroSCORE II) for risk prioritization in an Iranian population. Braz J Cardiovasc Surg. 2018;33:40-6.
11. Montague NT 3rd, Kouchoukos NT, Wilson TA, Bennett AL 3rd, Knott HW, Lochridge SK, et al. Morbidity and mortality of coronary bypass grafting in patients 70 years of age and older. Ann Thorac Surg. 1985;39:552-7.
12. Mullany CJ, Darling GE, Pluth JR, Orszulak TA, Schaff HV, Ilstrup DM, et al. Early and late results after isolated coronary artery bypass surgery in 159 patients aged 80 years and older. Circulation. 1990;82(5 Suppl):Iv229-36.
13. Cleveland JC Jr, Shroyer AL, Chen AY, Peterson E, Grover FL. Off-pump coronary artery bypass grafting decreases risk-adjusted mortality and morbidity. Ann Thorac Surg. 2001;72:1282-8; discussion 1288-9.
14. Plomondon ME, Cleveland JC Jr, Ludwig ST, Grunwald GK, Kiefe CI, Grover FL, et al. Off-pump coronary artery bypass is associated with improved risk-adjusted outcomes. Ann Thorac Surg. 2001;72:114-9.
15. Calafiore AM, Di Mauro M, Canosa C, Di Giammarco G, Iaco AL, Contini M. Early and late outcome of myocardial revascularization with and without cardiopulmonary bypass in high risk patients (EuroSCORE > or = 6). Eur J Cardiothorac Surg. 2003;23:360-7.
16. Horneffer PJ, Gardner TJ, Manolio TA, Hoff SJ, Rykiel MF, Pearson TA, et al. The effects of age on outcome after coronary bypass surgery. Circulation. 1987;76:V6-12.
17. Cernaianu AC, Vassilidze TV, Flum DR, Maurer M, Cilley JH Jr, Grosso MA, et al. Predictors of stroke after cardiac surgery. J Card Surg. 1995;10:334-9.
18. Rich MW, Sandza JG, Kleiger RE, Connors JP. Cardiac operations in patients over 80 years of age. J Thorac Cardiovasc Surg. 1985;90:56-60.
19. Shroyer AL, Grover FL, Hattler B, Collins JF, McDonald GO, Kozora E, et al. On-pump versus off-pump coronary-artery bypass surgery. N Engl J Med. 2009;361:1827-37.
20. Lamy A, Devereaux PJ, Prabhakaran D, Taggart DP, Hu S, Paolasso E, et al. Off-pump or on-pump coronary-artery bypass grafting at 30 days. N Engl J Med. 2012;366:1489-97.
21. Møller CH, Perko MJ, Lund JT, Andersen LW, Kelbaek H, Madsen JK, et al. No major differences in 30-day outcomes in high-risk patients randomized to off-pump versus on-pump coronary bypass surgery: the best bypass surgery trial. Circulation. 2010;121:498-504.
22. Dhurandhar V, Saxena A, Parikh R, Vallely MP, Wilson MK, Butcher JK, et al. Outcomes of on-pump versus off-pump coronary artery bypass graft surgery in the high Risk (AusSCORE > 5). Heart Lung Circ. 2015;24:1216-24.
23. Filardo G, Hamman BL, da Graca B, Sass DM, Machala NJ, Ismail S, et al. Efficacy and effectiveness of on- versus off-pump coronary artery bypass grafting: a meta-analysis of mortality and survival. J Thorac Cardiovasc Surg. 2018;155:172-9.e5.
24. Smart NA, Dieberg G, King N. Long-term outcomes of on- versus off-pump coronary artery bypass grafting. J Am Coll Cardiol. 2018;71:983-91.
25. Park SJ, Jo AJ, Kim HJ, Cho S, Ko MJ, Yun SC, et al. Real-world outcomes of on- vs off-pump coronary bypass surgery: result from korean nationwide cohort. Ann Thorac Surg. 2022;113:1989-98.
26. Squiers JJ, Schaffer JM, Banwait JK, Ryan WH, Mack MJ, DiMaio JM. Long-term survival after on-pump and off-pump coronary artery bypass grafting. Ann Thorac Surg. 2022;113:1943-52.

27. Chikwe J, Lee T, Itagaki S, Adams DH, Egorova NN. Long-term outcomes after off-pump versus on-pump coronary artery bypass grafting by experienced surgeons. J Am Coll Cardiol. 2018;72:1478-86.

28. Quin JA, Wagner TH, Hattler B, Carr BM, Collins J, Almassi GH, et al. Ten-year outcomes of off-pump vs on-pump coronary artery bypass grafting in the department of veterans affairs: a randomized clinical trial. JAMA Surg. 2022;157:303-10.

29. Shroyer AL, Hattler B, Wagner TH, Collins JF, Baltz JH, Quin JA, et al. Five-year outcomes after on-pump and off-pump coronary-artery bypass. N Engl J Med. 2017;377:623-32.

30. He L, Tiemuerniyazi X, Chen L, Yang Z, Huang S, Nan Y, et al. Clinical outcomes of on-pump versus off-pump coronary-artery bypass surgery: a meta-analysis. Int J Surg. 2024;110:5063-70.

Medical Journal of Bakırköy

Med J Bakirkoy 2025;21(3):305-309
DOI: 10.4274/BMJ.galenos.2025.2024.12-7

Research

Efficacy and Safety Analysis of Four Minimally Invasive Surgical Techniques in Carpal Tunnel Syndrome Treatment: A Comprehensive Cohort Study of 80 Patients

Karpal Tünel Sendromu Tedavisinde Dört Minimal İnvaziv Cerrahi Yöntemin Etkinlik ve Güvenilirlik Analizi: 80 Hastalık Kapsamlı Bir Kohort Çalışması

Kaan Tuğberk Özdemir, Hasan Armağan Uysal

Medicana International İzmir Hospital, Clinic of Neurosurgery, İzmir, Türkiye

ABSTRACT

Objective: The most prevalent peripheral neuropathy, carpal tunnel syndrome (CTS), may drastically impair a patient's function and quality of life. In an attempt to effectively alleviate symptoms with fewer adverse consequences, several less invasive surgical techniques have been created. However, there is a dearth of comparative data on the security and effectiveness of different approaches.

Methods: In a retrospective cohort study, we looked at 80 individuals who had been diagnosed with CTS. Percutaneous carpal tunnel release (PCTR), ultrasound-guided CTR (UGCTR), endoscopic CTR (ECTR), and open CTR (OCTR) are the four minimally invasive surgical procedures that we used. The outcomes were measured using clinical and electrophysiological assessments in addition to patient-reported health-related quality of life. The initial plan consisted of preoperative and postoperative checkups at one, three, six, and twelve months postoperative.

Results: All four methods were shown to be secure and successful in treating CTS, and they improved electrophysiological parameters, function, and symptoms. Nonetheless, differences in the incidence of complications and improvements in quality of life were seen. While OCTR had the lowest possibility of troubles, both ECTR and UGCTR achieved transactional recovery and shorter operation times. PCTR had a higher rate of nerve damage and recurrence, but it was also associated with a shorter hospital stay and lower expenses than the other approaches.

Conclusion: Our investigation confirmed the safety and efficacy of these four minimally invasive surgical techniques for the treatment of CTS. However, the process must be tailored to each approach's particular advantages and potential risks. Future research with larger sample numbers and longer follow-up trials may provide a better explanation for personalized CTS therapy.

Keywords: Open carpal tunnel release, endoscopic carpal tunnel release, ultrasound-guided carpal tunnel release, percutaneous carpal tunnel release, carpal tunnel syndrome, minimally invasive surgical techniques

ÖZ

Amaç: En yaygın periferik nöropati olan karpal tünel sendromu (KTS), bir hastanın işlevini ve yaşam kalitesini önemli ölçüde bozabilir. Daha az olumsuz sonuçla semptomları etkili bir şekilde tedavi etme girişimi olarak, birkaç daha az invaziv cerrahi teknik yaratılmıştır. Ancak, farklı yaklaşımların güvenliği ve etkinliği hakkında karşılaştırmalı veri eksikliği vardır.

Gereç ve Yöntem: Retrospektif bir kohort araştırmasında, KTS tanısı konmuş 80 kişiye bakıldı. Perkütan karpal tünel serbestleştirme (PKTS), ultrason kılavuzluğunda KTS (UGKTS), endoskopik KTS (EKTS) ve açık KTS (OKTS), kullandığımız dört minimal invaziv cerrahi prosedürdür. Hastanın bildirdiği sağlık ile ilgili yaşam kalitesine ek olarak klinik ve elektrofizyolojik değerlendirmeler kullanılarak ölçüldü. İlk plan, ameliyattan bir, üç, altı ve on iki ay sonra ameliyat öncesi ve sonrası kontrollerden oluşuyordu.

Bulgular: Dört yöntemin de KTS tedavisinde güvenli ve başarılı olduğu ve elektrofizyolojik parametreleri, işlevi ve semptomları iyileştirdiği gösterildi. Bununla birlikte, komplikasyon sıklığında farklılıklar ve yaşam kalitesinde iyileşmeler görüldü. OKTS'nin sorun olasılığı en düşükken, EKTS ve UGKTS işlemsel iyileşme ve daha kısa operasyon süreleri elde etti. PKTS'nin daha yüksek bir sinir hasarı ve tekrarlama oranı vardı, ancak aynı zamanda diğer yaklaşımlara göre daha kısa hastanede kalma süresi ve daha düşük masraflarla da ilişkilendirildi.

Sonuç: Araştırmamız, bu dört minimal invaziv cerrahi tekniğinin KTS tedavisi için güvenliğini ve etkinliğini doğruladı. Ancak, süreç her yaklaşımın özel avantajlarına ve potansiyel risklerine göre uyarlanmalıdır. Daha büyük örnek sayıları ve daha uzun takip denemeleri içeren gelecekteki araştırmalar, kişiselleştirilmiş KTS tedavisi için daha iyi bir açıklama sağlayabilir.

Anahtar Kelimeler: Açık karpal tünel, endoskopik karpal tünel, ultrason rehberliğinde karpal tünel, perkütan karpal tünel, karpal tünel sendromu, minimal invaziv cerrahi teknikleri

Address for Correspondence: Kaan Tuğberk Özdemir, MD, Medicana International İzmir Hospital, Clinic of Neurosurgery, İzmir, Türkiye
E-mail: tttugberk@gmail.com **ORCID ID:** orcid.org/0000-0001-8335-8689

Received: 31.01.2025
Accepted: 27.07.2025
Publication Date: 03.09.2025

Cite as: Özdemir KT, Uysal HA. Efficacy and safety analysis of four minimally invasive surgical techniques in carpal tunnel syndrome treatment: a comprehensive cohort study of 80 patients. Med J Bakirkoy. 2025;21(3):305-309

INTRODUCTION

Compression of the median nerve in the carpal tunnel generates carpal tunnel syndrome (CTS), the most common peripheral nerve entrapment neuropathy (1). One hand and finger numbness, discomfort, and weakness are classic signs of CTS (2). CTS is more frequent in women than in men, occurring at an overall frequency of 3-5% in the population (3). Conservative and surgical approaches are the two categories into which the three CTS treatment modalities may be separated (4). Minimally invasive surgical methods have gained popularity in recent years for addressing CTS (5). Shorter hospital stays, quicker recovery, and less postoperative pain are just a few advantages of these methods. Four minimally invasive surgical techniques: open carpal tunnel release (OCTR), endoscopic CTR (ECTR), ultrasound-guided CTR (UGCTR), and percutaneous CTR (PCTR) were assessed for efficacy and safety in this research (6). Over the years, OCTR has been the standard therapy for CTS (7). This therapy has a lower risk of problems than minimally invasive surgical techniques. ECTR is an efficient minimally invasive treatment with the benefits of a quicker functional recovery and a shorter operation time (8). The minimally invasive procedure known as UGCTR has gained popularity recently (9,10). Despite being less expensive and needing fewer hospital stays, PCTR has a higher risk of causing nerve damage and recurrence (11). The aim of this research is to compare the effectiveness and safety of the four different kinds of minimally invasive surgical techniques. Eighty people are expected to benefit from these therapies in terms of better function, reduced symptoms, and enhanced test scores. This study attempts to determine the distinctions among various strategies in relation to issues affecting quality of life. The findings of the study, which take into consideration both the positive and negative aspects of each potential method, may aid in choosing the best course of action and provide evidence for the efficacy and safety of minimally invasive surgical methods in the management of CTS. Future research with larger sample sizes and longer follow-up times may also encourage the use of these tactics and providing tailored treatment.

METHODS

Study Population and Selection Criteria

A retrospective cohort study examined data on 80 patients involving minimally invasive surgery from 2021 to 2023. The research participants were adults with electrophysiologically confirmed CTS symptoms (12). Twelve patients have to meet the following criteria in order to be eligible for study enrollment:
•Patients with CTS who are 18 years of age or older and have not improved with conservative measures.
•The diagnosis of CTS was confirmed by electrophysiological investigations. No prior CTS operations were carried out.
•Four minimally invasive surgical techniques were performed on the patients: OCTR, ECTR, UGCTR, and PCTR.

•Both the surgeons' own preferences and the specific medical needs of the patients were considered throughout the procedure selection process.

Evaluation Criteria

Patients underwent testing in the first, third, sixth, and twelfth months, both before and after the surgery, to monitor controls. Improvements in electrophysiological indicators, functionality, and symptom relief were assessed at each follow-up. In addition, they documented problems and evaluations of quality of life for each method.

Surgical Methods

The transverse carpal ligament is divided to open the carpal tunnel using the OCTR technique (13). A local anesthetic was used during this procedure. A standardized mixture of 10 mL of 2% lidocaine hydrochloride and 5 mL of 0.5% bupivacaine hydrochloride was administered to each patient. Patients underwent procedures under local anesthesia and were admitted to the hospital for a short postoperative stay.
•**Endoscopic release:** The transverse carpal ligament is sectioned using an endoscope (14). The patients had a short recuperation time after the procedure, which was performed under local anesthesia.
•**UGCTR:** The transverse carpal ligament is severed using ultrasound imaging (15). Following the surgery, which was carried out under local anesthesia, patients had a short period of postoperative recovery while in the hospital.
•**PCTR:** The transverse carpal ligament is transected using a needle and a small incision (16). The patients were given local anesthesia during the procedure, and they were admitted to the hospital briefly after the procedure.

Statistical Analysis

The SPSS statistical software package (IBM Corp., Armonk, NY, USA) was used for data analysis. The mean and standard deviation were used to describe continuous variables, whereas frequency and percentages were used to describe categorical variables. The independent samples t-test and chi-square test were used to evaluate group differences. The threshold for statistical significance was set at $p<0.05$.

Ethical Approval

Approval was obtained from the İzmir Bakırçay University Non-Interventional Clinical Research Ethics Committee and research permission was obtained from the institution where the study was conducted (approval number: 1004, date: 26.04.2023). The research was conducted in accordance with the Declaration of Helsinki.

RESULTS

To treat CTS, this research compares the efficacy of four minimally invasive surgical techniques. Eighty patients' data were evaluated for the research. Methods: four patient groups were established: twenty patients underwent PCTR, had ECTR, twenty underwent UGCTR, and underwent OCTR (Table 1). Evaluations before and after the surgery showed that all four surgical procedures resulted in excellent clinical and functional improvement. In the first, third, sixth, and 12 months after the procedure, improvements in electrophysiological parameters were seen across all techniques (17). There was no significant difference between the groups ($p>0.05$) (Table 2). The statistical findings showed that the OCTR technique, ECTR technique, UGCTR technique, and PCTR technique did not vary in terms of efficacy or safety ($p>0.05$) (Table 3) (18). There was no statistically significant difference in the number of complications across the four surgical techniques ($p>0.05$) (Table 4) (19). Additionally, each of these methods improved the quality of life for patients when included in quality of life evaluations. In terms of recovery timeframes, endoscopic and UGCTR procedures aided patients in recovering more rapidly, but open and percutaneous approaches had somewhat longer recovery times (20).

Table 1. Post-procedure symptom relief and functional recovery rates

Method	1. month	3. month	6. month	12. month
Open	85%	90%	95%	100%
Endoscopic	80%	85%	95%	100%
Ultrasound-guided	75%	90%	95%	100%
Percutaneous	80%	85%	90%	95%

Table 2. Change of electrophysiological parameters

Method	1. month	3. month	6. month	12. month
Open	85%	90%	95%	100%
Endoscopic	80%	85%	95%	100%
Ultrasound-guided	75%	90%	95%	100%
Percutaneous	80%	85%	90%	95%

DISCUSSION

The intent of this large-scale group investigation is to compare four different minimally invasive surgical procedures for CTS, and their potential side effects and safety for individual patients.

The results of the research show that OCTR, ECTR, UGCTR, and PCTR all have similar success and complication risk rates. These results are consistent with earlier studies reported in recent literature (21,22).

The most popular and conventional surgical method for treating CTS is OCTR. This approach is still regarded as the gold standard due to its excellent success rates and little risk of complications (23). Alternative minimally invasive surgical techniques have been developed as a result of lesser scars and the open method's longer recovery time (24).

The endoscopic approach to CTR gained popularity since it was more comfortable after surgery, and took less time to recuperate (25). Interestingly, the endoscopic technique used in this investigation had success rates and complication risks comparable to other techniques. Patients may return to work sooner because of the endoscopic procedure's improved recovery time and cosmetic outcomes (26).

In recent years, UGCTR has become a non-invasive treatment technique. Because of the high-quality images and real-time imaging capabilities, the surgeon can perform the surgery with more assurance (27). In this study, the endoscopic method and the ultrasound-guided procedure had similar success rates and complication risks.

One of the newest minimally invasive methods for treating CTS is PCTR. This approach has been linked to a greater risk of problems even though it reduces tissue damage and speeds up recovery (28).

Table 3. Complication rates

Method	Complication rate
Open	5%
Endoscopic	10%
Ultrasound-guided	5%
Percutaneous	15%

Table 4. Recovery times

Method	Average recovery time
Open	6 weeks
Endoscopic	3 weeks
Ultrasound-guided	3 weeks
Percutaneous	5 weeks

In this experiment, those who received percutaneous therapy had greater issues than those who received other modalities, despite similar success rates. However, by using this approach more skillfully and with more experience, the risks might be decreased (29).

Study Limitations

The study's limitations were a comparatively small cohort and the possible impact of past experience with various surgical procedures on success and complication rates. Furthermore, since the research is retrospective rather than randomized, the findings may not be significant. Larger sample sizes and randomized controlled designs will enable future research to more precisely evaluate the safety and effectiveness of such minimally invasive and surgical procedures.

Four minimally invasive modification techniques are safe and effective for treating CTS, according to the present study's findings. A successful treatment plan should be selected by weighing the benefits and drawbacks of each method, taking patient selection and surgical expertise into consideration. Further research in this area might lead to improved outcomes in the management of CTS and the creation of less invasive surgical methods.

CONCLUSION

The safety and efficacy of four minimally invasive surgical therapy modalities for CTS: OCTR, ECTR, UGCTR, and PCTR, are being assessed in this systematic cohort study. The research found that both strategies had comparable success rates and complication risks. Every surgical technique has pros and cons, and the field of surgery is always evolving. Because of this, it is essential to choose the most effective treatment option based on the surgeon's expertise and the right patient selection. This research and the corpus of current medical literature support the safety and effectiveness of minimally invasive surgical techniques in the treatment of CTS. Additional studies using larger sample sizes and randomized controlled-techniques will be needed to evaluate the relative efficacy and safety of different modalities, and to provide additional insight into this subject. Patients' quality of life may be enhanced, and productivity losses may be reduced with such treatment.

ETHICS

Ethics Committee Approval: Approval was obtained from the İzmir Bakırçay University Non-Interventional Clinical Research Ethics Committee and research permission was obtained from the institution where the study was conducted (approval number: 1004, date: 26.04.2023).

Informed Consent: Retrospective study.

FOOTNOTES

Authorship Contributions

Concept: H.A.U., Design: H.A.U., Data Collection or Processing: K.T.Ö., Analysis or Interpretation: K.T.Ö., H.A.U., Literature Search: K.T.Ö., H.A.U., Writing: K.T.Ö.

Conflict of Interest: No conflict of interest was declared by the authors.

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

REFERENCES

1. Atroshi I, Gummesson C, Johnsson R, Ornstein E, Ranstam J, Rosén I. Prevalence of carpal tunnel syndrome in a general population. JAMA. 1999;282:153-8.
2. Padua L, Coraci D, Erra C, Pazzaglia C, Paolasso I, Loreti C, et al. Carpal tunnel syndrome: clinical features, diagnosis, and management. Lancet Neurol. 2016;15:1273-84.
3. Bland JD. Carpal tunnel syndrome. BMJ. 2007;335:343-6.
4. Huisstede BM, Randsdorp MS, Coert JH, Glerum S, van Middelkoop M, Koes BW. Carpal tunnel syndrome. part I: effectiveness of nonsurgical treatments - a systematic review. Arch Phys Med Rehabil. 2010;91:981-1004.
5. Vasiliadis HS, Xenakis TA, Mitsionis G, Paschos NK, Georgoulas P, Papadopoulos G. Endoscopic versus open carpal tunnel release. Arthroscopy. 2014;30:159-64.
6. Chaise F, Bellemère P. Ultrasound-guided surgery for carpal tunnel syndrome. Chir Main. 2015;34:49-55.
7. Agee JM, McCarroll HR Jr, Tortosa RD, Berry DA, Szabo RM, Peimer CA. Endoscopic release of the carpal tunnel: a randomized prospective multicenter study. J Hand Surg Am. 1992;17:987-95.
8. Gerritsen AA, Uitdehaag BM, van Geldere D, Scholten RJ, de Vet HC, Bouter LM. Systematic review of randomized clinical trials of surgical treatment for carpal tunnel syndrome. Br J Surg. 2001;88:1285-95.
9. Erdmann MW, van Beek AL, Thornton SJ, Barnard AR. A comparison of open and endoscopic carpal tunnel release. Hand Clin. 1995;11:1-6.
10. Nakamichi K, Tachibana S. Ultrasonographically guided percutaneous carpal tunnel release. J Hand Surg Am. 1998;23:1006-12.
11. Kuschner SH, Lane CS. Endoscopic versus open carpal tunnel release: big deal or much ado about nothing? Am J Orthop (Belle Mead NJ). 1997;26:591-6.
12. Phalen GS. The carpal-tunnel syndrome. Seventeen years' experience in diagnosis and treatment of six hundred fifty-four hands. J Bone Joint Surg Am. 1966;48:211-28.
13. Cobb TK, Dalley BK, Posteraro RH, Lewis RC. Anatomy of the flexor retinaculum. J Hand Surg Am. 1993;18:91-9.
14. Agee JM, Peimer CA, Pyrek JD, Walsh WE. Endoscopic carpal tunnel release: a prospective study of complications and surgical experience. J Hand Surg Am. 1995;20:165-71.
15. Nakamichi K, Tachibana S. Percutaneous carpal tunnel release compared with mini-open release using ultrasonographic guidance for both techniques. J Hand Surg Am. 2010;35:437-45.

16. Chen L, Duan X, Huang X, Lv J, Peng K, Xiang Z. Effectiveness and safety of endoscopic versus open carpal tunnel decompression. Arch Orthop Trauma Surg. 2014;134:585-93.
17. Benson LS, Bare AA, Nagle DJ, Harder VS, Williams CS, Visotsky JL. Complications of endoscopic and open carpal tunnel release. Arthroscopy. 2006;22:919-24, 924.e1-2.
18. Lee Y, Youn H, Shin S, Chung Y. Minimally invasive carpal tunnel release using a hook knife through a small transverse carpal incision: technique and outcome. Clin Orthop Surg. 2023;15:318-26.
19. Trumble TE, Diao E, Abrams RA, Gilbert-Anderson MM. Single-portal endoscopic carpal tunnel release compared with open release: a prospective, randomized trial. J Bone Joint Surg Am. 2002;84:1107-15.
20. Chung KC, Burns PB, Kim HM, Burke FD, Wilgis EF, Fox DA. Long-term follow-up for recurrent carpal tunnel syndrome after open carpal tunnel release in a multicenter study. Plast Reconstr Surg. 2013;131:225-34.
21. Nakamichi K, Tachibana S. Ultrasonographically guided percutaneous carpal tunnel release. J Hand Surg Br. 2007;32:372-5.
22. Gurpinar T, Polat B, Polat AE, Carkci E, Kalyenci AS, Ozturkmen Y. Comparison of open and endoscopic carpal tunnel surgery regarding clinical outcomes, complication and return to daily life: a prospective comparative study. Pak J Med Sci. 2019;35:1532-7.
23. Chiu DT, Janecka IP, Krizek TJ, Wolff M, Lovelace RE. Autogenous venous nerve conduit: a preliminary report. J Hand Surg Am. 1982;7:388-93.
24. Yadav RL, Jain GK, Kumar S. A comparative study of open carpal tunnel release and percutaneous carpal tunnel release in patients with carpal tunnel syndrome. J Clin Orthop Trauma. 2015;6:1-5.
25. Chow JC. Endoscopic release of the carpal ligament for carpal tunnel syndrome: long-term results using the Chow technique. Arthroscopy. 1999;15:417-21.
26. Vasiliadis HS, Xenakis TA, Mitsionis G, Paschos N, Georgoulis A. Endoscopic versus open carpal tunnel release: a long-term comparative study. Int Orthop. 2010;34:1025-8.
27. Rotaru-Zavaleanu AD, Lungulescu CV, Bunescu MG, Vasile RC, Gheorman V, Gresita A, et al. Occupational carpal tunnel syndrome: a scoping review of causes, mechanisms, diagnosis, and intervention strategies. Front Public Health. 2024;12:1407302.
28. Palmer DH, Hanrahan LP. Social and economic costs of carpal tunnel surgery. Instr Course Lect. 1995;44:167-72.
29. Lane JCE, Craig RS, Rees JL, Gardiner MD, Green J, Prieto-Alhambra D, et al. Serious postoperative complications and reoperation after carpal tunnel decompression surgery in England: a nationwide cohort analysis. Lancet Rheumatol. 2020;3:e49-e57.

Medical Journal of Bakırköy

Med J Bakirkoy 2025;21(3):310-315
DOI: 10.4274/BMJ.galenos.2025.2025.5-9

Research

Evaluation of ChatGPT-4o's Responses to Questions about Myasthenia Gravis in English and Turkish

ChatGPT-4o'nun Miyastenia Gravis Hakkındaki İngilizce ve Türkçe Sorulara Verdiği Yanıtların Değerlendirilmesi

Berin İnan, Ömer Karadaş, Zeki Odabaşı

University of Health Sciences Türkiye, Gülhane Faculty of Medicine, Department of Neurology, Ankara, Türkiye

ABSTRACT

Objective: Large language models, such as Chat Generative Pre-Trained Transformer 4o (ChatGPT-4o), are increasingly used by both patients and medical professionals to access health-related information. Myasthenia gravis (MG) is a chronic autoimmune neuromuscular disorder requiring long-term treatment. Therefore, timely access to accurate medical information about MG is important. This study aimed to evaluate the accuracy, completeness, clarity, appropriateness for the target audience, risk of misinformation or harm, and readability of ChatGPT-4o-generated responses to queries about MG from patients and neurology residents, in both English and Turkish.

Methods: We developed four sets of 20 questions, frequently asked by patients and neurology residents about MG in both English and Turkish, covering pathophysiology and symptoms, diagnosis, treatment, prognosis, and daily management. ChatGPT-4o responses were generated in separate sessions on March 29, 2025. Two neurologists independently evaluated the responses using a 5-point Likert scale across five domains. Readability was assessed using the Flesch Reading Ease score, Flesch-Kincaid grade level, and Gunning-Fog index for English, and the Ateşman readability index for Turkish.

Results: Scores for accuracy, clarity, appropriateness, and risk of misinformation or harm were consistently above 4 in both languages, with clarity rated as 5 in all responses. Completeness received the lowest scores (3.5-5.0), particularly in Turkish responses to resident-directed questions. Readability was higher in Turkish. English responses to resident queries were extremely difficult to read, while patient-directed ones remained in the "difficult" to "very difficult" range. Several discrepancies were observed in specific contents between English and Turkish outputs, such as differences in differential diagnosis lists, treatment options, contraindicated medications, and thymectomy indications.

Conclusion: ChatGPT-4o produced high-quality responses overall to MG-related queries in both languages. However, language-specific inconsistencies and content omissions highlight the need for further model refinement, particularly in multilingual and professional-use contexts.

Keywords: Artificial intelligence, ChatGPT-4o, large language models, myasthenia gravis, neurology

ÖZ

Amaç: Chat Generative Pre-Trained Transformer 4o (ChatGPT-4o) gibi büyük dil modelleri, hem hastalar hem de sağlık profesyonelleri tarafından sağlıkla ilgili bilgilere erişmek amacıyla giderek daha fazla kullanılmaktadır. Miyastenia gravis (MG), uzun süreli tedavi gerektiren kronik bir otoimmün nöromüsküler hastalıktır. Bu nedenle, MG hakkında doğru tıbbi bilgilere zamanında erişim önemlidir. Bu çalışma, ChatGPT-4o tarafından MG ile ilgili olarak hastalar ve nöroloji asistanları tarafından yöneltilen İngilizce ve Türkçe sorulara verilen yanıtların doğruluk, bütünlük, açıklık, hedef kitleye uygunluk, yanlış bilgilendirme veya zarar riski ile okunabilirlik açısından değerlendirilmesini amaçlamıştır.

Gereç ve Yöntem: İngilizce ve Türkçe olarak, hastalar ve nöroloji asistanları tarafından MG hakkında sıkça sorulan sorulardan oluşan, patofizyoloji ve semptomlar, tanı, tedavi, prognoz ve günlük yaşam yönetimini kapsayan 20 soruluk dört soru seti oluşturuldu. Her bir set için ChatGPT-4o yanıtları, 29 Mart 2025 tarihinde ayrı oturumlarda üretildi. İki nörolog, yanıtları beş farklı alanda 5 puanlık Likert ölçeği kullanarak birbirinden bağımsız olarak değerlendirdi. Okunabilirlik; İngilizce için Flesch okuma kolaylığı skoru, Flesch-Kincaid sınıf düzeyi ve Gunning-Fog indeksi ile, Türkçe için ise Ateşman okunabilirlik indeksi ile değerlendirildi.

Bulgular: Doğruluk, açıklık, uygunluk ve yanlış bilgilendirme ya da zarar verme riski açısından puanlar her iki dilde de tutarlı şekilde 4'ün üzerinde olup, açıklık tüm yanıtlarda 5 olarak değerlendirilmiştir. Bütünlük ise, özellikle asistanlara yönelik Türkçe yanıtlarda en düşük puanları (3,5-5,0) almıştır. Okunabilirlik Türkçe'de daha yüksek bulunmuştur. Asistanlara yönelik İngilizce yanıtlar son derece zor okunabilirken, hastalara yönelik

Address for Correspondence: Berin İnan, MD, University of Health Sciences Türkiye, Gülhane Faculty of Medicine, Department of Neurology, Ankara, Türkiye
E-mail: berin.inan@yahoo.com **ORCID ID:** orcid.org/0000-0002-2758-0207

Cite as: İnan B, Karadaş Ö, Odabaşı Z. Evaluation of ChatGPT-4o's responses to questions about myasthenia gravis in English and Turkish. Med J Bakirkoy. 2025;21(3):310-315

Received: 22.05.2025
Accepted: 08.07.2025
Publication Date: 03.09.2025

olanlar "zor" ile "çok zor" arasında değişmiştir. İngilizce ve Türkçe yanıtlar arasında; ayırıcı tanı listesi, tedavi seçenekleri, kontrendike ilaçlar ve timektomi endikasyonları gibi belirli içeriklerde çeşitli tutarsızlıklar gözlemlenmiştir.

Sonuç: ChatGPT-4o, MG ile ilgili sorulara her iki dilde de genel olarak yüksek kaliteli yanıtlar üretmiştir. Ancak dile özgü tutarsızlıklar ve içerik eksiklikleri, özellikle çok dilli ve profesyonel kullanım bağlamlarında modelin daha da geliştirilmesi gerektiğini ortaya koymaktadır.

Anahtar Kelimeler: Yapay zeka, ChatGPT-4o, büyük dil modelleri, miyastenia gravis, nöroloji

INTRODUCTION

In recent years, artificial intelligence (AI) has been increasingly used in the medical field by both patients and healthcare professionals. In particular, large language models (LLMs), a type of machine learning model designed to understand, analyze, generate, and manipulate human language, have gained popularity as fast, easy, and accessible tools for addressing a broad range of inquiries, from everyday concerns to complex academic questions (1,2).

Myasthenia gravis (MG) is the most common disorder of the neuromuscular junction and is characterized by fatigable skeletal muscle weakness (3,4). As a chronic autoimmune condition that necessitates long-term treatment, MG significantly impacts patients' quality of life and requires effective patient education and engagement (5). Patients frequently use the internet to search for information about the signs and symptoms of the disease, available treatment options, and strategies for the daily management of myasthenic symptoms (6,7). Similarly, medical professionals increasingly rely on LLMs for rapid access to information about a variety of medical conditions, including MG (6,7). However, the accuracy and reliability of the AI-generated content vary, requiring careful evaluation, as LLMs may have a tendency to hallucinate, resulting in misinformation (8-11).

Chat Generative Pre-Trained Transformer 4 (ChatGPT-4), one of the widely used LLMs, was launched by OpenAI in March 2023 (1). ChatGPT-4 is capable of processing both text and image inputs and performing complex tasks (1). More recently, an advanced version, ChatGPT-4o, was released by OpenAI in May 2024 (12). ChatGPT-4o is superior to ChatGPT-4 in terms of speed, cost-efficiency, multimodal functionality, and multilingual performance (2,12,13). While ChatGPT-4o is available for free with usage limitations and also as a paid version with extended features, ChatGPT-4 is not freely accessible (13).

In this study, we aimed to evaluate and compare the accuracy, completeness, clarity, appropriateness for the target audience, risk of misinformation or harm, and readability of ChatGPT-4o-generated responses to queries about MG from patients and neurology residents, in both English and Turkish.

METHODS

Study Design and Analysis of Responses

In this cross-sectional study, we developed four sets of queries consisting of frequently asked questions about MG from patients and neurology residents, in both English and Turkish. The questions were identical across the English and Turkish versions, and the content of the questions was similar in the patient and neurology resident groups. However, while the questions were phrased in a scientific tone in the neurology resident group, plain language was preferred for the patient group. Each set of queries consisted of 20 questions, including five on pathophysiology and symptoms, three on diagnosis, four on treatment modalities, four on prognosis, and four on the daily management of MG.

Each set of queries was submitted to ChatGPT-4o separately in a new chat window on 29 March 2025. The responses generated by ChatGPT-4o were independently evaluated by two neurologists specialized in neuromuscular diseases, based on the current literature about MG (4). Evaluations were conducted across five domains: accuracy, completeness, clarity, appropriateness for the target audience, and risk of misinformation or harm. Each domain was assessed using a 5-point Likert scale, ranging from 1 (poor) to 5 (excellent). The mean of the scores given by the two experts for each domain was calculated and used for statistical analysis.

We analyzed the readability of the responses in English using Readable software (Readable.com, Horsham, United Kingdom) (14), applying the Flesch Reading Ease score (FRES), Flesch-Kincaid grade level, and Gunning-Fog index. The readability of the responses in Turkish was evaluated using the Ateşman readability index (15,16).

Statistical Analysis

The normality of the data was assessed using the Shapiro-Wilk test. Descriptive statistics were given as mean±standard deviation or median (minimum-maximum) for continuous variables, and as frequency (percentage) for categorical variables. Group comparisons were performed using the Mann-Whitney U test or the independent samples t-test. All statistical analyses were conducted using SPSS for Windows, version 23.0 (IBM Corp., Armonk, NY, USA). Statistical significance was set as a p-value <0.05.

Ethical Approval

This cross-sectional study did not involve human participants, human tissue, or individually identifiable data. Therefore, informed consent and ethical approval were waived in accordance with institutional and national guidelines.

RESULTS

The accuracy, clarity, appropriateness for the target audience, and low risk of misinformation or harm scores for all ChatGPT-4o-generated responses were above 4 (very good) in both Turkish and English. All responses received a score of 5 points (excellent) for clarity from both experts. Among all evaluated domains, completeness received the lowest scores, ranging from 3.5 to 5 points. Although statistical analysis was not performed for the subgroups (pathophysiology and symptoms, diagnosis, treatment modalities, prognosis, and daily life management) due to the small sample size, we observed that completeness scores were higher for responses related to prognosis and daily life management than for those addressing pathophysiology and symptoms, diagnosis, and treatment modalities.

The scores for accuracy, completeness, clarity, appropriateness for the target audience, and risk of misinformation or harm were comparable between the patient and resident groups in the English responses. However, in the Turkish responses, the patient group demonstrated higher completeness scores than the resident group (Tables 1 and 2).

In the assessment of readability in English, responses to the patient group were found to be easier to read (Table 1). Readability scores were similar between the patient and neurology resident groups in Turkish responses (Table 2).

When comparing English and Turkish responses, the scores for accuracy, completeness, clarity, appropriateness, and risk of misinformation or harm did not differ in responses to patient queries. However, the completeness scores for the resident group were significantly lower in Turkish than in English (Table 3).

Table 1. Analysis of ChatGPT-4o's responses in English

Parameters	For patients	For residents	p-value
Accuracy, median (min-max)	5.0 (4.0-5.0)	5.0 (4.0-5.0)	0.799
Completeness, median (min-max)	5.0 (3.5-5.0)	5.0 (4.0-5.0)	0.602
Clarity, median (min-max)	5.0 (5.0-5.0)	5.0 (5.0-5.0)	1.000
Appropriateness for audience, median (min-max)	5.0 (5.0-5.0)	5.0 (4.0-5.0)	0.799
Risk of misinformation or harm, median (min-max)	5.0 (4.5-5.0)	5.0 (5.0-5.0)	0.799
Readability scores			
FRES, mean±SD	28.1±22.6	1.0±26.4	**0.001**
FKGL, mean±SD	11.5±3.5	15.2±3.9	**0.003**
GFI, mean±SD	12.0±3.4	14.9±4.3	**0.024**

FKGL: Flesch-Kincaid grade level, FRES: Flesch Reading Ease score, GFI: Gunning fog index, SD: Standard deviation

Table 2. Analysis of ChatGPT-4o's responses in Turkish

Parameters	For patients	For residents	p-value
Accuracy, median (min-max)	5.0 (4.5-5.0)	5.0 (4.0-5.0)	0.799
Completeness, median (min-max)	5.0 (3.5-5.0)	5.0 (4.0-5.0)	0.602
Clarity, median (min-max)	5.0 (5.0-5.0)	5.0 (5.0-5.0)	1.000
Appropriateness for audience, median (min-max)	5.0 (4.5-5.0)	5.0 (4.0-5.0)	0.799
Risk of misinformation or harm, median (min-max)	5.0 (5.0-5.0)	5.0 (5.0-5.0)	0.799
Ateşman readability index, median (min-max)	69.9 (40.2-91.2)	78.2 (25.3-95.7)	0.068

Table 3. Comparison of ChatGPT-4o's responses in English and Turkish

	English	Turkish	p-value
For patients			
Accuracy, median (min-max)	5.0 (4.0-5.0)	5.0 (4.5-5.0)	0.968
Completeness, median (min-max)	5.0 (3.5-5.0)	5.0 (3.5-5.0)	0.620
Clarity, median (min-max)	5.0 (5.0-5.0)	5.0 (5.0-5.0)	1.000
Appropriateness for audience, median (min-max)	5.0 (5.0-5.0)	5.0 (4.5-5.0)	0.799
Risk of misinformation or harm, median (min-max)	5.0 (4.5-5.0)	5.0 (5.0-5.0)	0.799
For residents			
Accuracy, median (min-max)	5.0 (4.0-5.0)	5.0 (4.5-5.0)	0.904
Completeness, median (min-max)	5.0 (4.0-5.0)	4.5 (3.5-5.0)	**0.035**
Clarity, median (min-max)	5.0 (5.0-5.0)	5.0 (5.0-5.0)	1.000
Appropriateness for audience, median (min-max)	5.0 (4.0-5.0)	5.0 (4.0-5.0)	0.192
Risk of misinformation or harm, median (min-max)	5.0 (5.0-5.0)	5.0 (5.0-5.0)	1.000

DISCUSSION

In this study, we examined the performance of ChatGPT-4o in responding to frequently asked queries about MG, both in English and Turkish, by patients, and neurology residents. Rather than focusing solely on quantitative scores, we also analyzed qualitative aspects such as clarity, completeness of the content, contextual appropriateness for the target audiences, and linguistic differences.

ChatGPT-4o exhibited high overall response quality and reliability in our study, as reflected by high accuracy, clarity, appropriateness scores, and low risk of misinformation. However, the performance of ChatGPT models has shown variability depending on the complexity of content and disease subtypes (17-20). ChatGPT-3.5 achieved stronger results in peripheral nerve and cerebrovascular diseases, while its performance was weaker in neuromuscular junction disorders and multiple sclerosis (17). Although ChatGPT-4's performance matched or even surpassed physicians in multiple-choice and board-style exams (18,19), its capacity for clinical reasoning and higher-order decision making remained limited (18). Furthermore, concerns regarding overconfidence and factual inconsistency have been raised, particularly in open-ended questions or complex medical scenarios (18-20).

The completeness scores were the lowest among all domains, which exhibited a language-dependent variation, prompting further evaluation. Completeness scores were notably lower in Turkish responses to resident queries compared to their English counterparts. This finding is consistent with prior studies suggesting that ChatGPT-4 exhibits higher performance in English, likely due to the predominance of English-language training data in the model's training corpus (21-24). Studies evaluating the performance of ChatGPT-4 in bilingual examinations showed that the model achieved significantly higher accuracy in English than in Chinese (21), Arabic (23), and Korean (24). Language-related discrepancies in ChatGPT-4 performance have also been reported in clinical and public health domains. In a study evaluating ChatGPT's multilingual performance in clinical nutrition advice, ChatGPT-4 produced significantly lower quality outputs in Kazakh compared to English and Russian (22). These findings collectively suggest that language-based performance biases remain a challenge and underscore the need for multilingual fine-tuning and dataset diversification to address such disparities. Although ChatGPT-4o has been promoted as a multilingually improved version of its predecessors (2,12), to the best of our knowledge, no study has yet evaluated its multilingual performance in any medical context, including MG.

In addition to the differences in completeness scores between English and Turkish responses, our study also identified linguistic discrepancies in specific content elements. For instance, in response to the patient group's question, "Can MG be mistaken for other conditions?", the differential diagnoses listed differed between the English and Turkish versions. Similarly, the response to the question on treatment options included eculizumab in English but omitted it in Turkish. The list of contraindicated medications in MG also varied between the two languages. Similar discrepancies were observed in resident-directed queries. The differential diagnoses of MG were not consistent across languages, and the indications for thymectomy included different age thresholds in the English and Turkish responses. Moreover, the predictors of spontaneous remission varied

between languages. These findings suggest potential inconsistencies in how ChatGPT-4o retrieves and generates language-specific medical content.

Notably, in our study, ChatGPT-4o achieved its highest completeness scores when responding to questions related to prognosis and the daily management of MG, suggesting the model's strength in patient-centered communication. This finding is consistent with previous studies across various medical conditions (9,11,25). ChatGPT-4o has been shown to provide accurate and reliable responses in contexts such as keratoconus (25), prostate cancer (9), and postmenopausal osteoporosis (11), particularly when addressing follow-up care, treatment adherence, and lifestyle recommendations.

Readability remains a significant barrier to patient accessibility in AI-generated medical content. In our study, responses to both patient and resident-directed questions in Turkish were fairly easy to read. However, in their English counterparts, responses to resident-directed questions were extremely difficult to read based on the FRES, whereas responses to patient-directed questions showed higher scores, indicating relatively better accessibility. Nevertheless, the overall readability for responses to patient-directed questions remained within the "difficult" to "very difficult" range, aligning with previous studies that have highlighted the limited readability of ChatGPT-4o's outputs in various clinical contexts (8-11,25). Encouragingly, several studies have shown that prompting ChatGPT-4o to simplify its language can significantly improve readability without compromising accuracy (9,10). Moreover, patient perceptions of understandability may not always correspond to objective readability indices, suggesting that future models should integrate real-time feedback and personalization to improve communication effectiveness (9).

Study Limitations

This study has several limitations. First, the number of questions was limited, restricting the application of statistical analyses in certain comparisons. Second, the evaluation was based on predefined questions, which may not fully reflect the variability of real-world questions posed by patients or residents. Additionally, only two languages were assessed, limiting the generalizability of the findings to other languages. Finally, readability was evaluated using standard indices that may not accurately represent actual patient comprehension.

CONCLUSION

ChatGPT-4o demonstrated high overall performance in responding to medical queries about MG, providing accurate, clear, and contextually appropriate answers in both English and Turkish. Although minor language-related differences were observed, particularly in the completeness of complex responses and certain factual discrepancies, ChatGPT-4o shows strong potential as a supportive tool for both patient education and professional reference.

ETHICS

Ethics Committee Approval: This cross-sectional study did not involve human participants, human tissue, or individually identifiable data. Therefore, ethical approval were waived in accordance with institutional and national guidelines.

Informed Consent: Informed consent was not obtained.

FOOTNOTES

Authorship Contributions

Concept: B.İ., Z.O., Design: B.İ., Z.O., Data Collection or Processing: B.İ., Ö.K., Analysis or Interpretation: B.İ., Ö.K., Z.O., Literature Search: B.İ., Ö.K., Writing: B.İ.

Conflict of Interest: No conflict of interest was declared by the authors.

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

REFERENCES

1. OpenAI. GPT-4 technical report. arXiv Preprint. 2023; arXiv:2303.08774. Available from: https://arxiv.org/pdf/2303.08774
2. Shahriar S, Lund BD, Mannuru NR, Arshad MA, Hayawi K, Bevara RVK, et al. Putting GPT-4o to the sword: a comprehensive evaluation of language, vision, speech, and multimodal proficiency. Appl Sci. 2024;14:7782.
3. Gilhus NE, Tzartos S, Evoli A, Palace J, Burns TM, Verschuuren JJGM. Myasthenia gravis. Nat Rev Dis Primers. 2019;5:30.
4. Bird SJ. Myasthenia gravis [Internet]. UpToDate; 2025 [cited 2025 Mar 29]. Available from: https://www.uptodate.com/contents/overview-of-the-treatment-of-myasthenia-gravis?search=Overview%20of%20the%20treatment%20of%20myasthenia%20gravis&source=search_result&selectedTitle=1~150&usage_type=default&display_rank=1
5. Dewilde S, Philips G, Paci S, Beauchamp J, Chiroli S, Quinn C, et al. Patient-reported burden of myasthenia gravis: baseline results of the international prospective, observational, longitudinal real-world digital study MyRealWorld-MG. BMJ Open. 2023;13:e066445.
6. Jin JQ, Dobry AS. ChatGPT for healthcare providers and patients: practical implications within dermatology. J Am Acad Dermatol. 2023;89:870-1.

7. Tan S, Xin X, Wu D. ChatGPT in medicine: prospects and challenges: a review article. Int J Surg. 2024;110:3701-6.
8. Wang S, Wang Y, Jiang L, Chang Y, Zhang S, Zhao K, et al. Assessing the clinical support capabilities of ChatGPT 4o and ChatGPT 4o mini in managing lumbar disc herniation. Eur J Med Res. 2025;30:45.
9. Trapp C, Schmidt-Hegemann N, Keilholz M, Brose SF, Marschner SN, Schonecker S, et al. Patient- and clinician-based evaluation of large language models for patient education in prostate cancer radiotherapy. Strahlenther Onkol. 2025;201:333-42.
10. Li J, Chang C, Li Y, Cui S, Yuan F, Li Z, et al. Large language models' responses to spinal cord injury: a comparative study of performance. J Med Syst. 2025;49:39.
11. Liu R, Liu J, Yang J, Sun Z, Yan H. Comparative analysis of ChatGPT-4o mini, ChatGPT-4o and Gemini advanced in the treatment of postmenopausal osteoporosis. BMC Musculoskelet Disord. 2025;26:369.
12. OpenAI. GPT-4o System Card [Internet]. 2025 [cited 2025 Aug 8]. Available from: https://arxiv.org/pdf/2410.21276
13. Craig L. GPT-4o vs. GPT-4: How do they compare? [Internet]. 2025 [cited 2025]. Available from: https://www.techtarget.com/searchenterpriseai/feature/GPT-4o-vs-GPT-4-How-do-they-compare#:~:text=OpenAI's%20testing%20indicates%20that%20GPT,idioms%2C%20metaphors%20and%20cultural%20references.
14. Readable.com. Readability score [Internet]. [cited 2025]. Available from: https://app.readable.com/text/?demo&_ga=2.14124151.156675892.1746127328-404057203.1743896947
15. Ateşman E. Measurement of readability in Turkish. Language J [Internet]. 1997;58:71-4.
16. Türkçe okunabilirlik indeksi; 2025. Erişim adresi: http://okunabilirlikindeksi.com/
17. Altunisik E, Firat YE, Cengiz EK, Comruk GB. Artificial intelligence performance in clinical neurology queries: the ChatGPT model. Neurol Res. 2024;46:437-43.
18. Chen TC, Multala E, Kearns P, Delashaw J, Dumont A, Maraganore D, et al. Assessment of ChatGPT's performance on neurology written board examination questions. BMJ Neurol Open. 2023;5:e000530.
19. Ros-Arlanzón P, Perez-Sempere A. Evaluating AI competence in specialized medicine: comparative analysis of ChatGPT and neurologists in a neurology specialist examination in Spain. JMIR Med Educ. 2024;10:e56762.
20. García-Rudolph A, Sanchez-Pinsach D, Opisso E, Soler MD. Exploring new educational approaches in neuropathic pain: assessing accuracy and consistency of artificial intelligence responses from GPT-3.5 and GPT-4. Pain Med. 2025;26:48-50.
21. Wu Z, Gan W, Xue Z, Ni Z, Zheng X, Zhang Y. Performance of ChatGPT on nursing licensure examinations in the United States and China: cross-sectional study. JMIR Med Educ. 2024;10:e52746.
22. Adilmetova G, Nassyrov R, Meyerbekova A, Karabay A, Varol HA, Chan MY. Evaluating ChatGPT's multilingual performance in clinical nutrition advice using synthetic medical text: insights from central Asia. J Nutr. 2025;155:729-35.
23. Sallam M, Al-Mahzoum K, Alshuaib O, Alhajri H, Alotaibi F, Alkhurainej D, et al. Language discrepancies in the performance of generative artificial intelligence models: an examination of infectious disease queries in English and Arabic. BMC Infect Dis. 2024;24:799.
24. Song ES, Lee SP. Comparative analysis of the response accuracies of large language models in the Korean national dental hygienist examination across Korean and English questions. Int J Dent Hyg. 2025;23:267-76.
25. Balci AS, Cakmak S. Evaluating the accuracy and readability of ChatGPT-4o's responses to patient-based questions about keratoconus. Ophthalmic Epidemiol. 2025:1-6.

Medical Journal of Bakırköy

Med J Bakirkoy 2025;21(3):316-322
DOI: 10.4274/BMJ.galenos.2025.2025.4-4

Research

Percutan Crossed Dual Planar Pin Fixation Technique Provides Satisfactory Radiological and Clinical Results in Intra-articular Calcaneus Fractures with Low Complication Rates

Perkütan Çapraz Çift Planar Pin Fiksasyon Tekniği, Düşük Komplikasyon Oranlarıyla Eklem İçi Kalkaneus Kırıklarında Tatmin Edici Radyolojik ve Klinik Sonuçlar Sağlar

Berhan Bayram[1], Mustafa Gökhan Bilgili[2]

[1]Acıbadem Altunizade Hospital, Clinic of Orthopedic Surgery and Sport Medicine, İstanbul, Türkiye
[2]University of Health Sciences Türkiye, Bakırköy Dr. Sadi Konuk Training and Research Hospital, Clinic of Orthopedics and Traumatology, İstanbul, Türkiye

ABSTRACT

Objective: This study aimed to evaluate the clinical and radiological results of crossed double-plane pin fixation of intra-articular calcaneus fractures based on the Sanders classification.

Methods: The study was designed retrospectively. Patients who were operated on using the crossed double-plane pin fixation technique between 2013 and 2017 were included in the study. Patients under the age of 18 who had previously undergone surgery around the foot/ankle with extra-articular fractures with fractures on the ipsilateral or contralateral lower extremity were excluded from the study. All patients were operated on by a single surgeon in a single clinic. All patients were categorized according to the Sanders classification, with pre-operative computed tomography. The Böhler angle was used for radiological evaluation, and the American Orthopaedic Foot and Ankle Society score was used for clinical evaluation.

Results: After applying the inclusion and exclusion criteria, a total of 52 fractures of 46 patients were included in the study. Thirty-six (78.3%) of the patients were male and 10 (21.7%) were female. Twenty-one patients had right-sided fractures (45.7%), nineteen patients had left-sided fractures (41.3%), and six patients had bilateral fractures (13%). The mean patient age at the time of treatment was 39.6 years (15-64 years). Fractures were classified according to the Sanders classification system. According to this classification, eight of the fractures were type 2 (15.4%), thirty-five were type 3 (67.3%), and nine were type 4 (17.3%). Of the forty-five calcaneus fractures with abnormal preoperative Böhler values, thirty-three (73%) had normal Böhler values in postoperative control measurements. There was a statistically significant difference between preoperative and postoperative measurements in terms of Böhler distributions ($p<0.05$). While the baseline rate was low in the preoperative measurement, it was higher in the postoperative measurement. In addition, a statistically significant difference was found in the number of individuals returning to normal as per Sanders grouping (Fisher's exact $p<0.05$). The rate of those with Sanders 3 was found to be higher compared to others.

Conclusion: Double pin fixation technique in intra-articular calcaneus fractures provides satisfactory radiological and clinical results with low complication rates.

Keywords: Calcaneus fracture, percutaneous, cross, pin, fixation

Address for Correspondence: Berhan Bayram MD, Acıbadem Altunizade Hospital, Clinic of Orthopedic Surgery and Sport Medicine, İstanbul, Türkiye
E-mail: dr.berhanbayram@gmail.com **ORCID ID:** orcid.org/0000-0003-3777-5611

Cite as: Bayram B, Bilgili MG. Percutan crossed dual planar pin fixation technique provides satisfactory radiological and clinical results in intra-articular calcaneus fractures with low complication rates. Med J Bakirkoy. 2025;21(3):316-322

Received: 08.05.2025
Accepted: 31.05.2025
Publication Date: 03.09.2025

ÖZ

Amaç: Bu çalışma ile intra-artikuler kalkaneus kırıkların çapraz çift planlı pin fiksasyonun klinik ve radyolojik sonuçlarının Sanders sınıflandırması baz alınarak değerlendirilmesi amaçlanmıştır.

Gereç ve Yöntem: Çalışma retrospektif olarak dizayn edilmiştir. 2013-2017 yılları arasında çapraz çift planlı pin fiksasyonu tekniği kullanılarak opere edilen hastalar çalışmaya dahil edildi. On sekiz yaşın altındaki hastalar, daha önce ayak/ayak bileği çevresinden cerrahi geçiren hastalar, ekstra artiküler kırığı olan hastalar, aynı tarafta veya karşı taraf alt ekstremitesinde kırığı olan hastalar çalışma dışında bırakıldı. Hastaların tamamı tek bir klinikte tek bir cerrah tarafından opere edildi. Hastaların tamamı pre-operatif çekilen bilgisayarlı tomografi ile Sanders sınıflamasına göre kategorize edildi. Radyolojik değerlendirme için Böhler açısı, klinik değerlendirme içinse Amerikan Ortopedi Ayak ve Ayak Bileği Derneği skoru kullanıldı.

Bulgular: Dahil edilme ve dışlanma kriterlerinden sonra çalışmaya toplamda 46 hastanın 52 kırığı dahil edildi. Hastaların 36'sı (%78,3) erkek, 10'u (%21,7) kadındı. Yirmi bir hastada sağ taraf (%45,7), 19 hastada sol taraf (%41,3), altı hastada ise her iki tarafta (%13) kırık saptandı. Tedavi anında ortalama hasta yaşı 39,6 (15-64) idi. Kırıklar Sanders sınıflandırma sistemine göre sınıflandırıldı. Bu sınıflandırmaya göre kırıkların 8'i tip 2 (%15,4), 35'i tip 3 (%67,3), 9 tanesi de tip 4 (% 17,3) olarak saptandı. Preop Böhler değeri normal olmayan 45 kalkaneus kırığının, postop kontrol ölçümlerinde 33 tanesinin (%73) Böhler değerlerinin normale geldiği görüldü. Preop ve postop ölçümleri arasında Böhler dağılımları açısından istatistiksel anlamlı farklılık saptandı ($p<0,05$). Önce ölçümünde normal oranı düşükken sonra ölçümünde daha yüksek olduğu görüldü. Ayrıca Sanders gruplamasına göre normale geçenler açısından istatistiksel anlamlı farklılık saptandı (Fisher'ın kesin testi $p<0,05$). Sanders 3 olanların oranı daha yüksek bulunmuştur.

Sonuç: İntra-artiküler kalkaneus kırıklarında çift pin fiksasyon tekniği düşük komplikasyon oranları ile tatmin edici radyolojik ve klinik sonuçlar sağlar.

Anahtar Kelimeler: Kalkaneus kırığı, perkütan, çapraz, pin, fiksasyon

INTRODUCTION

The calcaneus is the most commonly fractured tarsal bone, representing approximately 60% of foot fractures and 1-2% of all fractures (1). Calcaneal fractures are associated with significant morbidity and a high rate of complications (2,3). Notably, over 70% of these fractures are intra-articular, affecting the subtalar joint (4). Displaced intra-articular calcaneal fractures (DIACF) can result in various functional issues, including hindfoot deformities, persistent pain, and chronic stiffness (5). When treated conservatively, DIACF is linked to poor clinical outcomes due to the development of subtalar arthritis and disruption of foot morphology (6).

The management of intra-articular calcaneal fractures is complex and presents numerous challenges. The primary objectives of treatment are to restore the calcaneal height and length to their pre-injury state and to reconstruct the posterior joint surface (7). Although the fundamental goals of treatment modalities are consistent, the optimal approach for managing intra-articular calcaneal fractures remains a topic of debate (8). These fractures are frequently addressed using open reduction and internal fixation (ORIF) via an extensile lateral approach. However, this method carries a considerable risk of wound complications and sural nerve injury (9,10). As a result, there has been a growing preference for less invasive techniques in the treatment of intra-articular calcaneal fractures. Among these approaches, percutaneous screw fixation has demonstrated the ability to reduce wound complications while achieving satisfactory reduction (11,12). Additionally, percutaneous fixation with K-wires offers benefits such as lower implant costs and shorter surgical times. Literature suggests that it can yield outcomes comparable to traditional ORIF (13,14). These encouraging findings have contributed to the increasing adoption of minimally invasive techniques in recent years. Nonetheless, there is a notable lack of comprehensive studies to identify which patient populations are most suitable for these methods.

In this study, we aimed to classify intra-articular calcaneal fractures treated with cross dual-planar pin fixation according to the Sanders classification (15) and to assess the impact of this technique on both radiological and clinical outcomes. By doing so, we sought to enhance the existing literature on patient selection for this approach. We hypothesized that patients with Sanders type 2 and type 3 fractures would experience better outcomes compared to those with type 4 fractures.

METHODS

Demographic Characteristics and Preoperative Plan

The study was designed as a retrospective study. Data collection began after approval from the Clinical Research Ethics Committee of the University of Health Sciences Türkiye, Bakırköy Dr. Sadi Konuk Training and Research Hospital (approval no: 2018-03-23, date: 12.02.2018). Patients who underwent surgery for intra-articular calcaneal fractures at our hospital between 2012 and 2017 were retrospectively reviewed. Among these, only patients operated on using the cross dual-planar pinning technique were included in the study. Exclusion criteria included patients under the age of 18, those with open calcaneal

fractures, extra-articular calcaneal fractures, ipsilateral or contralateral lower extremity fractures, pathological fractures, a history of prior foot or ankle surgery, a diagnosis of rheumatologic disease, or neurovascular deficits in the affected extremity. Patients presenting to the emergency department due to trauma underwent anteroposterior (AP), lateral, and axial radiographs of the ankle, in cases where a calcaneal fracture was suspected. For patients diagnosed with a calcaneal fracture based on radiographic imaging and clinical examination, a foot and ankle computed tomography (CT) scan was performed to classify the fracture. The Sanders classification system was used for all patients. The classification of CT scans for all cases was performed by the primary author of this study.

Surgical Technique

All surgical procedures were performed at a single center by the same surgeon. General or spinal anesthesia was administered to all patients for the surgical intervention. A tourniquet was not applied during the procedures. The surgeries were conducted with the patients the prone position. This surgical technique involves the use of two crossed Schanz screws, positioned posteromedially and posterolaterally.

Initially, a screw was inserted from the posteromedial aspect in a manner parallel to the collapsed joint fragment. This Schanz screw was directed to pass just beneath the fragment, requiring elevation from the tuber calcanei. The entry point and direction of the Schanz screw were determined using C-arm fluoroscopy. Once the screw was positioned appropriately, it was manipulated medially to laterally to correct the varus alignment. Simultaneously, the screw was advanced in a cranial-to-caudal direction to facilitate the reduction of the collapsed joint fragment. If adequate fracture reduction was achieved, the Schanz screw was advanced to the most distal part of the calcaneus without penetrating the calcaneocuboid joint. Following the placement of the first screw, the entry point for the second screw was also determined under C-arm fluoroscopic guidance. The posterolateral screw was directed from the tuber calcanei, passing beneath the residual impacted fragment and parallel to the collapsed joint fragment. At this stage, multiple radiographic views were required to confirm the correct direction of the screw. Unlike the first screw, the second screw was manipulated solely for joint fragment elevation before being advanced to the most distal aspect of the calcaneus. After the percutaneous placement of the screws, reduction was assessed using C-arm fluoroscopy. Finally, the screw entry points were closed with absorbable sutures, and all patients were immobilized in a long leg cast after forefoot manipulation. The dual shanz fixation technique was applied as described by Duramaz et al. (16). All patients received antibiotic prophylaxis during anesthesia induction, followed by two additional postoperative doses. Low-molecular-weight heparin was administered to all patients as prophylaxis against thromboembolic disease.

Postoperative Follow-up

Following discharge, all patients were scheduled for follow-up visits every two weeks during the first month, monthly thereafter, and annually after six months from the discharge, if no additional complaints were reported. At each follow-up, AP, lateral, and axial radiographs of the ankle were obtained for all patients. CT was performed only once, on the first postoperative day. Böhler's angle was measured from preoperative and postoperative day 1 radiographs for all patients. The measurements were independently conducted by two different orthopedic specialists (BB and MGB), and the mean values of both measurements were used for statistical analysis. For clinical evaluation, the American Orthopaedic Foot and Ankle Society (AOFAS) score (17) was assessed at the final follow-up. In cases of bilateral fractures, a separate AOFAS score was calculated for each foot. Pre-operative, intra-operative, and post-operative radiographs of one of the patients included in the study are shown in Figure 1.

Statistical Analysis

Descriptive statistics were used to describe continuous variables. The relationship between two independent categorical variables was analyzed using Fisher's exact test, while the relationship between two dependent categorical variables was assessed using the McNemar test. The level of statistical significance was set at 0.05. All statistical analyses were performed using MedCalc Statistical Software version 12.7.7 (MedCalc Software Ltd., Ostend, Belgium).

RESULTS

A total of 52 calcaneal fractures in 46 patients were included in the study. The majority of the patients were male (78.3%). The fractures were located on the right side in 21 patients (45.7%), on the left side in 19 patients (41.3%), and on the bilateral side in 6 patients (13%). The mean age of the patients was 39.6 years (range: 15-64 years).

Preoperative CT scans were used to classify the fractures according to the Sanders classification. Based on this classification, 8 fractures (15.4%) were categorized as Sanders type 2, 35 fractures (67.3%) as Sanders type 3, and 9 fractures (17.3%) as Sanders type 4. The distribution of patients according to sex, laterality, and Sanders classification is presented in Table 1.

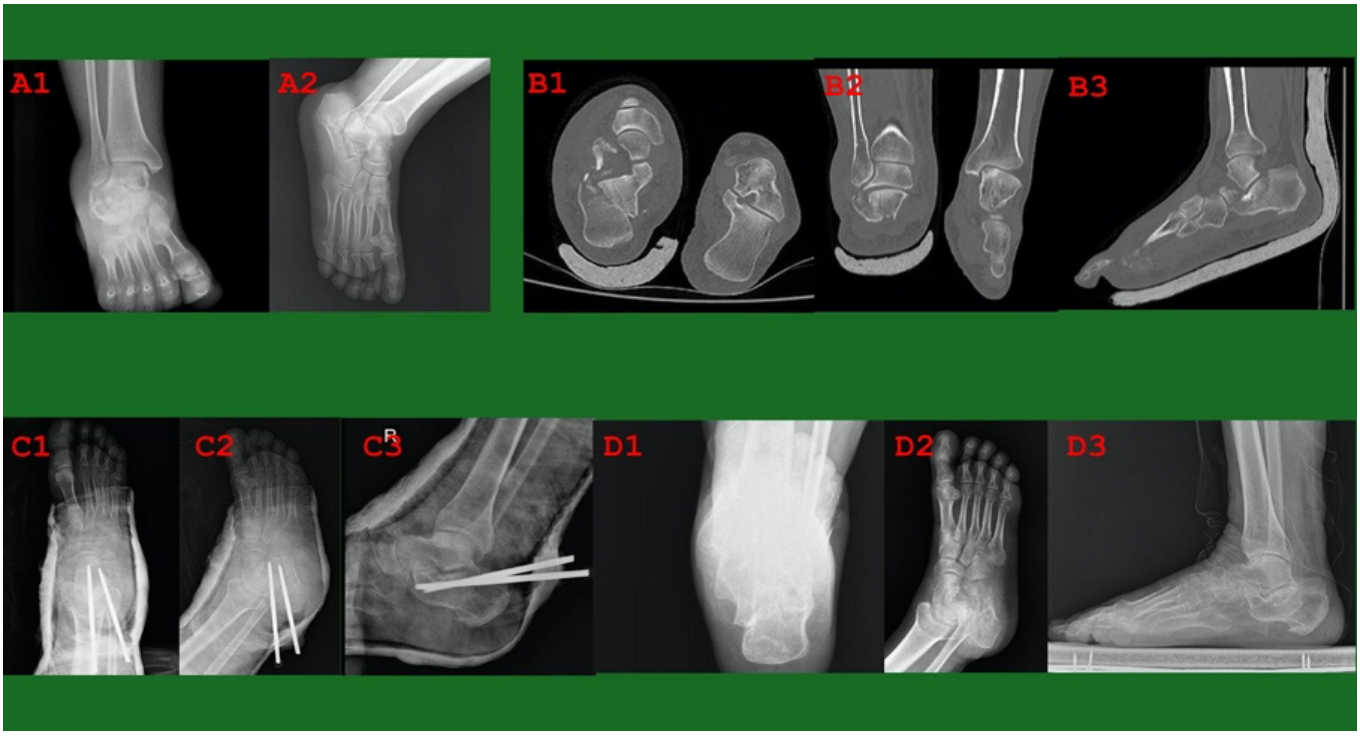

Figure 1. Fifty-five-year-old male admitted after falling from a height. **A1.** Pre-operative AP radiographs showing comminuted intra-articular calcaneus fracture. **A2.** Pre-operative lateral radiographs showing comminuted intra-articular calcaneus fracture. **B1.** Pre-operative computed tomography axial view showing displaced posterior facet of the calcaneus. **B2.** Pre-operative computed tomography coronal view showing displaced posterior facet of the calcaneus. **B3.** Pre-operative computed tomography sagittal view showing displaced posterior facet of the calcaneus. **C1.** Post-operative AP radiographs showing closed reduction and two crossed percutaneous Schanz pin fixation. **C2.** Post-operative oblique radiographs showing closed reduction and two crossed percutaneous Schanz pin fixation. **C3.** Post-operative lateral radiographs showing closed reduction and two crossed percutaneous Schanz pin fixation. **D1.** Control axial radiographs taken 1 year after surgery (AOFAS score: 91). **D2.** Control oblique radiographs taken 1 year after surgery (AOFAS score: 91). **D3.** Control lateral radiographs taken 1 year after surgery (AOFAS score: 91)

AP: Anteroposterior, AOFAS: American Orthopaedic Foot and Ankle Society

Table 1. Data regarding patients' gender, side and Sanders classification

Gender	**Female**	10	21.7
	Male	36	78.3
Side	**Right**	21	45.7
	Left	19	41.3
	Bilateral	6	13
Group preop sanders	**2**	8	15.4
	3	35	67.3
	4	9	17.3

The mean follow-up period for the patients was 47.5 months (range: 17.4-70.6 months). At the final follow-up, the mean AOFAS score was 83.25 (range: 47-96). The AOFAS scores were also analyzed separately according to the Sanders classification of fractures. The mean AOFAS scores were 74.3 for Sanders type 2 fractures, 84.7 for Sanders type 3 fractures, and 85.4 for Sanders type 4 fractures.

Among the 45 calcaneal fractures with an abnormal preoperative Böhler's angle, 33 cases (73%) achieved a postoperatively normalized Böhler's angle. In the 7 patients whose preoperative Böhler's angle was within normal limits, postoperative measurements also remained within the normal range. Each patient's preoperative and postoperative Böhler's angles were assessed individually, and a statistically significant improvement was observed between the preoperative and postoperative measurements ($p<0.05$).

The distribution of patients categorized according to Böhler's angle is presented in Table 2.

The relationship between fracture types and changes in Böhler's angle is presented in Table 3.

In the study, there was no statistically significant difference between the preoperative and postoperative Böhler's angles in patients with Sanders type 2 fractures ($p>0.05$). However, a statistically significant difference was observed in Sanders type 3 and Sanders type 4 fractures when comparing preoperative and postoperative Böhler's angle measurements ($p<0.05$).

Table 2. The distribution of patients according to Böhler's angle

All patients	Preop Böhler			
		Not normal, n (%)	Normal, n (%)	p-value
postBöhler	Not normal	12 (26.7)	0 (0.0)	**<0.001**
	Normal	33 (73.3)	7 (100.0)	

McNemar test

Table 3. Relationship between fracture types and changes in Böhler's angle

Sanders					
		2	3	4	p-value
		n (%)	n (%)	n (%)	
preBöhler	Normal	2 (28.6)	4 (57.1)	1 (14.3)	0.520
postBöhler	Normal	6 (15.0)	27 (67.5)	7 (17.5)	**<0.001**

The restoration of an abnormal Böhler's angle to normal values was also analyzed according to fracture type, and this improvement was found to be statistically significant across all fracture types. Among the three fracture types, the highest rate of normalization of Böhler's angle was observed in Sanders type 3 fractures.

No cases of nonunion, wound complications, thromboembolic events, or complications requiring revision surgery were encountered in any of the patients included in the study.

DISCUSSION

The management of DIACFs is complex and remains a subject of debate. The most important findings of the current study are that the crossed dual-pin fixation technique in a double-plane configuration for intra-articular calcaneal fractures provides both radiologically and clinically successful outcomes with low complication rates.

Conservative treatment of DIACFs is recommended for patients with medical comorbidities that pose a high surgical risk. However, this approach is associated with several complications, including subtalar arthritis, persistent pain, peroneal impingement syndrome, foot deformities, and flexor hallucis longus contracture (18,19). To mitigate these complications, ORIF techniques have been developed. However, ORIF is also associated with postoperative wound complications and nerve injuries (20), with reported complication rates ranging between 6% and 23% (21). Minimally invasive and percutaneous fixation techniques have been introduced to reduce the risk of complications. Various implant materials can be used for percutaneous fixation (22). In our study, Schanz screws, which are a more cost-effective alternative, were used as the fixation material.

Arora et al. (20) performed percutaneous fixation with K-wires in patients with DIACF and reported that, in radiological comparisons, postoperative Böhler's angle increased significantly in Sanders type 3 and type 4 fractures, while no statistically significant increase was observed in Sanders type 2 fractures. Chaniotakis et al. (23) described this technique as safe and effective, emphasizing that the quality of reduction is the most critical parameter influencing outcomes.

Mesregah et al. (24) treated Sanders type 2 and type 3 intra-articular calcaneal fractures using either closed or minimally open reduction followed by K-wire fixation. Their findings demonstrated excellent clinical outcomes with minimal soft tissue complications. Additionally, they reported no significant difference between fixation with K-wires and screws in terms of clinical results (24). Systematic reviews in the literature assessing the effectiveness of percutaneous fixation techniques in DIACF support the conclusion that this technique is a viable treatment option (25).

In our study, the crossed dual-pin fixation technique in a double-plane configuration was found to be an effective treatment method with low complication rates. This technique was particularly successful in the management of Sanders type 3 and type 4 fractures. However, in Sanders type 2 fractures, the outcomes were observed to be less satisfactory compared to the other fracture types.

The literature indicates that surgical restoration of Böhler's angle is associated with better long-term clinical outcomes (26). In our study, the AOFAS score was used, as it provides the best comparability across different studies (17).

In Arora et al. (20) study, the mean AOFAS score in patients who underwent fixation with K-wires was reported as 85.1. In contrast, El-Azab et al. (27) study found a mean AOFAS score of 75.8 in the K-wire fixation group. Another study

evaluating the short-term outcomes of DIACF treated with closed reduction and percutaneous K-wire fixation reported a satisfactory mean AOFAS score of 84 (28). Although numerical variations exist in the literature regarding the outcomes of this surgical technique, the majority of studies associate it with good to excellent results. In our study, the mean AOFAS score was 83.25, with more favorable outcomes observed in Sanders type 3 and type 4 fractures. This discrepancy was attributed to the less effective restoration of Böhler's angle in Sanders type 2 fractures compared to the other groups.

High rates of postoperative infection and wound complications have been reported in patients undergoing internal fixation via the extensile lateral approach. Sampath Kumar et al. (29) reported wound healing complications in 30.4% of their patients and deep infections in 13%. Similarly, in a larger case series involving 191 patients, the wound infection rate following ORIF was found to be 25% (30). In contrast to these high infection rates, various studies utilizing different percutaneous fixation techniques have reported significantly lower infection rates (12,20,28).

In our study, no cases of wound healing complications or infections were observed. This finding suggests that, with adequate preoperative preparation and strict adherence to sterile conditions, percutaneous fixation can achieve low complication rates, making it a safer alternative to open techniques.

Study Limitations

Our study has several limitations. The retrospective design and small sample size are among the primary limitations. Additionally, the use of only Böhler's angle as a radiological parameter and the relatively short follow-up period are constraints. Furthermore, the absence of a control or comparison group and the lack of a preoperative clinical scoring system for baseline assessment are additional limitations. However, a strength of this study is the categorization of outcomes based on the Sanders fracture classification, which enhances the clinical relevance of the findings. Additionally, all surgical procedures were performed at a single center by the same surgeon, representing another advantage and ensuring consistency in surgical technique and postoperative management.

CONCLUSION

In a conclusion, the crossed dual-pin fixation technique in a double-plane configuration can be effectively used in the treatment of DIACF, providing satisfactory radiological and clinical outcomes with low complication rates. This technique is particularly advantageous in Sanders type 3 and type 4 fractures, where better outcomes can be anticipated preoperatively.

ETHICS

Ethics Committee Approval: Data collection began after approval from the Clinical Research Ethics Committee of the University of Health Sciences Türkiye, Bakırköy Dr. Sadi Konuk Training and Research Hospital (approval no: 2018-03-23, date: 12.02.2018).

Informed Consent: Retrospective study.

FOOTNOTES

Authorship Contributions

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

Conflict of Interest: No conflict of interest was declared by the authors.

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

REFERENCES

1. Cao L, Weng W, Song S, Mao N, Li H, Cai Y, et al. Surgical treatment of calcaneal fractures of Sanders type II and III by a minimally invasive technique using a locking plate. J Foot Ankle Surg. 2015;54:76-81.
2. Hsu AR, Anderson RB, Cohen BE. Advances in surgical management of intra-articular calcaneus fractures. J Am Acad Orthop Surg. 2015;23:399-407.
3. Folk JW, Starr AJ, Early JS. Early wound complications of operative treatment of calcaneus fractures: analysis of 190 fractures. J Orthop Trauma. 1999;13:369-72.
4. Stulik J, Stehlik J, Rysavy M, Wozniak A. Minimally-invasive treatment of intra-articular fractures of the calcaneum. J Bone Joint Surg Br. 2006;88:1634-41.
5. Besch L, Schmidt I, Mueller M, Daniels-Wredenhagen M, Hilgert RE, Varoga D, et al. A biomechanical evaluation to optimize the configuration of a hinged external fixator for the primary treatment of severely displaced intraarticular calcaneus fractures with soft tissue damage. J Foot Ankle Surg. 2008;47:26-33.
6. Arastu M, Sheehan B, Buckley R. Minimally invasive reduction and fixation of displaced calcaneal fractures: surgical technique and radiographic analysis. Int Orthop. 2014;38:539-45.
7. Thermann H, Krettek C, Hüfner T, Schratt HE, Albrecht K, Tscherne H. Management of calcaneal fractures in adults. Conservative versus operative treatment. Clin Orthop Relat Res. 1998;353:107-24.
8. Barei DP, Bellabarba C, Sangeorzan BJ, Benirschke SK. Fractures of the calcaneus. Orthop Clin North Am. 2002;33:263-85.
9. Juliano P, Nguyen HV. Fractures of the calcaneus. Orthop Clin North Am. 2001;32:35-51.
10. Harnroongroj T, Tangmanasakul A, Choursamran N, Sudjai N, Harnroongroj T. Measurement technique of calcaneal varus from axial view radiograph. Indian J Orthop. 2015;49:223-6.

11. Dewall M, Henderson CE, McKinley TO, Phelps T, Dolan L, Marsh JL. Percutaneous reduction and fixation of displaced intra-articular calcaneus fractures. J Orthop Trauma. 2010;24:466-72.
12. Abdelgaid SM. Closed reduction and percutaneous cannulated screws fixation of displaced intra-articular calcaneus fractures. Foot Ankle Surg. 2012;18:164-79.
13. Schuberth JM, Cobb MD, Talarico RH. Minimally invasive arthroscopic-assisted reduction with percutaneous fixation in the management of intra-articular calcaneal fractures: a review of 24 cases. J Foot Ankle Surg. 2009;48:315-22.
14. Walde TA, Sauer B, Degreif J, Walde HJ. Closed reduction and percutaneous Kirschner wire fixation for the treatment of dislocated calcaneal fractures: surgical technique, complications, clinical and radiological results after 2-10 years. Arch Orthop Trauma Surg. 2008;128:585-91.
15. Sanders R, Vaupel ZM, Erdogan M, Downes K. Operative treatment of displaced intraarticular calcaneal fractures: long-term (10-20 years) results in 108 fractures using a prognostic CT classification. J Orthop Trauma. 2014;28:551-63.
16. Duramaz A, Polat Ö, İlter MH, Bayram B, Bayrak A, Baca E. Could percutaneous fixation with crossed Schanz pins be an alternative to open reduction in the treatment of intra-articular calcaneal fractures? Int Orthop. 2021;45:731-41.
17. Ibrahim T, Beiri A, Azzabi M, Best AJ, Taylor GJ, Menon DK. Reliability and validity of the subjective component of the American Orthopaedic Foot and Ankle Society clinical rating scales. J Foot Ankle Surg. 2007;46:65-74.
18. Rodriguez-Merchan EC, Galindo E. Intra-articular displaced fractures of the calcaneus. Operative vs non-operative treatment. Int Orthop. 1999;23:63-5.
19. Parmar HV, Triffitt PD, Gregg PJ. Intra-articular fractures of the calcaneum treated operatively or conservatively. A prospective study. J Bone Joint Surg Br. 1993;75:932-7.
20. Arora C, Jain AK, Dhammi IK. Outcome of percutaneous fixation of calcaneal fractures: a prospective analysis in an Indian population. J Foot Ankle Surg. 2019;58:502-7.
21. Yang Y, Zhao H, Zhou J, Yu G. Treatment of displaced intraarticular calcaneal fractures with or without bone grafts: a systematic review of the literature. Indian J Orthop. 2012;46:130-7.
22. Allegra PR, Rivera S, Desai SS, Aiyer A, Kaplan J, Gross CE. Intra-articular calcaneus fractures: current concepts review. Foot Ankle Orthop. 2020;5:2473011420927334.
23. Chaniotakis C, Genetzakis V, Samartzidis K, Siligardou MR, Stavrakakis I. Percutaneous fixation of displaced intraarticular fractures of the calcaneus. A retrospective case series study and a review of the literature. Injury. 2023;54:110966.
24. Mesregah MK, Shams A, Gamal O, Zaki EM. Clinical and radiological outcomes of minimally invasive reduction and percutaneous K-wire fixation for intra-articular calcaneal fractures. Orthopedics. 2020;43:97-101.
25. Wallin KJ, Cozzetto D, Russell L, Hallare DA, Lee DK. Evidence-based rationale for percutaneous fixation technique of displaced intra-articular calcaneal fractures: a systematic review of clinical outcomes. J Foot Ankle Surg. 2014;53:740-3.
26. Makki D, Alnajjar HM, Walkay S, Ramkumar U, Watson AJ, Allen PW. Osteosynthesis of displaced intra-articular fractures of the calcaneum: a long-term review of 47 cases. J Bone Joint Surg Br. 2010;92:693-700.
27. El-Azab H, Ahmed KFE, Khalefa AH, Marzouk AR. A prospective comparative study between percutaneous cannulated screws and Kirschner wires in treatment of displaced intra-articular calcaneal fractures. Int Orthop. 2022;46:2667-83.
28. Morsi IM, Khalifa AA, Hussien MA, Abdellatef A, Refae H. Evaluation of the short-term results of closed reduction and percutaneous K-wires fixation of displaced intra-articular calcaneal fractures (DIACF). Foot (Edinb). 2020;45:101740.
29. Sampath Kumar V, Marimuthu K, Subramani S, Sharma V, Bera J, Kotwal P. Prospective randomized trial comparing open reduction and internal fixation with minimally invasive reduction and percutaneous fixation in managing displaced intra-articular calcaneal fractures. Int Orthop. 2014;38:2505-12.
30. Backes M, Schepers T, Beerekamp MSH, Luitse JSK, Goslings JC, Schep NWL. Wound infections following open reduction and internal fixation of calcaneal fractures with an extended lateral approach. Int Orthop. 2014;38:767-73.

Medical Journal of Bakırköy

Med J Bakirkoy 2025;21(3):323-330
DOI: 10.4274/BMJ.galenos.2025.2025.7-7

Research

The Interplay Between Shoulder Pain, Upper Limb Function, and Respiratory Capacity in Wheelchair Athletes

Tekerlekli Sandalye Sporcularında Omuz Ağrısı, Üst Ekstremite Fonksiyonu ve Solunum Kapasitesi Arasındaki İlişki

Elif Tuğçe Durmuş[1], Ertuğrul Safran[2]

[1]Bezmialem Vakıf University Institute of Health Sciences, Department of Physiotherapy and Rehabilitation, İstanbul, Türkiye
[2]Bezmialem Vakıf University Faculty of Health Sciences, Department of Physiotherapy and Rehabilitation, İstanbul, Türkiye

ABSTRACT

Objective: To examine the relationship between shoulder pain, upper extremity functionality, and respiratory parameters in wheelchair athletes due to repetitive mechanical loading.

Methods: This cross-sectional study included 30 active wheelchair athletes. Shoulder pain was assessed using a pressure algometer. Upper extremity functionality was evaluated with the Quick Disabilities of the Arm, Shoulder and Hand (Quick-DASH) questionnaire, JAMAR hand dynamometer, and Blazepod reaction system. Respiratory function was measured using maximum inspiratory pressure (MIP), maximum expiratory pressure (MEP), and chest expansion.

Results: Significant positive correlations were found between algometry values and grip strength ($p<0.01$), as well as between grip strength and both MIP and MEP ($p<0.001$). A negative correlation was observed between Quick-DASH scores and grip strength ($p<0.01$), indicating that decreased upper extremity function is associated with reduced muscle strength. These findings suggest a close interaction between musculoskeletal pain, upper limb performance, and respiratory muscle strength.

Conclusion: The interplay between shoulder pain, upper limb function, and respiratory parameters highlights the importance of a multidimensional evaluation approach in wheelchair athletes. Addressing these interrelated factors may contribute to more effective training and rehabilitation strategies.

Keywords: Wheelchair sports, musculoskeletal pain, hand grip strength, respiratory function, disability evaluation, athletic performance

ÖZ

Amaç: Tekerlekli sandalye kullanan sporcularda tekrarlayan mekanik yüklenmeye bağlı olarak gelişebilen omuz ağrısı, üst ekstremite fonksiyonu ve solunum parametreleri arasındaki ilişkiyi incelemektir.

Gereç ve Yöntem: Kesitsel tasarımla yürütülen bu çalışmaya 30 aktif tekerlekli sandalye sporcusu dahil edildi. Omuz ağrısı basınç algometresiyle değerlendirildi. Üst ekstremite fonksiyonelliği Kol, Omuz ve El Engellilik Anketi-Kısa Form (QuickDASH), JAMAR el dinamometresi ve Blazepod reaksiyon sistemi ile ölçüldü. Solunum fonksiyonları; maksimum inspiratuvar basınç (MIP), maksimum ekspiratuvar basınç (MEP) ve göğüs çevresi ölçümleriyle değerlendirildi.

Bulgular: Algometre ve kavrama kuvveti arasında anlamlı pozitif korelasyon ($p<0,01$); kavrama kuvveti ile MIP ve MEP arasında ise daha güçlü pozitif korelasyonlar ($p<0,001$) bulundu. Quick-DASH skorları ile kavrama kuvveti arasında anlamlı negatif korelasyon ($p<0,01$) saptandı. Bu durum, üst ekstremite fonksiyonundaki azalmanın kas kuvvetinde düşüşle ilişkili olduğunu göstermektedir.

Sonuç: Tekerlekli sandalye sporcularında ağrı, üst ekstremite fonksiyonu ve solunum parametreleri arasındaki çok boyutlu etkileşim, değerlendirme ve rehabilitasyon süreçlerinde bütüncül bir yaklaşımın gerekliliğini ortaya koymaktadır. Bu alanlardaki ilişkilerin göz önünde bulundurulması, daha etkili sporcu sağlığı ve performans yönetimi açısından önemlidir.

Anahtar Kelimeler: Tekerlekli sandalye sporları, kas-iskelet sistemi ağrısı, el kavrama kuvveti, solunum fonksiyonu, özürlülük değerlendirmesi, atletik performans

Address for Correspondence: Ertuğrul Safran, Prof. Asst., Bezmialem Vakıf University Faculty of Health Sciences, Department of Physiotherapy and Rehabilitation, İstanbul, Türkiye
E-mail: esafran@bezmialem.edu.tr **ORCID ID:** orcid.org/0000-0002-6835-5428

Cite as: Durmuş ET, Safran E. The interplay between shoulder pain, upper limb function, and respiratory capacity in wheelchair athletes. Med J Bakirkoy. 2025;21(3):323-330

Received: 18.07.2025
Accepted: 07.08.2025
Publication Date: 03.09.2025

INTRODUCTION

Regular participation in physical activity by individuals who use wheelchairs has positive effects on various multidimensional health parameters, such as maintaining musculoskeletal health, enhancing cardiovascular endurance, supporting psychological well-being, and promoting social integration. Physical activity serves as a key tool in reducing secondary complications and increasing functional independence. Moreover, sport and exercise are not merely recreational pursuits but are considered fundamental components of the multidisciplinary rehabilitation process (1,2).

In this population, the upper extremities play a crucial role in ensuring mobility and facilitating activities of daily living. However, their intensive and repetitive use-especially during activities such as wheelchair propulsion, transfers, reaching, and maintaining balance-leads to mechanical overloading, particularly in the shoulder girdle. This can result in overuse syndromes, tendinopathies, musculoskeletal pain, and ultimately, functional loss and reduced quality of life (3,4). Upper extremity pain has been reported in 30% to 73% of individuals who use wheelchairs, with the shoulder, wrist, and elbow being the most affected regions. Among athletes, such pain disrupts training continuity, decreases performance capacity, and significantly impairs independent living skills (5,6).

Beyond pain, upper extremity functionality comprises not only muscle strength and joint range of motion, but also proprioception, reaction time, grip strength, and effective use in daily activities. Self-report questionnaires such as the Quick Disabilities of the Arm, Shoulder and Hand (Quick-DASH) are widely used to assess disability and functional capacity related to the upper limbs, offering valuable insights for guiding rehabilitation efforts (7,8).

In addition to their role in movement, upper extremity muscles contribute to trunk stabilization and function as accessory respiratory muscles. Key muscle groups—such as the pectoralis major, latissimus dorsi, and serratus anterior—are actively involved in both inspiratory and expiratory phases of respiration. Therefore, reduced upper extremity function may lead to weakened respiratory muscle strength and limited chest expansion. While pain, upper extremity functionality, and respiratory function have often been studied independently in the literature, a combined evaluation of these systems may form the basis of a more comprehensive clinical approach. Simultaneous assessment of these parameters allows for more individualized and balanced rehabilitation goals, while also improving understanding of inter-system interactions. In particular, the dual role of upper extremity muscles in both motor function and ventilation- in whom postural stabilization is compromised-highlights the significance of this relationship (9-12).

Recent studies emphasize that integrating pain, functionality, and respiratory parameters into the assessment process offers considerable advantages in clinical decision-making. Such a multidisciplinary evaluation approach supports more effective and personalized rehabilitation planning by elucidating the complex interactions among physiological systems (13,14).

In this context, a multidisciplinary evaluation approach is suggested, to enhance the clinical decision-making process in wheelchair users. The aim of this study is to identify the relationships between shoulder pain, upper extremity functionality, and respiratory parameters in wheelchair athletes, and to evaluate how these systems interact.

METHODS

Study Design and Participants

This study was conducted between August and October 2024 at the Bağcılar Municipality Disability Center in İstanbul. It was designed as a cross-sectional descriptive study and approval was received from the Bezmialem Vakıf University Non-Interventional Clinical Research Ethics Committee (approval no: 2024/260, date: 10.06.2024). The study was carried out in accordance with the principles of the Declaration of Helsinki and included 30 volunteer wheelchair athletes.

Inclusion criteria were: being between 18 and 50 years of age, using a wheelchair for at least one year, and actively participating in sports activities. Individuals who had undergone upper extremity surgery within the past year or had cognitive or communicative impairments were excluded from the study. Informed consent was obtained from all participants.

Pain Assessment

To evaluate musculoskeletal pain, a Baseline® digital algometer (dolorimeter) (Fabrication Enterprises Inc., New York, USA) was used. Measurements were performed bilaterally at the most sensitive points of the deltoid, biceps, and triceps muscles. The device applied mechanical pressure perpendicularly using a 1 cm^2 metal tip. When the participant first perceived pain, the application was stopped and the value displayed on the screen was recorded in Newtons (N). Three measurements were taken for each region, and the average value was used for analysis (15).

Functionality Assessment

Hand grip strength was measured using a Camry® digital hand dynamometer (Camry Electronic Co. Ltd., Guangdong,

China) as an objective performance indicator. Three trials were performed for each hand, and the highest value was used for analysis (Figure 1) (16).

Reaction time was assessed using the Blazepod® system (Blazepod Ltd., Tel Aviv, Israel), a validated and reliable tool that measures motor response time to visual stimuli (Figure 2) (17). Participants completed a 30-second light sensor test, during which their response time to the light stimuli and the number of missed light emitting diode (LED) signals were recorded.

Functionality was also subjectively assessed using the Quick-DASH questionnaire. The Quick-DASH is a valid and reliable short-form tool consisting of 11 items that evaluate functional limitations and symptoms related to the upper extremity (18,19).

Measurement of Respiratory Parameters

Respiratory muscle strength was assessed through maximum inspiratory pressure (MIP) and maximum expiratory pressure (MEP) measurements. The evaluation was conducted using a portable digital SONMOL respiratory muscle trainer device, in accordance with the guidelines of the American Thoracic Society and the European Respiratory Society (20). A nose clip was applied to each participant, and the test was performed using mouth breathing maneuvers. Among the three best attempts, the highest value was used for analysis.

Chest expansion was measured using a tape measure at the axillary, epigastric, and subcostal levels. The difference between maximum inspiration and expiration was recorded in centimeters. Each measurement was repeated three times, and the average was calculated for analysis (21).

Figure 1. Measurement of hand grip strength

Statistical Analysis

All data were analyzed using IBM SPSS Statistics version 21.0 (Armonk, New York, USA). The normality of continuous variables was assessed using the Shapiro-Wilk test. For variables with a parametric distribution, Pearson correlation analysis was used, while Spearman correlation analysis was applied for non-parametric variables. A significance level of $p<0.05$ was considered statistically significant. A priori power analysis was conducted based on Cohen's classification for a large effect size ($r=0.50$), an alpha level of 0.05, and a statistical power of 80%. The analysis indicated that a minimum of 30 participants would be sufficient for detecting significant correlations (22).

RESULTS

Demographic and Descriptive Data

Among the 30 athletes who participated in the study, 80% (n=24) were male and 20% (n=6) were female, indicating that the majority of participants were men. Regarding educational background, 46.67% (n=14) had completed high school, 20% (n=6) had completed elementary school, 16.67% (n=5) had completed an associate degree, 13.33% (n=4) had completed a bachelor's degree, and only 3.33% (n=1) held a doctoral degree. None of the participants held a master's degree. This distribution shows that the majority of participants had received education at the secondary level.

When categorized by sports disciplines, 33.33% (n=10), of the athletes participated in basketball and arm wrestling, with both groups represented equally. Archery and boccia

Figure 2. Measurement of reaction time

were each represented by 16.66% (n=5) of the participants. Regarding types of disabilities, 30% (n=9) had paraplegia, 26.67% (n=8) had cerebral palsy (CP), 26.67% (n=8) had spina bifida, 10% (n=3) had post-polio syndrome, and 3.33% (n=1) were an amputee. These results indicate that the majority of athletes in the study had congenital or early-onset neurologically based disabilities. Additional demographic information such as body mass index and duration of disability is presented in Table 1.

Algometer Values

Pain threshold values determined by algometric measurements of the deltoid, biceps, and triceps muscle groups were evaluated separately for the right and left sides among the participating athletes. In the deltoid muscle group, the mean pain threshold was found to be 7.9±2.42 kg/cm² on the right side and 7.8±2.26 kg/cm² on the left. The minimum values were 3.6 and 4.0 kg/cm², and the maximum values were 12.3 and 11.7 kg/cm², respectively. The similarity between the sides suggests no lateralized sensitivity difference in the deltoid muscle.

In the biceps muscle group, the mean pain threshold was 6.8±1.65 kg/cm² on the right and 6.5±1.78 kg/cm² on the left. While the pain threshold values of the right and left biceps were relatively close, the slightly lower average on the left side may indicate greater pain sensitivity in the non-dominant limb for some individuals.

For the triceps muscle group, the mean pain threshold was equal on both sides at 8.5 kg/cm². The standard deviations were ±2.59 on the right and ±2.5 on the left. This symmetrical distribution in the triceps suggests no significant difference in muscle strength or usage intensity between sides.

Upper Extremity Functionality Values

An evaluation of upper extremity functionality measurements among the participating athletes revealed that the mean hand grip strength was 35.1±15.14 kg for the right hand and 33.1±16.54 kg for the left hand. Minimum and maximum values ranged from 10 to 62.9 kg for the right hand and from 9 to 59.3 kg for the left hand.

Regarding reaction time data, participants extinguished an average of 32.8±10.75 LEDs with the right extremity (with a mean response time of 639.5±168.62 ms) and missed an average of 2.1±3.41 LEDs. With the left extremity, they extinguished an average of 33.7±9.77 LEDs, had an average response time of 635.3±136.61 ms, and missed 1.6±3.44 LEDs.

The average score obtained from the Quick-DASH questionnaire, which was used to assess functional status of the upper extremity, was 23.5±16.38. This average score indicates a low-to-moderate level of functional impairment. The high standard deviation suggests a wide variation in upper extremity functionality among participants, which may be influenced by differences in disability type and the nature of the sports disciplines practiced.

Respiratory Parameters

An analysis of the respiratory parameters of the participating athletes revealed that the mean MIP was 73.15±32.38 cmH_2O. This value falls within the reference range of 60-100 cmH_2O reported in the literature, suggesting that inspiratory muscle strength was generally within physiological limits. However, the minimum value was as low as 23.6 cmH_2O, indicating possible inspiratory muscle weakness in some individuals.

In contrast, the mean MEP was calculated as 58.4±26.42 cmH_2O. Since this average falls below the clinical threshold of 80 cmH_2O, it suggests insufficient expiratory muscle strength. The wide distribution range (17.3-141 cmH_2O) reveals notable inter-individual differences among participants. This finding is particularly important

Table 1. Descriptive statistics of athletes' demographic characteristics

	n		Min.	Max.	Mean±SD
Age (years)	30		18	49	31.2±9.69
Body weight (kg)	30		48	110	71.2±16.26
Height (cm)	30		132	189	168±12.44
BMI (kg/m²)	30		16.9	41.6	25.3±5.58
Years of athletic participation	30		1	23	7.3±6.22
Duration of disability (years)	30		6	47	22.3±10.21
Biceps circumference (cm)	30	Right Left	22 22	48 44	32.2±5.42 31.8±5.26
Forearm circumference (cm)	30	Right Left	21 20	35 36	27.5±3.41 26.9±3.67
Weekly training duration (hours)	30		10	20	15.3±4.79

Max.: Maximum, Min.: Minimum, n: Number of participants, SD: Standard deviation, BMI: Body mass index

in the context of weakened muscles responsible for trunk stabilization and effective coughing.

Thoracic mobility was assessed at the axillary, epigastric, and subcostal levels. The mean axillary expansion was 4.95±1.99 cm, slightly below the reference value (≥5 cm), but still close to the normal range. The mean epigastric expansion was measured at 4.33±1.87 cm, indicating abnormal thoracic mobility based on the <5 cm criterion. The mean subcostal expansion was 3.88±2.09 cm. This measurement not only falls within the abnormal range but also approaches the pathological threshold of 2.5 cm in some individuals.

Relationship Between Algometer, Upper Extremity Functional Parameters, and Respiratory Parameters

When examining the correlation between pain threshold values measured with the algometer and grip strength, significant positive correlations were observed across all muscle groups ($p<0.01$). Notably, a strong correlation was found between right deltoid ($r=0.655$, $p<0.001$) and right-hand grip strength, and between right biceps ($r=0.647$, $p<0.001$) and right-hand grip strength. Similarly, strong correlations were observed between left deltoid ($r=0.601$, $p<0.001$), left biceps ($r=0.618$, $p<0.001$), and left-hand grip strength. Although correlations in the triceps muscle groups were also significant (right triceps: $r=0.527$, left triceps: $r=0.500$), the correlation strength was slightly lower compared to the biceps and deltoid muscles (Table 2).

The correlation analyses revealed generally weak and statistically non-significant relationships between algometer-measured pain thresholds and reaction time or accuracy parameters ($p>0.05$), suggesting no strong direct association between pain perception and visual-motor responses. However, a statistically significant negative correlation was found specifically between the right biceps and the number of missed LEDs ($r=-0.369$, $p=0.045$).

Regarding the relationship between algometer values and Quick-DASH scores, all values showed significant correlations ($p<0.05$) except for the right triceps ($p>0.05$). Detailed results are presented in Table 3.

Table 2. Relationship between algometry and grip strength

	Grip strength (right) (r, p)	Grip strength (left) (r, p)
Deltoid (right)	0.655, <0.001	0.641, <0.001
Deltoid (left)	0.621, <0.001	0.601, <0.001
Biceps (right)	0.647, <0.001	0.704, <0.001
Biceps (left)	0.612, <0.001	0.618, <0.001
Triceps (right)	0.527, 0.003	0.533, 0.002
Triceps (left)	0.500, 0.005	0.498, 0.005

p: Statistical significance value, r: Pearson correlation coefficient

No significant correlations were observed between algometer parameters and respiratory muscle strength ($p>0.05$). Strong and significant positive correlations were found between both the right- and left-hand grip strength and MIP ($r=0.678$ and $r=0.660$, $p<0.001$) and MEP ($r=0.562$ and $r=0.592$, $p=0.001$).

Regarding Blazepod reaction time performance, moderate and significant negative correlations were found between MIP and MEP, and parameters such as right and left response time, and the number of missed LEDs. For example, right-side response time correlated with MIP ($r=-0.593$, $p=0.001$) and MEP ($r=-0.639$, $p=0.000$). These findings suggest that as respiratory muscle strength increases, reflexive response times and error rates decrease.

Finally, when analyzing correlations with chest expansion measurements (axillary, epigastric, and subcostal), a significant positive relationship was found particularly between subcostal expansion and right-hand grip strength ($r=0.421$, $p=0.021$) and left-hand grip strength ($r=0.579$, $p=0.001$). This suggests that trunk mobility may be associated with upper extremity strength.

DISCUSSION

In this study, the relationships between shoulder pain, upper extremity functionality, and respiratory parameters were comprehensively evaluated in wheelchair athletes. The findings revealed significant positive correlations between pain threshold and hand grip strength. This suggests that as muscle pain sensitivity decreases, grip strength increases; in other words, individuals who experience less pain may be able to utilize their motor functions more effectively (23,24). The negative correlation observed between Quick-DASH scores and pain threshold indicates that as subjective functional capacity decreases, pain sensitivity increases. This finding suggests that functional limitations in the upper extremity may be directly associated with pain. One of the most noteworthy findings of the study is the relationship established between respiratory muscle

Table 3. Relationship between algometry and quick DASH scores

	Quick DASH (r, p)
Deltoid (right)	-0.392, 0.032
Deltoid (left)	-0.472, 0.008
Biceps (right)	-0.482, 0.007
Biceps (left)	-0.477, 0.008
Triceps (right)	0.318, 0.087
Triceps (left)	-0.365, 0.047

p: Statistical significance value, r: Pearson correlation coefficient, DASH: Disabilities of the Arm, Shoulder and Hand

strength and upper extremity functionality. Significant and strong positive correlations were found between hand grip strength and both MIP and MEP. This supports the idea that upper extremity muscles play an active role not only in motor function but also in inspiratory and expiratory processes. In particular, the involvement of muscles such as the pectoralis major and latissimus dorsi in both grip strength and ventilatory function highlights the importance of a holistic, system-based approach to assessment.

Yalçın (25) reported that physical fitness parameters and sport-specific skills of the upper extremity differ among athletes with varying classification scores. In this context, the influence of disability types on classification scores in our study may be an important factor in explaining the variations observed in parameters such as motor skills and respiratory muscle strength (25). In a comparative study by Başar et al. (26), it was reported that individuals with CP demonstrated lower scores in muscle strength and anaerobic power compared to other groups. Considering the 26.7% prevalence of CP among participants in our study, this condition should be regarded as an important variable when interpreting findings related to upper extremity muscle strength and functionality. Additionally, Cömert et al. (27) in their study on the evaluation of upper extremity functions in wheelchair basketball players reported limitations in parameters such as grip strength, coordination, and range of motion in individuals with neurological disabilities such as CP and post-polio syndrome. These findings may directly influence our clinical outcomes, since these groups were represented by a substantial proportion—36.7%—in our study.

The respiratory parameter findings in our study are consistent with those reported by Pereira et al. (28), who stated that upper extremity muscles actively function as accessory respiratory muscles during inspiration and may be directly associated with respiratory function. At the same time, lower pain sensitivity in these muscles (i.e., higher pain thresholds) may contribute to more effective functional endurance and ventilatory support. Our findings are consistent with the literature emphasizing the dual role of shoulder girdle muscles in both motor performance and respiratory capacity. For instance, Curtis et al. (29) and Finley et al. (30) highlighted the limiting effects of shoulder pain on functional capacity, noting that muscle fatigue and pain in the deltoid and triceps muscle groups negatively impact wheelchair maneuverability and postural stability. The fact that these muscles also serve as accessory respiratory muscles suggests that increased pain levels may indirectly limit ventilatory performance. Although no significant relationship was found for the triceps muscle in our study (e.g., left triceps and MEP: p=0.555), this may indicate that the triceps muscle contributes less to both sensory and ventilatory functions compared to other muscle groups. The limited correlations observed between muscle pain thresholds and respiratory function suggest that this relationship is likely indirect and subject to individual variability. Samuelsson et al. (31) and Gironda et al. (32) reported that shoulder pain and upper extremity muscle dysfunction can restrict both daily living activities and functional capacity in individuals with spinal cord injury; and that upper extremity muscle dysfunction may also interact with ventilatory function.

In our study, when examining the relationship between grip strength and reaction time parameters in wheelchair basketball athletes, significant negative correlations were found, particularly between grip strength and the number of missed LEDs. For right-hand grip strength, significant correlations were identified with both the right and left missed LEDs ($r=-0.423$, $p=0.020$, and $r=-0.421$, $p=0.021$, respectively). These correlations were even more pronounced for left-hand grip strength ($r=-0.426$, $p=0.019$; $r=-0.477$, $p=0.008$). These findings suggest that grip strength is not only an indicator of physical capacity but also plays an important role in reactive tasks that require visual-motor integration and attention. The results are consistent with previous studies in the literature. Notably, Zacharakis (33) reported similar findings, demonstrating significant correlations between grip strength and technical skill tests. That study emphasized that grip strength is a key determinant of both muscle functionality and reaction time (33). Similarly, Molik et al. (34), in their evaluation of the relationship between anaerobic performance and on-court skill tests, found that upper extremity strength was positively associated with reaction time. Additionally, Yüksel (35), in a study conducted with elite athletes with physical disabilities, evaluated visual-reactive agility not merely as an output of the central nervous system, but also as a combination of motor response capacity and muscle control. This perspective supports the notion that the correlations observed in our study are not solely related to reflexive speed but are also associated with physiological power generation and the ability to apply that power effectively.

Study Limitations

This study has several limitations. As it was conducted with wheelchair athletes from a single rehabilitation and sports center, the findings may not fully reflect the diversity of the wider population of wheelchair users. Additionally, the use of performance-based and self-reported outcome measures may have been influenced by individual factors such as attention, motivation, or subjective perception.

Despite these limitations, the study aimed to provide a comprehensive, multidimensional evaluation of the interplay between pain, upper limb function, and respiratory capacity, and the results offer valuable insights for clinical and sports-related rehabilitation planning.

CONCLUSION

This study investigated the relationships among pain levels in the shoulder girdle muscles, upper extremity functionality, and respiratory parameters in wheelchair athletes using a holistic approach. The findings revealed significant positive correlations between pain thresholds and both grip strength and functional questionnaire scores. However, the relationships between pain levels and reaction time parameters were found to be weaker and more limited.

Strong correlations between upper extremity strength and respiratory muscle strength (MIP and MEP) suggest that these two systems are clinically interconnected. Moreover, the observed positive correlations between thoracic expansion and grip strength indicate that postural stability and trunk mobility may contribute to upper extremity performance. These results highlight the importance of considering these parameters collectively, especially in multidisciplinary rehabilitation processes.

Recommendations for Clinical Practice

When developing rehabilitation plans for wheelchair users, a comprehensive assessment should be conducted that includes not only pain evaluation but also respiratory function and upper extremity performance. Exercises aimed at strengthening respiratory muscles should be considered, as they may indirectly enhance upper extremity function. Regular monitoring of performance-based parameters such as grip strength and reaction time could improve training efficiency.

Recommendations for Future Research

Prospective studies with larger and more diverse samples may help clarify causal relationships between systems. Experimental research supported by objective tools such as electromyography, respiratory function tests, and postural analysis could enhance the validity of the findings. Subgroup analyses based on specific sports disciplines may help elucidate the effects of sport-specific loading patterns on the relationship between upper extremity function and overall health.

ETHICS

Ethics Committee Approval: It was designed as a cross-sectional descriptive study and approval was received from the Bezmialem Vakıf University Non-Interventional Clinical Research Ethics Committee (approval no: 2024/260, date: 10.06.2024).

Informed Consent: Informed consent was obtained from all participants.

FOOTNOTES

Authorship Contributions

Concept: E.T.D., E.S., Design: E.T.D., E.S., Data Collection or Processing: E.T.D., Analysis or Interpretation: E.T.D., E.S., Literature Search: E.T.D., E.S., Writing: E.T.D., E.S.

Conflict of Interest: No conflict of interest was declared by the authors.

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

REFERENCES

1. Perret C. Elite-adapted wheelchair sports performance: a systematic review. Disabil Rehabil. 2017;39:164-72.
2. Goosey-Tolfrey V. Wheelchair sport: a complete guide for athletes, coaches, and teachers: Human Kinetics; 2010.
3. Briley SJ, Vegter RJK, Goosey-Tolfrey VL, Mason BS. The longitudinal relationship between shoulder pain and altered wheelchair propulsion biomechanics of manual wheelchair users. J Biomech. 2021;126:110626.
4. Panpallı HG. Examination of shoulder pain, physical activity level, and quality of life in manual wheelchair eser athletes during the pandemic period (Master's thesis, Middle East Technical University (Turkey)).
5. Farì G, Latino F, Tafuri F, Dell'Anna L, Raele MV, Fai A, et al. Shoulder pain biomechanics, rehabilitation and prevention in wheelchair basketball players: a narrative review. Biomechanics. 2023;3:362-76.
6. Mayrhuber L, Rietveld T, de Vries W, van der Woude LH, de Groot S, Vegter RJ. A scoping review on shoulder injuries of wheelchair tennis players: potential risk-factors and musculoskeletal adaptations. Front Rehabil Sci. 2022;3:862233.
7. Galardini L, Coppari A, Pellicciari L, Ugolini A, Piscitelli D, La Porta F, et al. Minimal clinically important difference of the disabilities of the arm, shoulder and hand (DASH) and the shortened version of the DASH (QuickDASH) in people with musculoskeletal disorders: a systematic review and meta-analysis. Physical Therapy. 2024;104:pzae033.
8. Kennedy CA, Beaton DE, Smith P, Van Eerd D, Tang K, Inrig T, et al. Measurement properties of the Quick DASH (disabilities of the arm, shoulder and hand) outcome measure and cross-cultural adaptations of the Quick DASH: a systematic review. Qual Life Res. 2013;22:2509-47.
9. Maden Ç, Bayramlar K, Maden T, Sağlam M, Eltan SB, Keskin Ö. The relationships between upper extremity muscle strength and endurance with respiratory function, functional capacity and quality of life in children and adolescents with cystic fibrosis. Int J Ther Rehabil. 2023;30:1-9.
10. Sobierajska-Rek A, Wasilewska E, Śledzińska K, Jabłońska-Brudło J, Małgorzewicz S, Wasilewski A, et al. The association between the respiratory system and upper limb strength in males with Duchenne muscular dystrophy: a new field for intervention? Int J Environ Res Public Health. 2022;19:15675.

11. Lili L, Sunnerhagen KS, Rekand T, Alt Murphy M. Independence and upper extremity functioning after spinal cord injury: a cross-sectional study. Sci Rep. 2023;13:3148.
12. Woods A, Gustafson O, Williams M, Stiger R. The effects of inspiratory muscle training on inspiratory muscle strength, lung function and quality of life in adults with spinal cord injuries: a systematic review and Meta-analysis. Disabil Rehabil. 2023;45:2703-14.
13. Kovacs FM, Seco J, Royuela A, Barriga A, Zamora J. Prevalence and factors associated with a higher risk of neck and back pain among permanent wheelchair users: a cross-sectional study. Spinal Cord. 2018;56:392-405.
14. Soo Hoo JA, Kim H, Fram J, Lin YS, Page C, Easthausen I, et al. Shoulder pain and ultrasound findings: a comparison study of wheelchair athletes, nonathletic wheelchair users, and nonwheelchair users. PM R. 2022;14:551-60.
15. Fischer AA. Pressure algometry over normal muscles. Standard values, validity and reproducibility of pressure threshold. Pain. 1987;30:115-26.
16. Sánchez-Aranda L, Fernández-Ortega J, Martín-Fuentes I, Toval Á, Jurak G, Ruiz JR, et al. Reliability and concurrent validity of a low-cost handgrip dynamometer: the camry. Am J Occup Ther. 2025;79:7905205020.
17. de-Oliveira LA, Matos MV, Fernandes IG, Nascimento DA, da Silva-Grigoletto ME. Test-retest reliability of a visual-cognitive technology (BlazePod™) to measure response time. J Sports Sci Med. 2021;20:179-80.
18. Koldas Dogan S, Ay S, Evcik D, Baser O. Adaptation of Turkish version of the questionnaire Quick Disability of the arm, shoulder, and hand (Quick DASH) in patients with carpal tunnel syndrome. Clin Rheumatol. 2011;30:185-91.
19. Beaton DE, Wright JG, Katz JN, Group UEC. Development of the QuickDASH: comparison of three item-reduction approaches. J Bone Joint Surg Am. 2005;87:1038-46.
20. Society AT. ATS/ERS statement on respiratory muscle testing. Am J Respir Crit Care Med. 2002;166:518-624.
21. Debouche S, Pitance L, Robert A, Liistro G, Reychler G. Reliability and reproducibility of chest wall expansion measurement in young healthy adults. J Manipulative Physiol Ther. 2016;39:443-9.
22. Bujang MA, Omar ED, Foo DHP, Hon YK. Sample size determination for conducting a pilot study to assess reliability of a questionnaire. Restor Dent Endod. 2024;49:e3.
23. Hoo JS. Shoulder pain and the weight-bearing shoulder in the wheelchair athlete. Sports Med Arthrosc Rev. 2019;27:42-7.
24. Fullerton HD, Borckardt JJ, Alfano AP. Shoulder pain: a comparison of wheelchair athletes and nonathletic wheelchair users. Med Sci Sports Exerc. 2003 Dec;35(12):1958-61.
25. Yalçin Aİ. Farklı klasifikasyon puanlarına sahip tekerlekli sandalye basketbol oyuncularında üst ekstremite fiziksel uygunluk parametreleri ile spora özgü beceriler arasındaki ilişkinin incelenmesi: Sağlık Bilimleri Enstitüsü; 2015.
26. Başar S, Ergün N, Yiğiter Bayramlar K. A comparative study of muscle strength and anaerobic power of the young national and national junior wheelchair basketball players. Turk J Phys Med Rehabil. 2013;59.
27. Cömert E, Yıldırım NÜ, Ergun N. Evaluation of upper extremity functionality in wheelchair basketball sport. Turkiye Klinikleri J Sports Sci. 2010;2:62-9.
28. Pereira RN, Abreu MF, Gonçalves CB, Corrêa WF, Mizuhira DR, Moreno MA. Respiratory muscle strength and aerobic performance of wheelchair basketball players. Motriz: Revista de Educação Física. 2016;22:124-32.
29. Curtis KA, Drysdale GA, Lanza RD, Kolber M, Vitolo RS, West R. Shoulder pain in wheelchair users with tetraplegia and paraplegia. Arch Phys Med Rehabil. 1999;80:453-7.
30. Finley MA, Rasch EK, Keyser RE, Rodgers MM. The biomechanics of wheelchair propulsion in individuals with and without upper-limb impairment. J Rehabil Res Dev. 2004;41:385-94.
31. Samuelsson K, Tropp H, Gerdle B. Shoulder pain and its consequences in paraplegic spinal cord-injured, wheelchair users. Spinal Cord. 2004;42:41-6.
32. Gironda RJ, Clark M, Neugaard B, Nelson A. Upper limb pain in Anational sample of veterans with paraplegia. J Spinal Cord Med. 2004;27:120-7.
33. Zacharakis E. The effect of upper limb characteristics on palm strength, anaerobic power, and technical skills of wheelchair basketball players of varying classification. J Phys Educ Sport. 2020;20:584-91.
34. Molik B, Laskin J, Kosmol A, Marszałek J, Morgulec-Adamowicz N, Frick T. Relationships between anaerobic performance, field tests, and functional level of elite female wheelchair basketball athletes. Hum Mov. 2013;14:366-71.
35. Yüksel MF. Examination of reaction times of elite physically disabled badminton players. Int JSCS. 2017;5:319-27.

Medical Journal of Bakırköy

Med J Bakirkoy 2025;21(3):331-338
DOI: 10.4274/BMJ.galenos.2025.2025.2-7

Research

The Relationship Between University Students' Smoking Attitudes and Depression

Üniversite Öğrencilerinin Sigara İçme Tutumları ile Depresyon Arasındaki İlişki

Abdulkadir Aydın[1], Zeki Akkuş[2]

[1]Dicle University Atatürk Vocational School of Health Sciences, Department of Medical Services and Techniques, Diyarbakır, Türkiye

[2]Dicle University Faculty of Medicine, Department of Biostatistics and Medical İnformatics, Diyarbakır, Türkiye

ABSTRACT

Objective: It was aimed to investigate the smoking attitudes of the students, their relationship with the family, the reasons that increase smoking, their nicotine addiction levels, and to reveal whether smoking is related to depression and anxiety.

Methods: The population of the study is 412 students at the School of Health Vocational School. For this purpose, 343 (83.25%) students were reached. Data were collected by filling out surveys by students. Fagerström Test for Nicotine Dependence (FTND) was applied to evaluate students' demographic and smoking attitudes.

Results: We found that 76.11% of smoking students wanted to quit smoking within 1 year. We found a FTND total score of 4.17 and a moderate level of nicotine addiction. We found no statistical difference in Beck anxiety and depression scores in smokers between the two groups ($p>0.05$).

Conclusion: As a result, in this study, we found that the smoking rates of Vocational School students are lower than other universities in our country, but the smoking rates are similar compared to developed countries.

Keywords: Beck Inventory, Fagerström Test for Nicotine Dependence, reasons for smoking, smoking

ÖZ

Amaç: Bu çalışmada, üniversite öğrencilerinin sigara kullanımıyla ilgili tutumlarını, aile ilişkilerini, sigaraya başlamada etkili olan nedenleri, nikotin bağımlılık düzeylerini ve sigara kullanımının depresyon ve anksiyete ile ilişkili olup olmadığını incelemeyi amaçlandı.

Gereç ve Yöntem: Çalışmanın evrenini Sağlık Meslek Yüksekokulu'nda öğrenim gören 412 öğrenci oluşturmuş, bunlardan 343'üne (%83,25) ulaşılmıştır. Veriler, öğrenciler tarafından doldurulan anketler aracılığıyla toplanmıştır. Öğrencilerin sigara kullanım özellikleri ve bağımlılık düzeylerini değerlendirmek için Fagerström Nikotin Bağımlılık Testi (FNBT) uygulandı.

Bulgular: Sigara kullanan öğrencilerin %76,11'inin bir yıl içinde sigarayı bırakmak istediği bulundu. Ortalama FNBT puanı 4,17 olup, öğrencilerin orta düzeyde nikotin bağımlılığına sahip olduğu belirlendi. Sigara kullanan ve kullanmayan öğrenciler arasında Beck anksiyete ve depresyon skorları açısından istatistiksel olarak anlamlı bir fark bulunmadı ($p>0,05$).

Sonuç: Çalışma sonuçlarına göre, Sağlık Meslek Yüksekokulu öğrencilerinin sigara kullanma oranlarının ülkedeki diğer üniversitelere kıyasla daha düşük olduğu, ancak gelişmiş ülkelerle benzerlik gösterdiği görülmüştür.

Anahtar Kelimeler: Beck Envanteri, Fagerström Nikotin Bağımlılık Testi, sigara içme nedenleri, sigara içme

Address for Correspondence: Abdulkadir Aydın, Prof. Assoc., Dicle University Atatürk Vocational School of Health Sciences, Department of Medical Services and Techniques, Diyarbakır, Türkiye
E-mail: akcosut@hotmail.com **ORCID ID:** orcid.org/0000-0002-9316-448X

Cite as: Aydın A, Akkuş Z. The relationship between university students' smoking attitudes and depression. Med J Bakirkoy. 2025;21(3):331-338

Received: 20.02.2025
Accepted: 16.05.2025
Publication Date: 03.09.2025

INTRODUCTION

Today, tobacco is the single most important cause of death in the world. According to World Health Organization (WHO) data, tobacco causes the death of more than 5 million people a year. By 2030, this number will exceed 8 million. It is estimated that unless urgent measures are taken, tobacco use will cause the death of 1 billion people during this century (1). Although tobacco is used in the form of chewing, snuff, pipes, cigars, cigarettes, and hookahs, the most common form of consumption is cigarettes (2). Smoking causes diseases such as cancer, heart diseases, stroke, lung diseases, diabetes, and chronic obstructive pulmonary diseases (3). In this sense, although smoking is one of the most harmful habits, it is also considered a serious psychosocial problem in terms of its causes (4). According to the WHO 2018 health statistics report, smoking rates among men over the age of 15 are 65.1%, in Tunisia 53.3%, in Cuba 35.6%, in France 33.7%, in Japan 33.0%, in Germany 33.1%, in Türkiye 41.1% it is reported (5). In our country, 44.8% of men and 18.1% of women smoke (6). Although the age of starting smoking varies from society to society, studies have emphasized that, in general, 90% of smokers started smoking before the age of 18, and 99% started smoking before the age of 26 (7,8). Similarly, in our country, it is observed that the smoking trend begins in adolescence (9,10). There are manyreasons why young people between the ages of 14 and 26 start smoking. Adolescence, which includes the high school and university years, is a time of many stressful situations such as separation from the home environment and family members, adapting to the new environment, choosing friends, preparing for exams, becoming a candidate for a profession, finding a job after graduation, choosing a partner to share life with, economic problems, and improving social relations. It is considered a process in which depression and anxiety are experienced severely (11). In this process, the student starts smoking to cope with the problems he encounters. This attitude later turns into a physiological, psychological, and social addiction (12). For this purpose, there are many studies conducted in the world and in our country, especially on the use of tobacco and its products among university students, and their relationship with anxiety and depression (13,14). Smoking addiction is one of the rare health problems for which a "global epidemic alert" has been given by the WHO (15). For this purpose, many countries have aimed to take various measures. Since young people studying at universities are also the target group of the tobacco industry, the importance of intervention programs for this group is increasing (16). In our country, the "National Tobacco Control Program" was createdin 2006, and the use of tobacco and tobacco products was restricted in many public areas. Smoking was bannedespecially in places where the young population spends time,such as social and cultural venues, sports and entertainment places such as restaurants, coffee houses, cafeterias, beer halls, classrooms, private teaching institutions, and highway, railway, sea, and airline public transportation vehicles. With the adoption and implementation of this law, Türkiye became a country with smoke-free air space (17). In addition, as of 2019, new restrictions have been introduced on the packaging of tobacco products and the use of additives and stimulant compounds to reduce the use of tobacco and facilitate inhalation. It is one of the countries that has implemented all the necessary strategies within the scope of combating tobacco. In this study, we aimed to examine the smoking attitudes of university students studying health in nine different departments and to determine the relationship of these attitudes with anxiety and depression.

METHODS

The study was conducted in accordance with the World Medical Association Declaration of Helsinki and the Guideline for Good Clinical Practice. To conduct this study, approval was obtained from the Dicle University Faculty of Medicine Non-Interventional Clinical Research Ethics Committee (approval no: 64, date: 15.02.2018). This study was conducted as a cross-sectional and descriptiveon students of Dicle University Atatürk Vocational School of Health Services from March to June 2021. The population of the study was 412 first and second year students attending 9 associate degree programs at the School of Health Professions. However, we were able to reach 343 (83.25%) students. Written and verbal consent were obtained from the students who agreed to participate in the study. The questionnaire form consists of 3 parts: demographic questions and information questions about smoke. The Beck Depression Inventory (BDI) was used to evaluate the psychological state of all participants and Fagerström Test for Nicotine Dependence (FTND) used to determine the level of nicotine addiction. In the study, questions were asked about the demographic characteristics of the participating students, such as age and gender, as well as where they stayed, whether their parents smoke, their family income level, the number of smokers at home, and their awareness of the harms of smoking in order to evaluate their living environment and smoking status. If the participating student was a smoker, a survey was also conducted asking the age of starting smoking, the reason, the number of cigarettes smoked per day, whether smoking is harmful or not, andtheir desire to quit smoking. The BDI was used to

evaluate the psychological state of all participants in the last week to assess whether smoking had an impact on their psychological well-being. BDI is used to determine the subject's risk for depression and to measure changes in the level and severity of depressive symptoms. Its purpose is to determine the risk for depression and to measure the level and changes in severity of depressive symptoms. This form, which contains a total of 21 self-rating scales, provides a four-point Likert type measurement. Each item receives increasing points from 0 to 3 and the total score is obtained by adding them up. A high total score indicates a higher severity of depression (18). The adaptation, validity, and reliability of this inventory, which was developed by Beck et al. (18) for the Turkish population, was determined by Hisli (19).

We used the FTND to determine the level of nicotine addiction. FTND consists of 6 questions. Each question is given a score between 1 and 10 points. A high score indicates a high level of addiction. According to the total scores obtained from this test, nicotine addiction is evaluated in three groups as low (0-3 points), medium (4-6 points), and high (≥7 points) (20). In the study, the Turkish version of FTND was used by Uysal et al. (21).

Statistical Analysis

SPSS 25.0 statistical package (IBM Corp., Armonk, NY, USA) was used for analyses. Quantitative variables were presented as mean±standard deviation, and categorical variables were presented as number and percentage (%). The Kolmogorov-Smirnov test was used to check whether the data were normally distributed. Parametric tests were used for normally distributed data, and non-parametric tests were used for non-normally distributed data. The chi-square method was used for categorical variables. By checking whether the data comply with a normal distribution, Student's t-test was used if normal distribution was observed, and Mann-Whitney U test was used if normal distribution was not observed. Hypotheses were bidirectional, and $p<0.05$ was considered statistically significant.

RESULTS

Of the 343 students participating in the study, 54.8% (188) were female and 45.2% (155) were male, while the average age was 20.91±1.894. While 64.7% (222) of the students were staying at home, 35.3% (121) were staying in dormitories. We found that 77.0% (264) of the students smoked at home, while 52.5% (180) had fathers who smoked at home and 25.1% (86) had mothers who smoked at home. Only 31.2% (107) of the participating students had a normal or above family income level. 96.8% (332) of the students thought that smoking was harmful (Table 1). We found that 17.9% (12) of the students who smoked started smoking for the first time under the age of 10, while 35.8% (24) started smoking after the age of 19. We found that while 53.7% (36) of the students were regular drinkers, 46.3% (31) drank occasionally. While 59.7% (40) of the students smoked 10 cigarettes or less per day, only 4.5% (3) smoked . While 83.6% (56) of the students who smoke thought that smoking was harmful, 76.1% (51) stated that they were thinking of quitting smoking (Table 2).

Table 1. Demographic characteristics of students

	Smoker, n (%), (n=67)	Non-smoker, n (%), (n=276)	Total, n (%), (n=343)
Age	21.70+2.153 18/27	20.71+1.777 18/28	20.91+1.894 18/28
Gender			
Female Male	24 (12.8%) 43 (27.7%)	164 (87.2%) 112 (72.3%)	188 (54.8%) 155 (45.2%)
Place of residence			
Home Dormitory	44 (65.7%) 23 (34.3%)	178 (64.5%) 98 (35.5%)	222 (64.7%) 121 (35.3%)
Class			
1. Class 2. Class	28 (41.8%) 39 (58.2%)	122 (44.2%) 154 (55.8%)	150 (43.7%) 193 (56.3%)
Smoking status at home			
Yes No	63 (94.0%) 4 (6.0%)	201 (72.8%) 75 (27.2%)	264 (77.0%) 79 (23.0%)
Does the father smoke?			
Yes No	49 (73.1%) 18 (26.9%)	131 (47.4%) 145 (52.6%)	180 (52.5%) 163 (47.5%)

Table 1. Continued

	Smoker, n (%), (n=67)	Non-smoker, n (%), (n=276)	Total, n (%), (n=343)
Does the mother smoke?			
Yes	19 (28.4%)	67 (24.2%)	86 (25.1%)
No	48 (71.6%)	209 (75.8%)	257 (74.9%)
Number of smokers at home			
0	0 (0.00%)	73 (26.4%)	73 (21.3%)
1	8 (11.9%)	101 (36.5%)	109 (31.8%)
2	34 (50.7%)	48 (17.3%)	82 (23.9%)
3	8 (11.9%)	31 (11.2%)	39 (11.4%)
4	14 (20.9%)	13 (4.7%)	27 (7.9%)
5	3 (4.5%)	10 (3.6%)	13 (3.8%)
Family income level			
0-1400 TL	23 (34.3%)	73 (26.4%)	96 (28.0%)
1401-2800TL	30 (44.8%)	110 (39.9%)	140 (40.8%)
2801-4200 TL	14 (20.9%)	69 (25.00%)	83 (24.2%)
4200TL over	0 (0.00%)	24 (8.7%)	24 (7.0%)
Do you think smoking is harmful?			
Yes	56 (83.6%)	276 (100%)	332 (96.8%)
No	11 (16.4%)	0 (0.00%)	11 (3.2%)

TL: Turkish lira

We found the Fagerström total score to be 4.17 (moderate nicotine dependence). The Beck anxiety total score was 9.22±3.16 in smokers and 8.94±3.33 in non-smokers. The Beck depression total scorewas 6.73±1.95 in smokers and 6.51±3.00 in nonsmokers. We found no statistical difference in anxiety and depression scores between the two groups ($p>0.05$). We found total anxiety and stress scores to be similar and borderline in the two groups (Table 3). While there was no correlation between the FTND and the Beck Anxiety Inventory (BAI), we found a positive correlation between the FTND and the BDI ($p<0.05$) (Table 4). When the demographic characteristics of the students were examined regarding smoking, we found statistically that males smoke more than females, students whose fathers smoke smoke more than those whose fathers do not smoke, and students whose fathers smoke at home smoke more than those who do not smoke at home ($p<0.05$). In addition, we found that Beck depression-anxiety, income level and mother's smoking did not have a statistically significant effect on students' smoking ($p>0.05$) (Table 5). We found that 43.3% of smokers started smoking due to stress and sadness, 82% said smoking calmed them down, and 73.1% had an increased desire to smoke with tea or coffee, and after meals. We found that the factor most strongly associated with smoking was stress, with 50.7% (34) (Table 6).

Table 2. Students' opinions about smoking

	n	%
What age did you smoke for the first time?		
5-10 age	12	17.9
11-15 age	14	20.9
16-18 yaş	17	25.4
19 age and over	24	35.8
Being a regular drinker		
Regular	36	53.7
Sometimes	31	46.3
Number of cigarettes smoked per day		
0-10 piece	40	59.7
11-20 piece	18	26.9
21-30 piece	6	9.0
31 and piece	3	4.5
Changes in smoking amount over the years		
İncreased	27	40.3
Decreased	20	29.9
Not change	20	29.9
Do you think smoking is harmful?		
Yes	56	83.6
No	11	16.4
Do you want to quit smoking within a year?		
Yes	51	76.1
No	16	23.9

Table 3. Addiction, anxiety and depression levels of smoking and non-smoking students

	Smoker	Non-smoker	p-value
FTND	4.17*< 5	0.00	
Beck Anxiety Scale	9.22<10	8.94<10	0.772
Beck Depression Scale	6.73<7	6.50<7	0.811

*Low level of nicotine addiction, FTND: Fagerström Test for Nicotine Dependence

Table 4. Correlation between anxiety-depression and addiction

	FTND
Beck Anxiety Scale	0.071*
Beck Depression Scale	0.044*

*Spearman's rho correlation, FTND: Fagerström Test for Nicotine Dependence

Table 5. Evaluation of smoking status and demographic, clinical, anxiety and depression conditions

	Smoking status (p)
Gender (female-male)	0.000
Father's smoking status (smoker-non-smoker)	0.000
Mother's smoking status (smoker-non-smoker)	0.489
Family income level (smoker or non-smoker)	0.052
Beck anxiety (smoker non-smoker)	0.531
Beck depression (smoker non-smoker)	0.455
Smoking status at home (smoker-non-smoker)	0.000

Pearson chi-square test, Student's t-test

DISCUSSION

In recent years, there has been an increase in smoking rates among university students due to various reasons, and in addition, smoking attitudes among high school and secondary school students have become more positive. The prevalence of smoking among university students in Türkiye has been reported as 18-48%, and in Health Vocational Schools as 37.5-55% (22,23). It was reported that in studies evaluating tobacco addiction among medical faculty students in our country, the smoking rate was between 17% and 52.6%, and in studies conducted in European countries, the smoking rate of medical faculty students was between 11% and 61% (24). In a study conducted in 5 countries (Belarus, Lithuania, Poland, Russia, Slovakia), between 2017-2018, and examining the smoking rates of 14,352 students, including 8,800 medical students, it was found that the overall smoking rate was 66.1%, while the smoking rate among medical students was 68.9%. For students studying in non-medical departments, this rate was determined as 61.8% (25). In a study investigating the smoking rate of nursing students studying in France, United States of America, Spain, Australia, and Italy, the prevalence of current smoking was reported as 26.6% with a confidence interval of (22.9-30.4) (26).

Studies, emphasizing, that individuals typically begin smoking during high school (27). In addition, a study on the age of starting smoking emphasized that male students started smoking at 14.84 years, whilefemale students started at 16.13 years (28). n our study, we found that 64.2% of the participating students started smoking under the age of 18. We also found that 17.9% of individuals started smoking by age 10 or younger. These values we found are parallel to the findings in the literature. The group under the age of 18 in our study represents high school and secondary school

Table 6. Examining the reasons for starting to smoke and increasing smoking

	C	EP	Prove yourself	W	RTB	S-S	F
Reason for starting smoking	20.9% (14)	4.5% (3)	4.5% (3)	14.9% (10)	9.0% (6)	43.3% (29)	32.8% (22)
	Sedation		**Attention gathering**		**Losing weight**	**Relieve loneliness**	
Reason to continue smoking	82.1% (55)		17.9% (12)		4.5% (3)	14.9% (10)	
	Tea	**After meal**	**Stress**		**Coffee**	**Alcohol**	**Friend**
Reasons that increase the desire to smoke	41.8% (28)	47.8% (32)	50.7% (34)		31.3% (21)	3% (2)	13.4% (9)

C: Curiosity, EP: Environmental pressure, W: Wannabe, RTB: Reaction to the ban, S-S: Stress-Sadness, F: Friend

students. Reasons for starting smoking are wide-ranging. In his study, Korkmaz et al. (4) stated that the reasons for starting smoking were 28.6% for social influence, 22% for curiosity, 18.2% for friend encouragement, and 20.4% for other reasons. Another study conducted on university students identified the reasons for starting smoking as 46.3% family problems, 40% school problems, and 33.9% curiosity (29). In another study, it was reported that one of the main reasons for smoking among students was influence from friends, with a rate of 24.4% (30). Similarly, in another study conducted in our country, the first reason why students start smoking is because their close relatives or friends smoke (42.6%), followed by "stress" (24.6%) and "curiosity, wannabe" (18.9%) (23). In our study, we stated that participants could choose more than one option when the reason for starting smoking was questioned. Our study found that the main reason for smoking at 43.3% was stress and sadness. However, we found that 32.8% were close relatives-friends, 20.9% were curiosities, and 14.9% were wannabes.

When the literature is examined, the rate of male students who smoke has been found to be higher than that of female students in many studies (31). In a study of university students, the smoking rate was found to be 18.9% among male students and 5% among female students. In addition, the smoking rate of 504 students, which consists entirely of female students, in the nursing-midwifery department of the vocational school of health was reported to be 12.3% (29). Although these reported findings are similar, the smoking rates of both male and female students were found to be lower than those in our study. In our study, we found the smoking rate to be 27.7% among male students and 12.8% among female students.

The smoking status of role model mothers and fathers in the family can have negative effects on their children who are studying. In this sense, a study showed that children whose parentssmoke 4 times more than those whose parents do not smoke (32). Similarly, in a meta-analysis that included 58 studies, the effect of parental smoking on the child's initiation of smoking was examined, and it was shown that the rate of smoking initiation in children increased by 2.73 times when both parents smoked (33). In our study, we questioned the smoking status of the mother and father separately. We found that 73.1% of the fathers and 28.4% of the mothers of the students smoke, while 47.4% of the fathers and 24.2% of the mothers of the students do not smoke. According to the results of our study, we can say that both mothers' and fathers' smoking have negative effects on other members of the family and that we obtained similarities with the results in the literature. In addition, when the results of the smoking attitudes of mothers and fathers are examined in our study, it can be said that students are more negatively affected by their fathers' smoking.

Contradictory results are obtained when studies examining whether smoking causes a change in anxiety and depression levels are reviewed (34-37). In a study conducted for this purpose, it was shown that there was no significant difference in anxiety, depression, perceived fatigue levels, attention deficit and hyperactivity disorder between smoking and non-smoking university students. However, although high smoking rates have been reported in people with anxiety and depression disorders (35), it has been emphasized that the relationship between smoking and anxiety and depression may be two-way, and that smoking occasionally is used to relieve symptoms; over time, this condition may worsen (38). In other studies, smoking is thought to alleviate the negative effects of smoking and therefore smoking may cause symptoms of anxiety and depression (36,37). In a systematic review including 148 studies, almost half of the studies reported that initial depression/anxiety was associated with some form of later smoking behavior, and more than a third of the studies reported that smoking exposure was associated with subsequent depression/anxiety (34).However, it has also been reported that there are few studies that support a bidirectional model of smoking and anxiety, and the results of these studies are invalid (34). We used the BAI or BDI in our study. According to the data we obtained in our study, we could not find a statistical difference in the levels of anxiety and depression between smoking and non-smoking students. In addition, according to the data of the FTND, which we used to evaluate the addiction levels of smoking students we found the level to be low (4.17<5). One of the two most important outcomes of our study is that 83.6% of the students who smoke think that smoking is harmful; the other is that 76.1% of them want to quit smoking within a year.

CONCLUSION

As a result of our study and the literature we reviewed, we found that smoking is common among university students. University students are seen as the group with the highest risk of exposure to smoke. First of all, it is necessary to produce solutions to prevent all other substance use, especially smoking, throughout the education system. Data from previous studies suggest that some of the measures takenin our country are effective. We believe that compulsory or elective courses covering the harms of tobaccoand other substance use in university curricula can be a protective measure. We also think that meetings, symposiums and

raising awareness for families that emphasize the negative effects of parents who are role models in the family, on students will be beneficial.

ETHICS

Ethics Committee Approval: To conduct this study, approval was obtained from the Dicle University Faculty of Medicine Non-Interventional Clinical Research Ethics Committee (approval no: 64, date: 15.02.2018).

Informed Consent: Written and verbal consent were obtained from the students who agreed to participate in the study.

FOOTNOTES

Authorship Contributions

Concept: Z.A., Design: A.A., Data Collection or Processing: A.A., Analysis or Interpretation: Z.A., Literature Search: A.A., Writing: A.A.

Conflict of Interest: No conflict of interest was declared by the authors.

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

REFERENCES

1. World Health Organization. WHO report on the global tobacco epidemic, 2020: protecting youth from industry manipulation and preventing them from tobacco and nicotine use. Geneva: World Health Organization; 2020.
2. Aktaş A, Hıdıroğlu S, Karavuş, M. Üniversite öğrencilerinin nargile içme konusundaki bilgi, tutum ve davranışları. Fırat Tıp Derg. 2018;23:68-72.
3. World Health Organization. Who report on the global tobacco epidemic. 2017. Access address: https://www. who. int/news-room/factsheets/detail/tobacco.
4. Korkmaz M, Ersoy S, Özkahraman Ş, Duran ET, Uslusoy EÇ, Orak S, Orhan H. Süleyman Demirel Üniversitesi öğrencilerinin tütün mamulleri-alkol kullanım durumları ve sigaraya yaklaşımları. Med J Suleyman Demirel University. 2013;20.
5. World Health Organization. World health statistics 2018. Geneva: World Health Organization; 2018.
6. Sezer RE. Türkiye'de yetişkinlerin tütün ürünü kullanma durumlarında değişim (1993–2019). In: Türkiye Sağlık Raporu. Ankara: Hipokrat Yayınevi; 2020:1017-9.
7. US Department of Health Human Services. Preventing tobacco use among youth and young adults: a report of the Surgeon General. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention; 2012.
8. Rigotti NA, Lee JE, Wechsler H. US college students' use of tobacco products: results of a national survey. JAMA. 2000;284:699-705.
9. Boyacı H, Çorapçıoğlu A, Ilgazlı A, Başyiğit İ, Yıldız F. Kocaeli Üniversitesi öğrencilerinin sigara içme alışkanlıklarının değerlendirilmesi. Solunum Hastalıkları. 2003;14:169-75.
10. Tanrikulu AÇ, Çarman KB, Palanci Y, Çetin D, Karaca M. The prevalence of cigarette smoking among university students in Kars and risk factors. Thorac Res Pract. 2009;10:101.
11. Villatte A, Marcotte D, Potvin A. Correlates of depression in first-year college students. Can J High Educ. 2017;47:114-36.
12. Çilingir D, Hintistan S, Öztürk H. Smoking habits and affecting factors of students of health school. Gümüşhane Univ J Health Sci. 2012;1:69-87.
13. Vatan İ, Ocakoğlu H, İrgil E. Uludağ Üniversitesi Tıp Fakültesi öğrencilerinde sigara içme durumunun değerlendirilmesi. TAF Prev Med Bull. 2009;8.
14. Kartal M, Mıdık Ö,Büyükakkuş A. Tobacco smoking and its effect on quality of life of medical students in Ondokuz Mayıs University. Thorac Res Pract. 2012;13:11-7.
15. World Health Organization. The WHO framework convention on tobacco control: an overview [Internet]. Geneva: WHO; [cited 2020 Nov 21]. Available from: https://www.who.int/fctc/en/
16. Aslan D. Tütün kontrolü için "tütünsüz" üniversite modelleri. Sürekli Tıp Eğitimi Derg. 2019;28:23-5.
17. Saraçoğlu S, Öztürk F. An evaluation on tobacco control policies and tobacco consumption in Turkey. Politik Ekonomik Kuram. 2020;4:20-44.
18. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry. 1961;4:561-71.
19. Hisli N. Beck depresyon envanterinin universite ogrencileri icin gecerliligi, guvenilirligi (a reliability and validity study of Beck Depression Inventory in a university student sample). J Psychol. 1989;7:3-13.
20. Göçmen H, Savaş NN, Ceylan E, Ursavaş A, Karadağ M, Coşkun F, et al. Bursa'da sigara içimi ve nikotin bağımlılığı ile ilişkili faktörler. İzmir Göğ Hast Derg. 2011;25:1-8.
21. Uysal MA, Kadakal F, Karşidağ C, Bayram NG, Uysal O, Yilmaz V. Fagerstrom test for nicotine dependence: reliability in a Turkish sample and factor analysis. Tuberk Toraks. 2004;52:115-21.
22. Hassoy H, Ergin I, Davas A, Durusoy R, Karababa AO. Sağlık meslek yüksek okulu öğrencilerinde sigara, nargile, sarma tütün kullanımını etkileyen faktörlerin belirlenmesi ve öğrencilerin sigara, nargile, sarma tütüne başlama ve sürdürme konusundaki görüşleri. Solunum Derg. *2011;13:91-9.*
23. Kutlu R, Vatansev C, Demirbaş N, Taşer S. Tıp fakültesi öğrencilerinde tütün ve tütün ürünleri kullanım sıklığı. TJFMPC. 2019;13:219-6.
24. Çilekar Ş, Dumanlı A, Öz G, Günay E. Hastanemizde tıp fakültesi öğrencilerinin tütün kullanma durumları. Bozok Tıp Derg. 2019;9(4):22-26.
25. Brożek GM, Jankowski M, Lawson JA, Shpakou A, Poznański M, Zielonka TM, et al. The prevalence of cigarette and e-cigarette smoking among students in central and eastern europe-results of the YUPESS study. Int J Environ Res Public Health. 2019;16:2297.
26. Nilan K, McKeever TM, McNeill A, Raw M, Murray RL. Prevalence of tobacco use in healthcare workers: a systematic review and meta-analysis. PLoS One. 2019;14:e0220168.
27. Öz B, Alkevli A. Öğrencilerin madde kullanımı ve bağımlılığında etkili olan faktörlere bakışının demografik özelliklere göre incelenmesi: Çukurova Üniversitesi örneği. Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Derg. 2018;13:29-43.
28. Yılmaz G, İbiş S, Aktuğ ZB. Üniversite öğrencilerinin sigara, alkol ve uyuşturucu madde kullanma durumlari. Niğde Ömer Halisdemir Üniversitesi Sosyal Bilimler Enstitüsü Derg. 2020;2:73-80.

29. Koca B, Oğuzöncül AF. The effects of smoking, alcohol, drugs, the factors effecting drug usage, and social family support to the students at health institution of higher education at Inonu University. Med J Kocaeli. 2015;4:4-13.
30. Oğuz S, Çamcı G, Kazan M. The prevalence of cigarette smoking and knowing status for diseases caused by smoking among university students. Med J Van. 2018;5:332-7.
31. Nasser AMA, Zhang X. Knowledge and factors related to smoking among university students at Hodeidah University, Yemen. Tob Induc Dis. 2019;17:42.
32. den Exter Blokland EA, Engels RC, Hale WW 3rd, Meeus W, Willemsen MC. Lifetime parental smoking history and cessation and early adolescent smoking behavior. Prev Med. 2004;38:359-68.
33. Leonardi-Bee J, Jere ML, Britton J. Exposure to parental and sibling smoking and the risk of smoking uptake in childhood and adolescence: a systematic review and meta-analysis. Thorax. 2011;66:847-55.
34. Fluharty M, Taylor AE, Grabski M, Munafò MR. The association of cigarette smoking with depression and anxiety: a systematic review. Nicotine Tob Res. 2017;19:3-13.
35. Mykletun A, Overland S, Aarø LE, Liabø HM, Stewart R. Smoking in relation to anxiety and depression: evidence from a large population survey: the HUNT study. Eur Psychiatry. 2008;23:77-84.
36. Chaiton MO, Cohen JE, O'Loughlin J, Rehm J. A systematic review of longitudinal studies on the association between depression and smoking in adolescents. BMC Public Health. 2009;9:356.
37. Boden JM, Fergusson DM, Horwood LJ. Cigarette smoking and depression: tests of causal linkages using a longitudinal birth cohort. Br J Psychiatry. 2010;196:440-6.
38. Munafò MR, Araya R. Cigarette smoking and depression: a question of causation. Br J Psychiatry. 2010;196:425-6.

Medical Journal of Bakırköy

Med J Bakirkoy 2025;21(3):339-342
DOI: 10.4274/BMJ.galenos.2025.2025.1-2

Case Report

Hypereosinophilic Syndrome Masquerading as a Dialyzer Reaction: A Great Mimicker

Diyalizör Reaksiyonu Gibi Görünen Hipereozinofilik Sendrom: Harika Bir Taklitçi

Yusra Hashim[1], Siti Nur Hakimah Hashim[1], Kuan Yee Lim[1,2], Nur Ezzaty Mohammad Kazmin[3], Muhammad Yusuf Abu Shamsi[2], Lydia Kamaruzaman[1,2], Rozita Mohd[1,2]

[1]Universiti Kebangsaan Malaysia Faculty of Medicine, Department of Internal Medicine, Bangi Selangor, Malaysia
[2]Universiti Kebangsaan Malaysia, Hospital Canselor Tuanku Muhriz, Department of Internal Medicine, Kuala Lumpur, Malaysia
[3]Universiti Sains Islam Malaysia Faculty of Medicine, Department of Internal Medicine, Negeri Sembilan, Malaysia

ABSTRACT

Hypereosinophilic syndrome is a disorder characterized by the persistent marked elevation of eosinophils, which eventually leads to the inflammatory mediator release resulting in eosinophilic infiltrations in target organs. The occurrence of hypereosinophilic syndrome among end stage renal disease (ESRD) patients is rarely documented. As hypereosinophilic syndrome most commonly presents as intradialytic hypotension, it may mimic dialyzer reactions among ESRD patients with regular haemodialysis; hence, making the diagnosis and initiation of treatment challenging for attending physicians. We report a case of hypereosinophilic syndrome that presented with recurrent episodes of intradialytic hypotension mimicking a dialyzer reaction.

Keywords: Dialyzer reaction, ESRD, haemodialysis, hypereosinophilic syndrome, mimicker

ÖZ

Hipereozinofilik sendrom, eozinofillerin kalıcı belirgin yükselmesiyle karakterize bir hastalıktır ve bu da sonunda hedef organlarda eozinofilik enfiltrasyonlarla sonuçlanan enflamatuvar mediatör salınımına yol açar. Son dönem böbrek hastalığı (SDBH) hastalarında hipereozinofilik sendromun ortaya çıkması nadiren belgelenir ve raporlanır. Hipereozinofilik sendrom en sık intradiyalitik hipotansiyon olarak ortaya çıktığından, düzenli hemodiyaliz uygulanan SDBH hastalarında diyalizör reaksiyonlarını taklit edebilir ve bu nedenle tanıyı ve tedavinin başlatılmasını ilgili hekimler için zorlaştırır. Diyalizör reaksiyonunu taklit eden tekrarlayan intradiyalitik hipotansiyon ataklarıyla ortaya çıkan bir hipereozinofilik sendrom olgusunu bildiriyoruz.

Anahtar Kelimeler: Diyalizör reaksiyonu, SDBH, hemodiyaliz, hipereozinofilik sendrom, taklitçi

Address for Correspondence: Kuan Yee Lim MD, Universiti Kebangsaan Malaysia Faculty of Medicine, Department of Internal Medicine, Bangi Selangor; Universiti Kebangsaan Malaysia, Hospital Canselor Tuanku Muhriz, Department of Internal Medicine, Kuala Lumpur, Malaysia
E-mail: limkuanyee@gmail.com **ORCID ID:** orcid.org/0000-0003-2671-3722

Cite as: Hashim Y, Hashim SNH, Lim KY, Kazmin NEM, Shamsi MYA, Kamaruzaman L, et al. Hypereosinophilic syndrome masquerading as a dialyzer reaction: a great mimicker. Med J Bakirkoy. 2025;21(3):339-342

Received: 03.01.2025
Accepted: 28.04.2025
Publication Date: 03.09.2025

INTRODUCTION

Hypereosinophilic syndrome is a rare disorder characterized by hypereosinophilia with absolute eosinophil count of more than 1.5x10^9/L (1). It is an umbrella term used to describe diseases leading to peripheral blood eosinophilia, the causes of which include eosinophilic neoplasm, drug hypersensitivity reactions, helminth infections, and idiopathic causes (2). Hypereosinophilic syndrome has a wide range of manifestations, from non-specific subtle pruritus, rashes, and flushing, to life-threatening angioedema and shortness of breath, and may lead to multiorgan failure due to eosinophilic infiltrations (3). This can mimic intradialytic hypotension due to dialyzer reactions during haemodialysis in end stage renal disease (ESRD) patients (4).

CASE REPORT

A 51-year-old woman presented with hypotension ten minutes into haemodialysis, preceded by a week-long duration of generalized pruritus and skin rash. She was a known case of ESRD secondary to diabetic kidney disease, who had been on long-term regular haemodialysis for a month, using a polyethersulfone dialyzer. She had no known food or drug allergies and there was no previous episode of anaphylaxis. Her physical examination revealed a generalized pruritic macular erythematous skin rash. Her blood pressure was 138/80 mmHg, and her pulse rate was 98 beats per minute. Other systemic examinations were unremarkable. Her initial blood investigations revealed a significantly elevated eosinophil count of 18.4x10^9/L and a deranged renal profile that is consistent with ESRD. Other blood investigations, including liver function tests, were unremarkable.

She was admitted and treated for acute endogenous eczema with topical steroids, resulting in a significant resolution of her skin rashes. However, she had recurrent episodes of intradialytic hypotension with severe nausea and hot flushes after five-to-ten minutes into each haemodialysis session, with a persistently elevated eosinophil count ranging from 14.1 to 18.4x10^9/L (Figure 1). Further investigations to look for secondary causes of hypereosinophilia, including stool analysis for parasites and peripheral blood film, were unremarkable.

Despite using gentler modalities of haemodialysis including sustained low efficacy dialysis and continuous veno-venous haemodialysis, and various types of dialyzers including a polysulfone membrane (FX 1.4), polyethersulfone and polyvinylpyrrolidone blend (Theranova 400) and copolymer of acrylonitrile and sodium methallyl sulfonate (AN69), she persistently experienced hypotension, severe nausea, and hot flushes five-to-ten minutes into each haemodialysis session. Eventually, she was treated with oral prednisolone at a dose of 1 mg/kg for idiopathic hypereosinophilic syndrome, which drastically reduced her eosinophil count to a normal range. A trial of haemodialysis with a polyethersulfone dialyzer, conducted two days later, was uneventful. She was discharged with a tapering dose of prednisolone, and was able to tolerate the subsequent haemodialysis sessions without hypotension and rebound eosinophilia.

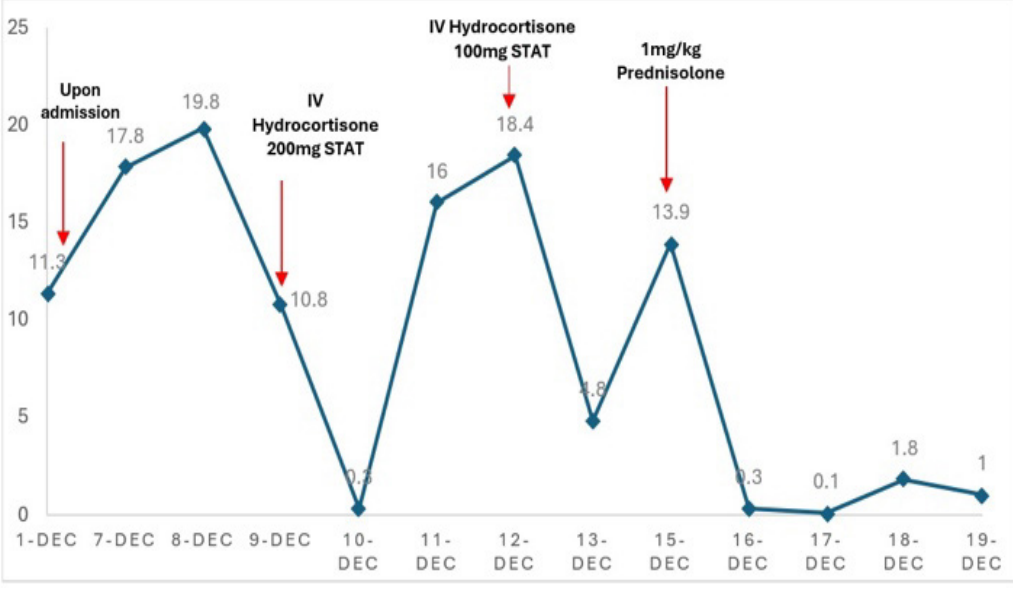

Figure 1. Eosinophil count trend throughout admission

She has given consent for this case to be reported and published.

DISCUSSION

Hypereosinophilia is a relatively common finding in daily clinical practice, and it is frequently associated with constellations of signs and symptoms that provide valuable clues for establishing the diagnosis. Hypereosinophilia is not a rare occurrence among ESRD patients receiving haemodialysis due to the reactions towards the haemodialysis circuit. It is often associated with the type of dialyzer used. With the emergence of modern membranes, hypereosinophilia among ESRD patients has become less common (5).

The cause of recurrent hypotension and gastrointestinal symptoms immediately after the initiation of dialysis in our case had been initially postulated to be due to hypersensitivity reactions to the dialyzer. Dialyzer reactions refer to every abnormal sequela following the initiation of haemodialysis and are a result of the interactions between blood and its constituents with the dialyzer membrane. Dialyzer reaction can be divided into two types: type A reaction usually occurs immediately after the commencement of dialysis and attributed to preformed antibodies against an allergen leading to immunoglobulin E (IgE)-mediated reactions, whereas type B reaction is delayed and non-IgE-mediated, resulting in milder and self-limiting symptoms (6,7).

The most common dialysis membrane reactions occur with ethylene oxide membranes and less commonly with biocompatible membranes. The relative risk of hypersensitivity reaction with synthetic membranes is 10-20% higher than that of cellulose membranes (8). Our case had been using a polyethylsulfone membrane for haemodialysis and she developed immediate hypersensitivity reactions five to ten minutes into haemodialysis. However, despite changes in dialyzer, her symptoms did not improve. Due to the persistent elevation of eosinophils, the diagnosis of hypereosinophilic syndrome was made in the absence of secondary causes.

Among ESRD patients, the common manifestations of hypereosinophilic syndrome include intradialytic hypotension with myriad other target organ damage such as chest pain, abdominal cramps, and generalized pruritus, which can be treated with systemic corticosteroids (9). Systemic corticosteroids are the gold standard treatment for hypereosinophilic syndrome with resolution of symptoms and normalization of blood parameters in 85% of cases, but the exact dosing and duration of corticosteroids remain controversial (2).

Although corticosteroids remain the first-line treatment for the management of idiopathic hypereosinophilic syndrome for induction and maintenance during acute settings, there are certain groups of patients who cannot tolerate high-dose corticosteroids, and will require initiation of a second-line agent. The use of hydroxyurea, methotrexate, cyclosporine, and pegylated interferon alpha has been reported in the successful treatment of idiopathic hypereosinophilic syndrome, even though the clinical evidence is scarce (10).

CONCLUSION

Persistent hypereosinophilia in the absence of any secondary causes should prompt the suspicion of idiopathic hypereosinophilic syndrome. Hypereosinophilic syndrome in ESRD patients, with regular haemodialysis, may complicate diagnosis by mimicking haemodialysis intolerance due to dialyzer reactions.

ETHICS

Informed Consent: She has given consent for this case to be reported and published.

FOOTNOTES

Authorship Contributions

Surgical and Medical Practices: Y.H., S.N.H.H., K.Y.L., N.E.M.K., M.Y.A.S., L.K., R.M., Consept: Y.H., S.N.H.H., K.Y.L., M.Y.A.S., Design: Y.H., S.N.H.H., K.Y.L., N.E.M.K., L.K., R.M., Data Collection or Processing: Y.H., K.Y.L., N.E.M.K., Analysis or Interpretation: Y.H., K.Y.L., N.E.M.K., M.Y.A.S., L.K., R.M., Literature Search: Y.H., S.N.H.H., K.Y.L., Writing: Y.H., S.N.H.H., K.Y.L.

Conflict of Interest: No conflict of interest was declared by the authors.

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

REFERENCES

1. Valent P, Klion AD, Horny HP, Roufosse F, Gotlib J, Weller PF, et al. Contemporary consensus proposal on criteria and classification of eosinophilic disorders and related syndromes. J Allergy Clin Immunol. 2012;130:607-12.
2. Ogbogu PU, Bochner BS, Butterfield JH, Gleich GJ, Huss-Marp J, Kahn JE, et al. Hypereosinophilic syndrome: a multicenter, retrospective analysis of clinical characteristics and response to therapy. J Allergy Clin Immunol. 2009;124:1319-25.
3. Mutsuyoshi Y, Hirai K, Morino J, Kaneko S, Minato S, Yanai K, et al. Idiopathic hypereosinophilic syndrome in hemodialysis patients. Medicine. 2021;100:e25164.

4. Gauckler P, Shin JI, Mayer G, Kronbichler A. Eosinophilia and kidney disease: more than just an incidental finding? J Clin Med. 2018;7:529.
5. Daugirdas JT, Ing TS. First-use reactions during hemodialysis: a definition of subtypes. Kidney Int Suppl. 1988;24:S37-43.
6. Chen DP, Flythe JE. Dialysis-associated allergic reactions during continuous renal replacement therapy and hemodialysis: a case report. Hemodial Int. 2020;24:E5-9.
7. Mukaya JE, Jacobson MS, Esprit D, Ajayi T. Allergic reaction to polysulphone membrane dialyser masquerading as infection. BMJ Case Rep. 2015;2015:bcr2014208591.
8. Simon P, Potier J, Thebaud HE. Facteurs de risque des réactions aiguës d'hypersensibilité en hémodialyse: enquête prospective multicentrique sur six mois dans l'Ouest de la France [Risk factors for acute hypersensitivity reactions in hemodialysis]. Nephrologie. 1996;17:163-70.
9. Gotlib J. Eosinophilic myeloproliferative disorders. In: Abutalib SA, ed. Cancer consult: expertise for clinical practice. 1st ed. New York: McGraw-Hill Education; 2014:167-75.
10. Caminati M, Carpagnano LF, Alberti C, Amaddeo F, Bixio R, Caldart F, et al. Idiopathic hypereosinophilic syndromes and rare dysimmune conditions associated with hyper-eosinophilia in practice: an innovative multidisciplinary approach. World Allergy Organ J. 2024;17:100928.