The Medical Bulletin of Haseki

E-ISSN: 2147-2688

2024
Volume 62
Issue
September 4

www.hasekidergisi.com

The Medical Bulletin of Haseki

Editorial Board

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

Online Publishing Date: September 2024
E-ISSN: 2147-2688
International scientific journal published quarterly.

The Medical Bulletin of Haseki

Scientific Advisory Board

The Medical Bulletin of Haseki

Scientific Advisory Board

The Medical Bulletin of Haseki

Scientific Advisory Board

The Medical Bulletin of Haseki

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

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

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

Owner: Mine GÜRSAÇ ÇELİK on Behalf of Haseki Training and Research Hospital

Responsible Manager: Akif ERBİN

The Medical Bulletin of Haseki

Contents

Original Articles

Case Report

Original Article

DOI: 10.4274/haseki.galenos.2024.9945
Med Bull Haseki 2024;62:184-190

Predictive Value of PSA Density in Pathological Discordance Terms in Patients who Undergo Robotic Surgery for Low-risk Prostate Cancer: An Analytic Cross-sectional Study of a Tertiary Reference Center

Emin Taha Keskin, Osman Can, Yigit Can Filtekin, Harun Ozdemir, Metin Savun, Abdulmuttalip Simsek, Halil Lutfi Canat

University of Health Sciences Turkey, Basaksehir Cam and Sakura City Hospital, Clinic of Urology, Istanbul, Turkey

Abstract

Aim: Pathological discordance between biopsy and radical prostatectomy (RP) remains a critical issue in determining appropriate treatment for prostate cancer (PCa). With this study, we aimed to evaluate the role of prostate specific antigen (PSA) density in predicting pathological discordance in low-risk PCa.

Methods: Data from 95 patients who underwent RP for low-risk PCa with Prostate Imaging Reporting and Data System 1-2 on multiparametric magnetic resonance imaging (MRI) were retrospectively analyzed in this cross-sectional study conducted between January and December 2023. The patients were divided into two groups according on biopsy and robotic-assisted laparoscopic prostatectomy pathology. The "compatible group" was defined as patients with no difference in International Society of Urological Pathology grade and tumor stage, other patients were defined as "incompatible group". The cut-off value for PSA density to predict the presence of pathological discordance was calculated by receiver operating-characteristic curve.

Results: Thirty eight (40%) patients were in the compatible group. No difference was found in serum PSA value between the groups (p=0.440), and a significant difference was found in prostate volume and PSA density (p=0.04 and p=0.001, respectively). The predictive cut-off value of PSA density was calculated as 0.088 ng/mL/cc (area under the curve: 0.729) (p<0.001). The sensitivity, specificity, positive and negative predictive values for this 0.088 ng/mL/cc value were 75.4%, 63.2%, 67.2% and 70.3%, respectively.

Conclusion: Prostate specific antigen density was found to have good performance in predicting pathological discordance in low-risk PCa patients with no pathological lesions detected by multiparametric MRI.

Keywords: Biopsy, prostatic neoplasms, prostatectomy, prostate-specific antigen, PSA density, robotics

Introduction

Early diagnosis of prostate cancer (PCa), the most common type of solid organ cancer in men, is important to ensure high treatment rates and local disease control. With the application of prostate specific antigen (PSA) screening for early diagnosis, there has been a dramatic increase in the number of biopsies and the rate of PCa (1,2). This rise also results in detection of tumors that might have no clinical significance and in unnecessary early diagnosis. In the modern era, in addition to imaging techniques, a number of factors are used in order to prevent unnecessary biopsies and unnecessary early diagnosis. Prostate specific antigen density (PSAD), a parameter identified by Benson et al. (3) in the early 1990s, is defined as the ratio of the serum PSA value to the volume of the prostate. This defined PSAD value has become known for its potential applications in the detection of clinically significant PCa and the prediction of high-risk disease, in addition to helping biopsy decisions (4,5).

Address for Correspondence: Emin Taha Keskin, University of Health Sciences Turkey, Basaksehir Cam and Sakura City Hospital, Clinic of Urology, Istanbul, Turkey
Phone: +90 212 909 60 00 **E-mail:** emintaha90@hotmail.com **ORCID:** orcid.org/0000-0002-1222-6424
Received: 09.06.2024 **Accepted:** 02.09.2024

Numerous studies have shown the possibility of pathological incompatibility between radical prostatectomy (RP) pathology and the pathology from prostate biopsies (6). Keskin et al. (7) found that compliance was only detected in 56% of the patients, and nearly half of the patients showed pathological non-compliance. The high rate of pathological non-compliance results in half of the patients receiving either inadequate or excessive treatment. In order to predict pathological incompatibility and prevent unnecessary biopsies, current guidelines recommend multiparametric magnetic resonance imaging (mp-MRI) for nearly all patients prior to biopsy (8). Multiparametric magnetic resonance imaging has a good sensitivity in detecting lesions classified as International Society of Urological Pathology (ISUP) grade group 1<. However, the efficacy of mp-MRI in detecting ISUP grade group 1 lesions remains less than 30% (9). Thus, despite the fact that we are in the era of MRI, the significance of other parameters, such as PSAD, remains due to the low success rate of MRI in ISUP grade group 1 lesions.

With this study, we aimed to investigate the role of PSAD value on the prediction of pathological discordance in patients diagnosed with low-risk ISUP grade group 1 PCa who did not have any lesions detected on mp-MRI.

Methods

After receiving ethics committee approval from the University of Health Sciences Turkey, Basaksehir Cam and Sakura City Hospital Ethics Committee (approval no.: KAEK/12.06.2024.08, date: June 28, 2024), the data of patients who were diagnosed with PCa by biopsy and underwent robot-assisted laparoscopic prostatectomy pathology (RALRP) was retrospectively analyzed with this cross-sectional study between January 2023 and December 2023 in our tertiary reference center.

Patients who were diagnosed with low-risk PCa according to the D'amico classification using serum PSA value, biopsy pathology [transrectal ultrasound-guided systematic prostate biopsy (12-core)] and digital rectal examinations, and no pathological findings were detected in mp-MRI evaluation [Prostate Imaging Reporting and Data System (PI-RADS) 1 and 2] were included in our study. Patients with intermediate and high risk PCa diagnosis, PI-RADS 2< lesion on mp-MRI, longer than 6 months between the biopsy date and the operation date, parameters that may affect tumor aggressiveness such as the presence of lymphovascular invasion, presence of variant pathology, and cribriform pattern in the biopsy pathology were excluded (Figure 1).

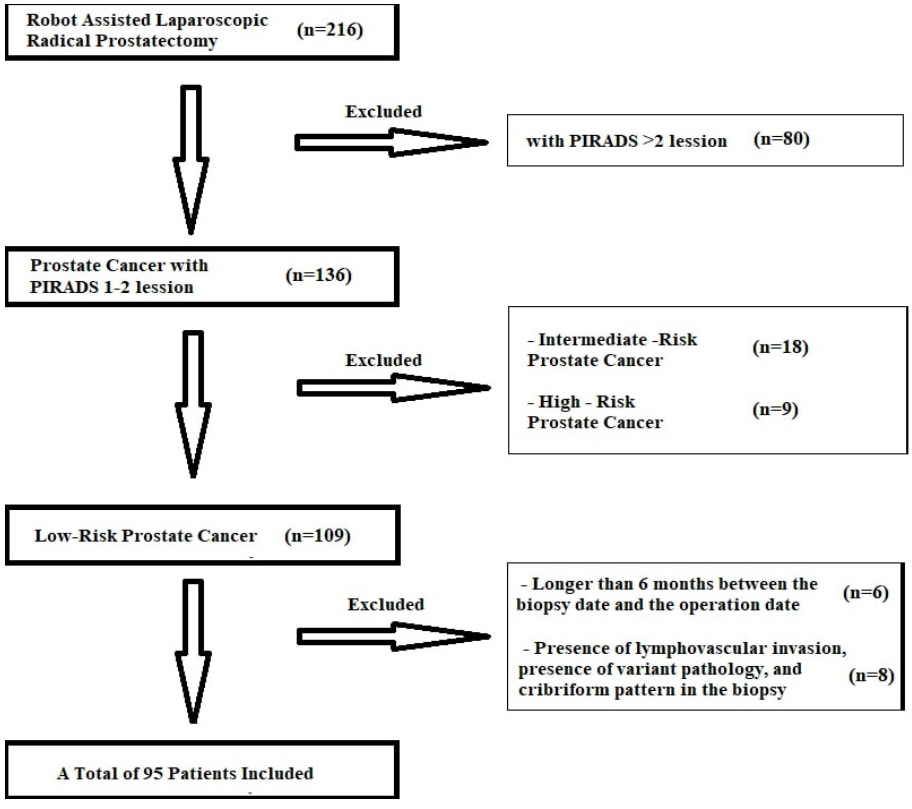


Figure 1. Flowchart of the study
PI-RADS: Prostate Imaging Reporting and Data System

Prostate volume was calculated in cc units by uroradiologists with 5 years of experience using mp-MRI images of the patients. Pathological evaluations of both biopsy and RALRP specimens were performed by uropathologists with at least 5 years of experience. Prostate specific antigen density was calculated in ng/mL/cc unit by dividing the serum PSA value (ng/mL) by the prostate volume (cc).

Gleason scores (GS) in the biopsy and RALRP specimen pathologies of the patients were expressed as ISUP grade scores. An increase in the ISUP grade score in the robot-assisted laparoscopic radical prostatectomy (RYLRP) specimen compared to the biopsy was considered an upgrade. In the pathological examination of the RYLRP specimen, the presence of extracapsular tumor extension, seminal vesicle invasion, and surrounding organ invasion was noted; if at least one of these was present, it was considered a pathological tumor stage increase (upstage).

When compared with the biopsy result, patients in the ISUP grade group and without pathological tumor stage change in the pathology of the RALRP specimen were recorded as the "compatible group", and the other patients were recorded as the "non-compatible group". Within the non-compatible group, patients who have only an elevation in ISUP [upgrade, upgrading patients (UG)] are defined to as "UG only". Similarly, those who alone exhibit an increase in pathological stage [upstage, upstaging patients (US)] are defined as "US only". Lastly, individuals who present both an increase in ISUP and stage are defined as "UG + US".

This study was designed as a retrospective study using only hospital records; informed consent is not obtained.

Statistical Analysis

The data obtained from the patients were analyzed via Statistical Package Program for Social Sciences 22.0. Numbers and percentages were used for descriptive statistics for categorical variables. The mean, minimum, and maximum values were used to describe numerical variables. The distribution of the data was tested using the Kolmogorov-Smirnov test for normality. T-test was used for parameters to compare groups with normal distribution. The cut-off value for PSA density indicating pathological discordance was calculated with the receiver operating-characteristic curve. P-value was considered <0.05 within the 95% confidence interval.

Results

Data from 95 patients were used in our study. The mean PSA value, PSA density and prostate volume (cc) of all patients were calculated as 5.7±1.6 ng/dL, 0.09±0.03 ng/mL/cc and 66.1±34 cc, respectively. All patients (100%, n=95) were found to have PI-RADS<3 lesions on mp-MRI and biopsy pathology was ISUP 1 (GS 3+3=6). The clinical and pathological stages and RALRP pathological datas of the patients were shown in Table 1. Only 40% (n=38) of the patients were found to be in the compatible group. Analysis of the data from patients in the non-compatible group compared to all patients showed that 37.9% (n=36) had only UG, 1.1% (n=1) had US, and 21% (n=20) had both UG and US.

Table 1. Demographic and pathological data

	All patients (n=95)
	Mean±SD
Age	62.5±5.5
BMI (kg/m²)	27.3±2.8
PSA value (ng/dL)	5.7±1.6
PSA density (ng/mL/cc)	0.09±0.03
Prostate volume (cc)	66.1±34
	n (%)
Clinical T-stage	
T1c	73 (76.8%)
T2a	22 (23.2%)
Pathological results of RALRP	
Gleason score	
3+3	40 (42.1%)
3+4	39 (41.1%)
4+3	9 (9.5%)
4+4	1 (1.1%)
4+5	5 (5.3%)
5+3	1 (1.1%)
ISUP grade	
1	40 (42.1%)
2	39 (41.1%)
3	9 (9.5%)
4	2 (2.1%)
5	5 (5.3%)
Pathological T-stage	
T2	74 (77.9%)
T3a	17 (17.9%)
T3b	4 (4.2%)
Compatible group	38 (40%)
Non-compatible group	57 (60%)
UG only	36 (37.9%)
US only	1 (1.1%)
UG + US	20 (21%)

BMI: Body mass index, ISUP: International Society of Urological Pathology, PI-RADS: Prostate Imaging Reporting and Data System, PSA: Prostate specific antigen, PALRP: Robot assisted laparoscopic radical prostatectomy, SD: Standard deviation, UG: Upgrading patients, US: Upstaging patients

There was no difference in PSA values (p-value=0.440) between the patients in the non-compatible group (UG group, US group, and US + UG group) and the patients in the compatible group, but there was a statistically significant difference in prostate volumes and PSA densities (p-value=0.04 and 0.001, respectively) (Table 2).

The cut-off value for PSA density in predicting pathological discordance was calculated as 0.088 ng/mL/cc (area under the curve: 0.729) (p<0.001). For the value of 0.088 ng/mL/cc, sensitivity, specificity, and positive and negative predictive values were calculated as 75.4%, 63.2%, 67.2%, and 70.3%, respectively (Table 3, Figure 2).

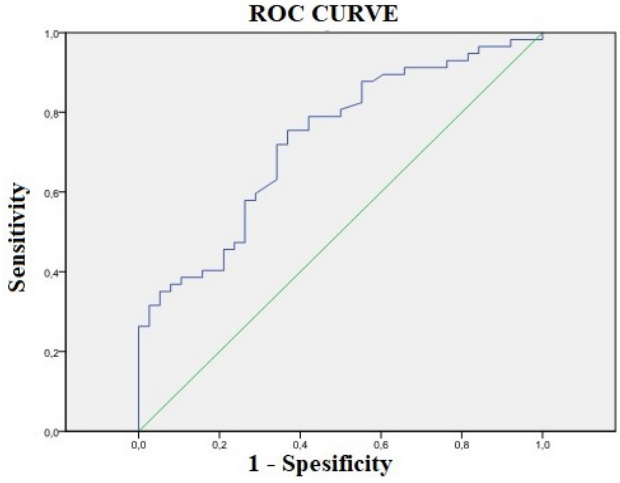


Figure 2. Receiver operating characteristic curve of PSA density predicting pathological progression (area under the curve: 0.729)
PSA: Prostate specific antigen, ROC: Receiver operating characteristic curve

Discussion

Tumor stage (as defined by the tumor-node-metastasis classification) and grade, GS, and ISUP grade classification have been accepted to be useful and reliable factors in predicting the prognosis of PCa. In PCa, tumor stage is determined by digital rectal examination and radiological imaging, and tumor grade is also determined by pathological examination of biopsy cores. Using these parameters, which are indicators of tumor aggressiveness, in addition to the PSA value, patients are classified as low, intermediate, or high risk, and treatment is planned according to this risk classification. Pathological results from biopsies are used to classify patients' risks and plan their treatment, but there can be pathological discordance between the specimen from a RP and the pathology of biopsies from patients who had a RP, depending on the risk class chosen based on the biopsy result. Upgrading is detected at a rate of approximately 30% in RP pathology, depending on the biopsy grade (10). In an another study, the upgrading and upstaging rates were found 42% and 24%, respectively (11).

Considering that more conservative treatment protocols, such as active surveillance, are predominantly applied to low-risk patients, this high pathological discordance rate becomes more clinically important. Failure to determine the correct PCa aggressiveness may lead to inadequate treatment and inappropriate follow-up of aggressive tumors. Since, it appears that some of the low-risk patients who are considered to be in the clinically localized disease group actually have a more aggressive malignancy. Various parameters have been investigated to predict this discordance, and a recent meta-analysis identified age, prostate volume, PSA value, PSAD, number and percentage of positive cores, PI-RADS score, clinical

Table 2. Comparison of pathologically compatible and non-compatible groups

	Compatible group (n=38)	Non-compatible group (n=57)					
		UG only (n=36)	p-value	US only (n=1)	p-value	US + UG (n=20)	p-value
PSA value (ng/dL)	5.65±1.5	5.71±1.9	0.880*	5.6	n/a	5.97±1.2	0.440*
Prostate volume (cc)	78.8±39.0	61.6±31.5	**0.041***	35	n/a	51.7±16.8	**0.004***
PSA density (ng/mL/cc)	0.081±0.028	0.103±0.034	**0.004***	0.16	n/a	0.124±0.044	**0.001***

Bold values refer to statistical significance. *t-test
PSA: Prostate specific antigen, UG: Upgrading patients, US: Upstaging patients

Table 3. Cut-off value of PSA density in predicting pathological discordance

	Cut-off value	Sensitivity (%)	Specificity (%)	PPV	NPV	AUC (95% CI)	p-value
PSA density (ng/mL/cc)	0.088	75.4%	63.2%	67.2%	70.3%	0.729	<0.001

AUC: Area under the curve, NPV: Negative predictive value, PPV: Positive predictive value, PSA: Prostate specific antigen, CI: confidence interval

T-stage, surgical margin status, and pathological T-stage as independent factors predicting upgrading following RP (6).

Several research have been published in the literature to reduce high rates of upgrading. It has been shown in the literature that the upgrading rate decreases as the number of sample cores taken in the biopsy increases. King et al. (12) showed that extended biopsy reduced GS upgrading from 66.7% to 36.8% in patients with a biopsy GS of 6. Capitanio et al. (13) study also showed that taking more than 18 core biopsy samples in low-risk PCa patients reduced the GS upgrading rate from 47.9% to 23.5%. Since the question of what the optimal number of biopsy cores should be for any biopsy technique is still unclear, Chambó et al. (14) recommend that at least 10 biopsy samples be obtained in patients with low-risk PCa. The European Urology Association guideline recommends that taking samples from more than 12 cores does not have an additional contribution to diagnosis and suggests a systematic 12-core biopsy for patients with suspicions of PCa who don't have any lesions on MRI (PI-RADS 1-2) (15). For this reason, it is even more important to predict pathological discordance in patients who have no lesion detected on MRI (PI-RADS 1-2) and who undergo systematic 12-core biopsy with suspicion of PCa. It seems that several parameters are needed other than increasing the number of cores to predict pathological compatibility and reduce possible non-compatible, especially in this patient group. We designed our analysis to include only patients who had a systematic 12-core biopsy in order to analyze the importance of the parameters in this group. Some studies in the current literature have shown that the percentage of positive cores among the all the sampled cores may affect the pathological concordance between biopsy and RP. A few studies have shown that an increase in the number of positive cores is associated with an increase in the GS upgrading rate, but there are also studies in the literature showing that there is no relationship between them (11,16). Since there is still no clear information on this issue, we did not evaluate the effect of positive core rate on compliance in this study.

The relationship between preoperative serum PSA value and GS upgrading also varies in the literature, similar to other parameters. Moussa et al. (18) showed that PSA level was a statistically significant determinant of GS upgrading, since the studies of Mian et al. (17), Jin et al. (11), and King et al. (12) did not show any relationship. In our study, no significant relationship was found between PSA level and GS upgrading.

Increased prostatic volume has been shown to lower the risk of GS upgrading (19). The relationship between prostatic volume and upgrading has been tried to be explained by the fact that the presence of a small prostate volume is an indicator of low *in vivo* androgenicity and that PCa, an androgen-dependent cancer, can develop despite this low *in vivo* androgenicity. This may indicate that cancer cells developing in the small-volume prostate may be a more aggressive tumor (20). Jin et al. (11) showed that patients with upgrading had significantly lower prostate volume, but in the same study, this low volume was not shown to be a predictor for upgrading in regression analysis. In this study, we found that patients with upgrading had lower prostate volume, similar to the literature. The choice of tool for calculating prostate volume affects the reliability of PSAD. Ultrasonography (USG) and MRI are the most commonly used methods. Transabdominal measurements generally yield higher prostate volumes than transrectal USG, which tends to underestimate prostate volume compared to mp-MRI. Additionally, mp-MRI-based PSAD calculations have shown a higher detection rate for PCa than transrectal USG (21,22). In our study, to minimize this bias, we used prostate volume calculated from mp-MRI for all measurements.

The PSAD was initially introduced as a more accurate predictor of PCa than PSA, but its use has been inconsistent in daily practice over the years. However, it was found to be associated not only with cancer detection but also with cancer aggressiveness (23). Several studies on PSAD-based upgrade prediction have been published in the literature. Corcoran et al. (24) showed that 58.3% of patients diagnosed with low-risk PCa increased to higher GS in RP pathology and that PSA density was a significant predictor of upgrading in ISUP group 1 patients. Similar results were shown by Kojima et al. (25) and Magheli et al. (26). Jin et al. (11) found 0.13 ng/mL as a significant predictive value for PSAD, with a sensitivity and specificity rate of 40% and 92%, respectively, to predict upgrading, and in the same study, they recommend using PSAD together with other predictive factors, considering the complexity of PCa. Sfoungaristos et al. (27) determined this value as 0.15 ng/mL. However, in the current study by Ozkaya et al. (28), the PSAD value was 0.18 ng/mL in the upgraded group and 0.16 ng/mL in the non-upgraded group, with no significant difference observed between the two, contrary to these previous studies. Our study showed a significant predictive value in predicting pathological discordance for 0.088 ng/mL, with a higher sensitivity of 75.4% and a lower specificity of 63.2%, contrary to the literature.

Study Limitations

There are some limitations in our study. First of all, the retrospective design of our study and the low number of patients are the main limitations of our study. In addition, the presence of more than one uroradiologist and uropathologist in both the evaluation of MRI images and pathological examination may lead to interobserver

differences, and this is considered another important limitation of our study. There is no clear period in the literature to prevent changes in tumor stage and grade between biopsy and RP operation in PCa patients. Therefore, in our study, we limited this period to 6 months, depending on our high clinical patient load. Yet, it is predictable that different outcomes may arise within shorter or longer periods. Hence, this duration interval between two processes acts as a further limitation in our study. Despite these limitations, using only MRI for prostate volume calculation is a strength of our study.

Conclusion

Our study highlights the ongoing significance of PSAD in predicting the consistency between biopsy and RP material in patients diagnosed with low-risk ISUP grade 1 PCa, particularly when mp-MRI detects no pathological lesion. Prostate specific antigen density is easy to use and calculable; despite recent technological developments, its significance remains valuable and it is a good predictor for upgrading and upstaging in PCa.

Footnote

Ethics Committee Approval: The study was initiated after receiving approval from the University of Health Sciences Turkey, Basaksehir Cam and Sakura City Hospital Ethics Committee (approval no.: KAEK/12.06.2024.08, date: June 28, 2024).

Informed Consent: This study was designed as a retrospective study using only hospital records; informed consent is not obtained.

Authorship Contributions

Surgical and Medical Practices: E.T.K., O.C., Y.C.F., H.O., M.S., A.S., H.L.C., Concept: E.T.K., O.C., A.S., H.L.C., Design: E.T.K., A.S., H.L.C., Data Collection or Processing: E.T.K., Analysis or Interpretation: E.T.K., O.C., Y.C.F., H.O., M.S., Literature Search: E.T.K., Writing: E.T.K., O.C., Y.C.F., H.O., M.S.

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

Financial Disclosure: This study received no financial support.

References

1. Boyle P, Ferlay J. Cancer incidence and mortality in Europe, 2004. Ann Oncol. 2005;16:481-88.
2. Siegel RL, Miller KD, Fuchs HE, et al. Cancer statistics, 2022. CA Cancer J Clin. 2022;72:7-33.
3. Benson MC, Whang IS, Pantuck A, et al. Prostate specific antigen density: a means of distinguishing benign prostatic hypertrophy and prostate cancer. J Urol. 1992;147:815-6.
4. Daryanto B, Trianto R, Seputra KP, et al. Evaluation of cutoff point prostate specific antigen (PSA) and prostate specific antigen density (PSAD) in patients with suspected prostate cancer. Med Arch. 2024;78:12-5.
5. Zou BZ, Wen H, Luo HJ, et al. Value of serum free prostate-specific antigen density in the diagnosis of prostate cancer. Ir J Med Sci. 2023;192:2681-7.
6. Wang Y, Chen X, Liu K, et al. Predictive factors for gleason score upgrading in patients with prostate cancer after radical prostatectomy: a systematic review and meta-analysis. Urol Int. 2023;107:460-79.
7. Keskin ET, Yikilmaz TN, Öztürk E, et al. Effect of prostate biopsies core number on correlation of needle biopsies and radical prostatectomy's gleason score. The New Journal of Urology. 2017;12:6-10.
8. Klotz L, Chin J, Black PC, et al. Comparison of multiparametric magnetic resonance imaging–targeted biopsy with systematic transrectal ultrasonography biopsy for biopsy-naive men at risk for prostate cancer: a phase 3 randomized clinical trial. JAMA Oncol. 2021;7:534-42.
9. Drost FJH, Osses DF, Nieboer D, et al. Prostate MRI, with or without MRI-targeted biopsy, and systematic biopsy for detecting prostate cancer. Cochrane Database Syst Rev. 2019;4:CD012663.
10. Cohen MS, Hanley RS, Kurteva T, et al. Comparing the Gleason prostate biopsy and Gleason prostatectomy grading system: the Lahey Clinic Medical Center experience and an international meta-analysis. Eur Urol. 2008;54:371-81.
11. Jin BS, Kang SH, Kim DY, et al. Pathological upgrading in prostate cancer patients eligible for active surveillance: Does prostate-specific antigen density matter? Korean J Urol. 2015;56:624-9.
12. King CR, McNeal JE, Gill H, et al. Extended prostate biopsy scheme improves reliability of Gleason grading: implications for radiotherapy patients. Int J Radiat Oncol Biol Phys. 2004;59:386-91.
13. Capitanio U, Karakiewicz PI, Valiquette L, et al. Biopsy core number represents one of foremost predictors of clinically significant gleason sum upgrading in patients with low-risk prostate cancer. Urology. 2009;73:1087-91.
14. Chambó RC, Tsuji FH, de Oliveira Lima F, et al. What is the ideal core number for ultrasound-guided prostate biopsy? Korean J Urol. 2014;55:725-31.
15. Eichler K, Hempel S, Wilby J, et al. Diagnostic value of systematic biopsy methods in the investigation of prostate cancer: a systematic review. J Urol. 2006;175:1605-12.
16. Hong SK, Han BK, Lee ST, et al. Prediction of Gleason score upgrading in low-risk prostate cancers diagnosed via multi (≥ 12)-core prostate biopsy. World J Urol. 2009;27:271-76.
17. Mian BM, Lehr DJ, Moore CK, et al. Role of prostate biopsy schemes in accurate prediction of Gleason scores. Urology. 2006;67:379-83.
18. Moussa AS, Li J, Soriano M, et al. Prostate biopsy clinical and pathological variables that predict significant grading changes

in patients with intermediate and high grade prostate cancer. BJU Int. 2009;103:43-8.

19. Kassouf W, Nakanishi H, Ochiai A, et al. Effect of prostate volume on tumor grade in patients undergoing radical prostatectomy in the era of extended prostatic biopsies. J Urol. 2007;178:111-4.
20. Freedland SJ, Isaacs WB, Platz EA, et al. Prostate size and risk of high-grade, advanced prostate cancer and biochemical progression after radical prostatectomy: a search database study. J Clin Oncol. 2005;23:7546-54.
21. Choe S, Patel HD, Lanzotti N, et al. MRI vs transrectal ultrasound to estimate prostate volume and PSAD: impact on prostate cancer detection. Urology. 2023;171:172-8.
22. de Vos II, Drost FH, Bokhorst LP, et al. Prostate cancer risk assessment by the primary care physician and urologist: transabdominal- versus transrectal ultrasound prostate volume-based use of the Rotterdam Prostate Cancer Risk Calculator. Transl Androl Urol. 2023;12:241-8.
23. Yusim I, Krenawi M, Mazor E, et al. The use of prostate specific antigen density to predict clinically significant prostate cancer. Sci Rep. 2020;10:20015.
24. Corcoran NM, Casey RG, Hong MK, et al. The ability of prostate-specific antigen (PSA) density to predict an upgrade in Gleason score between initial prostate biopsy and prostatectomy diminishes with increasing tumour grade due to reduced PSA secretion per unit tumour volume. BJU Int. 2012;110:36-42.
25. Kojima M, Troncoso P, Babaian RJ. Use of prostate-specific antigen and tumor volume in predicting needle biopsy grading error. Urology. 1995;45:807-12.
26. Magheli A, Hinz S, Hege C, et al. Prostate specific antigen density to predict prostate cancer upgrading in a contemporary radical prostatectomy series: a single center experience. J Urol. 2010;183:126-32.
27. Sfoungaristos S, Katafigiotis I, Perimenis P. The role of PSA density to predict a pathological tumour upgrade between needle biopsy and radical prostatectomy for low risk clinical prostate cancer in the modified Gleason system era. Can Urol Assoc J. 2013;7:E722-7.
28. Ozkaya M, Simsekoglu MF, Kalender G, et al. Clinical and histopathological parameters in transrectal ultrasound-guided biopsies associated with tumor upgrading after radical prostatectomy: A comparative analysis of risk groups. Prostate. 2024;84:1146-56.

Original Article

DOI: 10.4274/haseki.galenos.2024.9939
Med Bull Haseki 2024;62:191-200

Evaluation of the Quality of YouTube Videos for Age-related Macular Degeneration in Turkey

Mine Ozturk*, Ibrahim Cagri Turker**, Gurcan Dogukan Arslan***

*University of Health Sciences Turkey, Istanbul Haseki Training and Research Hospital, Clinic of Ophthalmology, Istanbul, Turkey
**University of Health Sciences Turkey, Sisli Hamidiye Etfal Training and Research Hospital, Clinic of Ophthalmology, Istanbul, Turkey
***Atlas University Medicine Hospital, Clinic of Ophthalmology, Istanbul, Turkey

Abstract

Aim: Patients frequently use YouTube for information. This study aimed to evaluate the quality of Turkish YouTube videos on age-related macular degeneration (AMD).

Methods: Turkish translations of "AMD" and "yellow spot disease" were searched on August 16, 2023. The first 100 videos were grouped as useful, misleading, or irrelevant. Video properties and quality scores [Journal of American Medical Association (JAMA), Global Quality Score (GQS), DISCERN] were evaluated.

Results: We classified 74 videos as useful, 16 as misleading, and 10 as irrelevant. The most common uploaders were television shows and social media. The mean JAMA score was 1.94±1.09, the mean GQS score was 2.49±1.29, and the mean DISCERN score was 36.82±16.53. The mean JAMA, GQS, and DISCERN scores were higher in the useful group than in the misleading group (p=0.0001, p=0.001, p=0.0001). The number of likes and view ratio were higher in the misleading group than in the useful group (p=0.019, p=0.037). The main content of the misleading group was treatment, and the uploaders were related to industry and commercial interests.

Conclusion: The quality of the AMD videos was insufficient. Although misleading videos had lower quality, they were more popular. This may be due to patients' hope for new treatments, as misleading videos have discussed this topic.

Keywords: YouTube, age-related macular degeneration, DISCERN score, global quality scale score, Journal of the American Medical Association score, social media, Turkish

Introduction

Age-related macular degeneration, or AMD, was first described in 1985 as a disease in people over 50 that causes a loss of central vision and changes in the macula's color and structure (1). In developed countries, it is the most important cause of visual impairment in people aged >60 years, accounting for 8.7% of legal blindness cases worldwide. The prevalence of AMD is believed to reach approximately 288 million by 2040 (2). With the prolonged life expectancy, the incidence of AMD has also increased.

Visual disorders, which have a negative impact on quality of life due to their physical and psychological effects, can affect patients on a daily basis, ranging from social relations to work concentration. With emotions such as tension, fear, and anxiety, patients with vision loss turn to online platforms where they can obtain information and visit a physician.

Due to high rates of advanced literacy, researchers looked into health-related internet use habits and found a strong relationship between health literacy and internet use (3).

Social networking sites, such as Facebook, YouTube, and Twitter, have opened new avenues for disseminating health-related information. Studies have been conducted to evaluate the content of health-related videos on different social networking sites (4,5). YouTube is a free video-sharing platform with more than 1 billion users. Millions of users upload videos on YouTube, and uncontrolled sharing

Address for Correspondence: Mine Ozturk, University of Health Sciences Turkey, Istanbul Haseki Training and Research Hospital, Clinic of Ophthalmology, Istanbul, Turkey
Phone: +90 532 411 53 55 **E-mail:** dr_mine@yahoo.com **ORCID:** orcid.org/0000-0003-4544-1021
Received: 06.06.2024 **Accepted:** 02.09.2024

can cause information pollution that contradicts relevant standards, especially health-related standards. Studies evaluating the quality of health-related videos on YouTube have reported that nearly 16-30% present misleading or low-quality information (6,7).

Video popularity is calculated using the video power index and view rate. The Journal of American Medical Association (JAMA), DISCERN, and Global Quality Score (GQS) scoring systems are used to evaluate the educational quality of videos. There are concerns regarding the quality and reliability of online medical information because uncontrolled individual video uploads can result in inaccurate and misleading information. It has become possible to access high-quality educational videos through scoring systems such as JAMA, DISCERN, and GQS.

In ophthalmology, numerous studies have evaluated the information quality and reliability of YouTube videos on different topics. Kayabaşı et al. (8) analyzed YouTube videos about myopia and concluded that they were of weak to moderate quality. The study of Tanyıldız and Oklar (9) concluded that YouTube videos about uveitis were poor in reliability and quality and were not sufficiently educational for patients. Ozturkmen and Berhuni (10) evaluated YouTube videos about pterygium surgery and reported that they were of low quality and inadequate in informing patients. Kaptı and Erdem (11) reported the quality of YouTube videos about congenital nasolacrimal duct obstruction as "average". Unlike most other studies, İlhan et al. (12) examined YouTube videos on thyroid orbitopathy, and the quality was good in most videos.

This study aimed to evaluate the reliability, popularity, and quality of Turkish YouTube videos on AMD to examine the status of Turkey, compare low, medium, and high-quality videos in terms of general video characteristics, and determine the relationship between general video characteristics and video reliability and quality evaluation.

Materials and Methods

Compliance with Ethical Standards

The data collected for this study were acquired from publicly accessible YouTube videos. This study was not required to obtain Institutional Review Board approval or ethical approval as it involved only public access data.

Search Strategy and Data Collection

In this cross-sectional, record-based study, YouTube (www.youtube.com) was searched using the Turkish keywords "yaşa bağlı makula dejenerasyonu and sarı nokta hastalığı" on August 16, 2023. "Age related macular degeneration" and "Yellow spot disease" are translations for "Yaşa bağlı makula dejenerasyonu" and "sarı nokta hastalığı" in the Turkish language. Video search was performed after clearing the browser's entire search history without the user logging in to prevent misdirection. The top 100 videos were selected for evaluation based on their "relevance", determined by YouTube's algorithm.

Video Categorization and Characteristics

Two ophthalmologist examiners (GDA, MO) watched and analyzed the videos.

All videos were grouped based on their content as useful, misleading, or irrelevant.

1. Useful: Video showing scientifically accurate and correct information regarding any aspect of the disease.

2. Misleading: The video presented inaccurate or unproven information based on available scientific evidence.

3. Irrelevant: This condition presents information that is not relevant or related to AMD.

Irrelevant videos were excluded from the study, and useful and misleading groups were included in the statistical analysis (Figure 1).

For each video, video metrics, including video length, time since upload, and number of likes, were recorded, and the view ratio was evaluated to assess popularity. The view ratio was determined by dividing the number of views by the time since upload.

The video and audio qualities of the videos were evaluated according to criteria previously described by Young et al. (13). Video quality was defined as good if it was professionally produced with excellent quality and effects, moderate if it was a home video, and poor if it was grainy, affecting the ability to see presentation details. Similarly, the audio quality was considered good if all words could be clearly heard without significant background noise or distracting audio effects, moderate if most words were understandable, with minimal background noise, and poor if it limited the understanding of the material.

Videos were also categorized into four groups based on their upload source: (1) ophthalmologist, (2) TV show/social media, (3) industry/commercial interest, and (4) medical school/academic center. The main contents discussed in the video were noted, and groups were classified according to the main content under the following headings: (1) Only treatment, (2) General information about disease + diagnosis, and (3) General information about disease + diagnosis + treatment (all aspects).

In addition, we recorded whether the video included real procedures or animation. Video participants were also recorded.

Video Quality Scoring Systems

The information quality of each video was evaluated using the DISCERN, JAMA, and GQS scores.

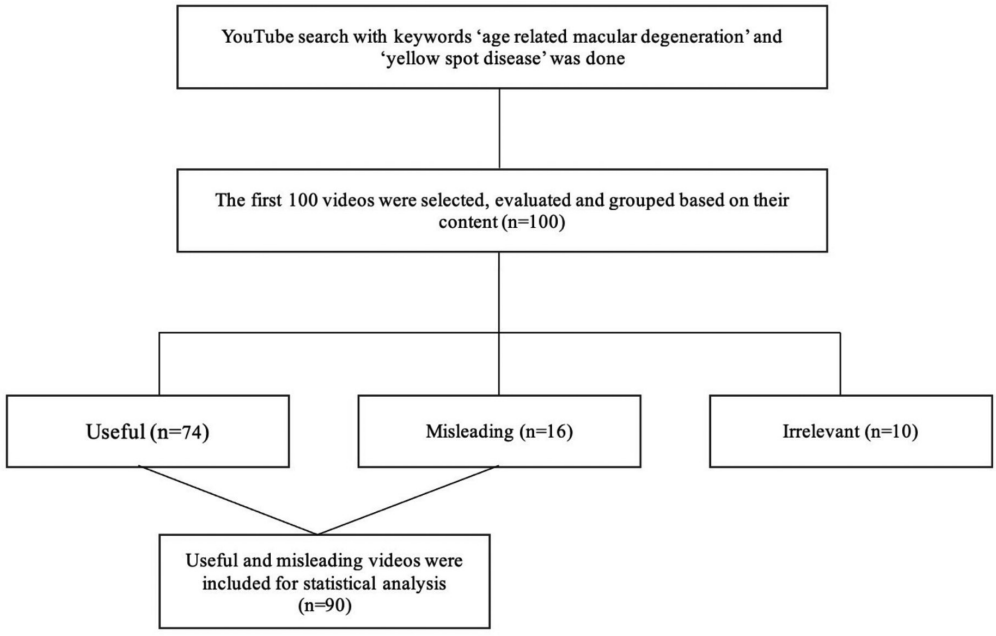


Figure 1. Flowchart of the study

The DISCERN score was developed to evaluate the educational quality and reliability of medical information, particularly treatment options available to the patient (14). The questionnaire contains three sections with 16 questions, and a higher score indicates superior quality. The initial eight questions related to reliability, whereas the latter seven assessed specific details of the treatments received. The final question relates to the overall quality of a publication (Appendix A). According to the DISCERN scoring system, the videos were grouped into excellent quality (63-80 points), good quality (51-62 points), fair quality (39-50 points), poor quality (27-38 points), and very poor quality (16-26 points).

The JAMA scoring system evaluates the quality of online health-related information according to four criteria: currency, authorship, disclosure, and attribution, each scored as 0 or 1 (Appendix B) (15). A high score on this scale indicates that the information was of good quality. A score of 4 indicates excellent quality, and a score of 0 indicates poor quality.

The GQS system is a 5-point scale used to evaluate the ease of use, overall flow, and accessibility of the information delivered in the video (Appendix C) (16). A score of 5 indicated excellent quality, and a score of 1 indicated poor quality.

Statistical Analysis

Statistical analyses were performed using the NCSS (Number Cruncher Statistical System) 2007 Statistical Software (Utah, USA). In evaluating the data, in addition to descriptive statistical methods (mean, standard deviation, median, interquartile range), the distribution of the variables was examined using the Shapiro-Wilk normality test. The Kruskal-Wallis test was used for intergroup comparisons of variables that did not show a normal distribution; Dunn's multiple comparison test was used for comparisons of subgroups; and Dunn's multiple comparison test was used for paired group comparisons. Mann-Whitney U and chi-square tests were used to compare qualitative data, and the Pearson correlation test was used to determine the relationships between variables. The results were evaluated at a significance level of $p<0.05$.

Results

Video Categorization and Characteristics

A total of 100 videos were initially included in the study. Of these, 74 (74%) were classified as useful, 16 (16%) as misleading, and 10 (10%) as irrelevant. Irrelevant videos were excluded, and 74 useful and 16 misleading videos related to AMD were included for further statistical analysis.

The video characteristics are summarized in Table 1. Both the audio and video qualities were generally good. Most videos (40%) were uploaded by television shows and social media platforms. More than half of the videos (57.8%) discussed AMD in terms of all aspects.

Table 2 summarizes the video metrics of the 90 videos analyzed. The mean JAMA score was 1.94±1.09 (moderate quality), the mean GQS score was 2.49±1.29 (moderate

Table 1. Video characteristics			
Video characteristics		**n**	**%**
Video quality	Good	72	80%
	Moderate	3	3.3%
	Poor	15	16.7%
Audio quality	Good	71	78.9%
	Moderate	10	11.1%
	Poor	9	10%
Uploader	Ophthalmologist	23	25.6%
	Industry/Commercial interest	29	32.2%
	Medical school/Academic center	2	2.2%
	TV show/Social media	36	40%
Video participants	Healthcare provider	88	97.8%
	Patients	2	2.2%
Video content	General information about disease + Diagnosis	17	18.9%
	General information about disease + Diagnosis + Treatment	52	57.8%
	Treatment	21	23.3%
Includes animation	No	68	75.6%
	Yes	22	24.4%
Includes real procedure	No	81	90%
	Yes	9	10%

Table 2. Video metrics		
	Mean±SD	**Range**
Time since upload (day)	1622.37±969.66	144-4077
Views (n)	4720.22±10842.85	36-80167
Likes (n)	29.21±62.92	0-476
View ratio	4.10±7.32	0.01-47.77
Length (min)	6.03±8.11	0.32-44.34
JAMA	1.94±1.09	0-4
GQS	2.49±1.29	1-5
DISCERN	36.82±16.53	16-68

JAMA: Journal of American Medical Association, GQS: Global Quality Score, SD: Standard deviation

quality), and the mean DISCERN score was 36.82±16.53 (poor quality). When the videos were evaluated according to the uploader, no significant differences were observed in JAMA, GQS, DISCERN scores, and view ratio (Table 3).

However, the JAMA and DISCERN scores differed significantly according to video content (p=0.045 and p=0.021, respectively) (Table 4). Overall, the highest video quality scores were obtained for videos that discussed the disease in all aspects (general information about the disease, diagnosis, and treatment group). When Dunn's multiple comparison test was performed, the JAMA averages of general information about the disease + diagnosis + treatment group were significantly higher than those of the treatment group (p=0.021). DISCERN averages of general information about the disease + diagnosis + treatment group were significantly higher than those of the treatment group (p=0.011). The view ratio was similar between the groups (p=0.061).

Correlation analysis showed a positive correlation between JAMA and GQS scores, between JAMA and DISCERN scores, and between GQS and DISCERN scores (r=0.862 p=0.0001; r=0.863 p=0.0001; r=0.874 p=0.0001), respectively. However, the scores were not correlated with the other parameters (Table 5).

After evaluating the data for all videos, the useful and misleading video groups were assessed separately and compared. Table 6 summarizes the video metrics of the two groups. Videos in the misleading group were newer than those in the useful group (p=0.006). The number of likes and view ratio were higher in the misleading group than in the useful group (p=0.019 and p=0.037, respectively). The JAMA, GQS, and DISCERN scores were statistically significantly higher in the useful group than in the misleading group (p=0.0001, p=0.001, and p=0.0001, respectively).

When the video and audio qualities and other characteristics of useful and misleading videos were compared, video content was the only significant difference between the two groups (p=0.0001). Videos giving all the details about the disease were most common in the useful group, whereas the most shared content in the misleading group was treatment (Table 7).

Table 3. Comparison of the JAMA, GQS and DISCERN scores and view ratio according to the uploader						
Uploader		**Ophthalmologist**	**Industry/Commercial interest**	**Medical school/Academic center**	**TV show/Social media**	**p‡**
JAMA	Median (IQR)	3 (1-3)	2 (1-2)	2 (1.5-1.52)	1.5 (1-3)	0.319
GQS	Median (IQR)	3 (2-4)	2 (1-3)	3.5 (2.25-3.02)	2 (1-3)	0.119
DISCERN	Median (IQR)	46 (20-54)	31 (18-47.5)	46 (33-36.02)	37 (20-50.75)	0.321
View ratio	Median (IQR)	0.53 (0.08-1.72)	1.13 (0.34-11.28)	6.2 (1.17-8.14)	1.17 (0.12-4.3)	0.203

‡Kruskal-Wallis test
JAMA: Journal of American Medical Association, GQS: Global Quality Score, IQR: Interquartile range

Table 4. Comparison of the JAMA, GQS and DISCERN scores and view ratio according to the video content					
Video content		**General information about disease + Diagnosis**	**General information about disease + Diagnosis + Treatment**	**Treatment**	**p†**
JAMA	Median (IQR)	2 (1-2.5)	2 (1-3)	1 (1-2)	0.045
GQS	Median (IQR)	2 (1-3)	3 (2-3.75)	2 (1-3)	0.086
DISCERN	Median (IQR)	32 (19.5-45)	44.5 (23.5-53.5)	20 (16-49)	0.021
View ratio	Median (IQR)	0.16 (0.07-1.36)	1.22 (0.24-5.89)	0.97 (0.33-6.59)	0.061
Dunn's multiple comparison test		**JAMA**	**DISCERN**		
General information about disease +diagnosis/General information about disease + diagnosis + treatment		0.156	0.114		
General information about disease + diagnosis/Treatment		0.432	0.222		
General information about disease + diagnosis + treatment/Treatment		0.021	0.011		

†Kruskal-Wallis test
JAMA: Journal of American Medical Association, GQS: Global Quality Score, IQR: Interquartile range

Table 5. Comparison of the JAMA, GQS and DISCERN scores, view ratio, and other video parameters									
		JAMA	**GQS**	**DISCERN**	**Time since upload (day)**	**Views (n)**	**Likes (n)**	**View ratio**	**Length (min)**
JAMA	r		0.862	0.863	0.105	-0.151	-0.167	-0.16	0.094
	p		0.0001	0.0001	0.325	0.156	0.116	0.132	0.377
GQS	r	0.862		0.874	0.064	-0.145	-0.16	-0.114	0.041
	p	0.0001		0.0001	0.548	0.174	0.133	0.286	0.698
DISCERN	r	0.863	0.874		0.14	-0.105	-0.121	-0.116	0.135
	p	0.0001	0.0001		0.187	0.325	0.256	0.278	0.206
Time since upload (day)	r	0.105	0.064	0.14		0.035	-0.095	-0.271	-0.035
	p	0.325	0.548	0.187		0.74	0.374	0.01	0.740
Views (n)	r	-0.151	-0.145	-0.105	0.035		0.947	0.844	0.039
	p	0.156	0.174	0.325	0.74		0.0001	0.0001	0.717
Likes (n)	r	-0.167	-0.16	-0.121	-0.095	0.947		0.897	0.122
	p	0,116	0.133	0.256	0.374	0.0001		0.0001	0.253
View ratio	r	-0.16	-0.114	-0.116	-0.271	0.844	0.897		0.108
	p	0.132	0.286	0.278	0.01	0.0001	0.0001		0.312
Length (min)	r	0.094	0.041	0.135	-0.035	0.039	0.122	0.108	
	p	0.377	0.698	0.206	0.74	0.717	0.253	0.312	

JAMA: Journal of American Medical Association, GQS: Global Quality Score

Table 6. Video metrics of useful and misleading video groups				
		Useful (n=74)	**Misleading (n=16)**	**p†**
Time since upload (day)	Median (IQR)	1965 (774.25-2351.75)	1127.5 (734.5-1253)	0.006
Views (n)	Median (IQR)	912.5 (259.25-4649)	2197.5 (701-6010)	0.126
Likes (n)	Median (IQR)	6 (1-26.25)	14 (9.5-54)	0.019
View ratio	Median (IQR)	0.99 (0.12-2.95)	3.24 (0.59-11.77)	0.037
Length (min)	Median (IQR)	2.56 (1.37-6.29)	3.87 (0.65-10.04)	0.768
JAMA	Median (IQR)	2 (1-3)	1 (0-2)	0.0001
GQS	Median (IQR)	3 (1.75-4)	1 (1-2)	0.001
DISCERN	Median (IQR)	44.5 (28-53.25)	16 (14.25-18.75)	0.0001

†Mann-Whitney U test
JAMA: Journal of American Medical Association, GQS: Global Quality Score, IQR: Interquartile range

Table 7. Characteristics of the useful and misleading videos						
		Useful (n=74)		**Misleading (n=16)**		**p+**
Audio quality	Good	59	79.73%	12	75.00%	0.379
	Moderate	9	12.16%	1	6.25%	
	Poor	6	8.11%	3	18.75%	
Video quality	Good	61	82.43%	11	68.75%	0.446
	Moderate	11	14.86%	4	25.00%	
	Home video	2	2.70%	1	6.25%	
Uploader	Ophthalmologist	19	25.68%	4	25.00%	0.657
	Industry/Commercial interest	22	29.73%	7	43.75%	
	Medical school/Academic center	2	2.70%	0	0.00%	
	TV show/Social media	31	41.89%	5	31.25%	
Includes animation	No	54	72.97%	14	87.50%	0.220
	Yes	20	27.03%	2	12.50%	
Includes real procedure	No	66	89.19%	15	93.75%	0.581
	Yes	8	10.81%	1	6.25%	
Video participants	Healthcare provider	73	98.65%	15	93.75%	0.228
	Specifically for patients	1	1.35%	1	6.25%	
Content	General information about disease + Diagnosis	17	22.97%	0	0.00%	0.0001
	General information about disease + Diagnosis + Treatment	48	64.86%	4	25.00%	
	Treatment	9	12.16%	12	75.00%	

+Chi-square test

Discussion

The primary finding of this study is the poor to moderate quality and reliability of Turkish YouTube videos on AMD. When the factors that may affect the quality scores were examined, content was considered a main factor. The group that mentioned only treatment had lower quality than the other groups. When the useful and misleading groups were compared in terms of video quality, all video quality scores of the misleading group were lower than those of the useful group. Contrary to lower quality scores, the like and view rates of the misleading group were higher than the useful group. Interestingly, the most popular videos did not necessarily have the highest quality and reliability.

YouTube is a video-sharing website where patients can gain information about their diagnoses and treatments recently through open-access health videos. Although the number of people seeking health-related information has increased in recent years because of the ease of access to such information, it has been revealed that approximately one-third of patients do not trust such content (17). In addition to patients, healthcare professionals often watch YouTube videos. In particular, surgical videos are viewed

by physicians on a learning curve or who want to discover and learn new techniques (18-20).

One of YouTube's key features is that anyone can upload videos, regardless of background, medical qualifications, professionalism, or purpose. Therefore, health-related information on YouTube can be quite wide-ranging, including inaccurate or highly qualified information. In this process, obtaining online information from the appropriate sources is crucial because it can even change patient compliance with treatment (15). In our study, we found that 10% of videos were irrelevant to the subject, and 18% of the related videos were misleading. This result is consistent with studies conducted in different countries and with different disease titles in the literature (6,7,21,22).

People often overlook the important factor of sound and image quality in YouTube videos. Good sound quality allows visitors to connect better with the video and engage more with the content. However, poor sound quality can reduce the video view rate even with the highest visual content. In our study, both the sound and video qualities were good. This may have occurred because most of the videos examined in our study were uploaded by TV shows and industries and, therefore, were prepared more professionally.

Video quality must be at the highest level, particularly for health-related videos. For this reason, several video quality questionnaires, such as HONcode, JAMA, DISCERN, and GQS, were used to evaluate the quality and reliability of the video content. In this study, the JAMA, DISCERN, and GQS scoring systems were used, and the results suggest that the YouTube videos on AMD are of "poor to moderate" quality. When the scoring and uploading individuals were compared, it was observed that the highest JAMA score was obtained by ophthalmologists, and the highest DISCERN and GQS scores were obtained by medical faculties and academic centers, indicating that videos uploaded by users outside the health sector are of lower quality than those added by health professionals. In our study, the low rate of ophthalmologists (25.6%) and medical school/academic center unloaders (2.2%) compared with the total might be one reason for the low quality scores. Our findings are compatible with the results of other studies in the literature (23,24). Non-profit organizations and academic-sourced videos are known to have the highest value for information. The high quality and reliability of healthcare professional videos can be attributed to several characteristics. First, healthcare experts base their recommendations on scientific evidence and clinical guidelines, which increases the accuracy and credibility of their information. Second, professional videos typically provide extensive coverage of the medical subject, encompassing various aspects, such as general information, diagnosis, and treatment alternatives, in a detailed manner. Third, accurate medical terminology enhances clarity, accuracy, and overall production quality. Hence, we advocate prioritizing videos produced by ophthalmologists or academic centers as the ideal educational resource for patients with AMD.

In this study, correlation analysis revealed a strong positive correlation between the DISCERN, JAMA, and GQS scores. This finding shows that the scales used in this study provide parallel results that reflect the reliability and educational quality of the videos.

The number of views indicates the popularity of videos, and the daily view rates reflect their currency (25). Our study found no correlation between JAMA, DISCERN, and GQS values and the number of views or view ratio; thus, high- and low-quality videos were watched at a similar rate. Our research shows that a video's popularity level does not necessarily correlate with its quality. This situation may be caused by two different reasons. First, patients may not be equipped to evaluate video content quality and accuracy and may watch videos without making a choice. Second, the number of views on YouTube was considered an important parameter indicating video popularity. It can be considered that videos with a high view rate among people have high reliability (26). Today, video discoverability and reliability can be increased artificially by purchasing views. This situation may cause patients to think that these videos are better. Because patients cannot be consciously selective within this pool of videos, ophthalmologists and academic centers need to increase the number of videos that contain accurate information.

We did not find a significant difference in the view ratio between the videos uploaded by academic and non-academic ones, indicating that patients do not prioritize videos uploaded by healthcare professionals when searching. This result may be related to several factors. First, commercially purchasing views and likes, which are mostly used by non-academic uploaders, can increase the discoverability of videos. Although this created the impression that the video was artificially better, the situation may be the opposite. Second, despite their medical knowledge, physicians may struggle to explain complex medical information in a simple manner that patients can easily understand. The medical language and complexity of videos uploaded primarily by physicians and healthcare institutions may make them less accessible to the general public.

When health-related videos on YouTube are evaluated, not only videos that provide general information about the disease but also videos about diagnosis and treatment can attract more attention (27). When useful and misleading videos were compared in our study, we found that misleading videos mainly discussed treatment. Misleading

videos might be more prominent than useful videos because they often share new, speculative, and promising content, whereas useful videos typically discuss proven and tried treatments. There is currently no treatment that can completely and permanently eliminate AMD, and there is no way to restore vision loss in advanced stages. Searching for new and alternative treatment methods may prompt patients to highlight new, misleading content. However, turning to unproven and unreliable treatment methods with the hope of a cure may harm patients both financially and healthfully.

Study Limitations

Our study has some limitations. Among this study's weaknesses is the subjective nature of the analysis, which was attempted to minimize by having two experienced observers rate the videos and confirm the inter-rater reliability. Another limitation is that although terms in both everyday spoken language and academic language (yellow spot disease and age-related macular degeneration) were used in the search section, the search words used by the authors may differ from the search words of the patients and the videos they encounter. In addition, it is known that the keywords typed into the search engine may vary according to the geographic region of the searcher and the search history. As YouTube is an interactive and dynamic video-sharing platform, rearranging and uploading new videos may change the results. In this study, only Turkish videos related to AMD were evaluated. Evaluating Turkish videos and videos from other languages may change the results. However, because our primary target group was Turkish patients, we aimed to evaluate the quality of Turkish videos and the situation in Turkey by using videos in Turkish, which is the mother tongue.

Despite these limitations, examining the videos by two different ophthalmologists and the consistency of the scores across these evaluations were important in the present study. Furthermore, using multiple scales to assess various facets of the videos' content and quality makes the study's results comprehensive. Finally, our study reveals Turkey's reality. To the best of our knowledge, this is the first study to evaluate the quality of Turkish YouTube videos on AMD published in Turkey.

Conclusion

Our findings demonstrate that the information quality of Turkish AMD-related videos on YouTube is poor to moderate. The number of irrelevant and misleading videos was high. Patients may not be able to distinguish between useful and misleading information. The potential effects of incomplete or unreliable YouTube videos on the patient-physician relationship, as well as patients' perceptions and understanding of their disease, are important. In particular, patients who are in search of a new treatment for their disease may lose confidence in their current treatments with unproven options. Although the quality indices were lower than useful videos, it is noteworthy that patients watched misleading videos that mainly mentioned new and speculative treatments. Publicizing more health-related videos created professionally by health professionals may increase public health awareness, and the Internet may be a valuable tool for delivering high-quality and reliable information to the public.

Footnote

Ethics Committee Approval: The data collected for this study were acquired from publically accessible Youtube videos. This study was not required to obtain Institutional Review Board Approval as it involved only the public access data.

Informed Consent: Not required.

Authorship Contributions

Concept: M.O., I.C.T., G.D.A., Design: M.O., I.C.T., G.D.A., Data Collection or Processing: M.O., I.C.T., G.D.A., Analysis or Interpretation: M.O., I.C.T., G.D.A., Literature Search: M.O., I.C.T., G.D.A., Writing: M.O., I.C.T., G.D.A.

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

Financial Disclosure: This study received no financial support.

References

1. Haab O. Erkrankungen der Macula Lutea. Centralblatt Augenheilkd. 1985;9:384-91.
2. Wong WL, Su X, Li X, et al. Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and metaanalysis. Lancet Glob Health. 2014;2:106-16.
3. Estacio EV, Whittle R, Protheroe J. The digital divide: Examining socio-demographic factors associated with health literacy, access and use of internet to seek health information. J Health Psychol. 2019;24:1668-75.
4. McMullan M. Patients using the Internet to obtain health information: How this affects the patient-health professional relationship. Patient Educ Couns. 2006;63:24-8.
5. Paul M, Dredze M. You Are What You Tweet: Analyzing Twitter for Public Health. Proc Int AAAI Conf Web Soc Media. 2021;5:265-72.
6. Garg N, Venkatraman A, Pandey A, Kumar N. YouTube as a source of information on dialysis: a content analysis. Nephrology (Carlton). 2015;20:315-20.
7. Kumar N, Pandey A, Venkatraman A, Garg N. Are video sharing web sites a useful source of information on hypertension? J Am Soc Hypertens. 2014;8:481-90.

8. Kayabaşı M, Köksaldı S, Durmaz Engin C. Evaluating the quality and reliability of YouTube videos on myopia: a video content analysis. Int Ophthalmol. 2024;44:329.
9. Tanyıldız B, Oklar M. Evaluating the quality, utility, and reliability of the information in uveitis videos shared on YouTube. Int Ophthalmol. 2023;43:549-55.
10. Ozturkmen C, Berhuni M. YouTube as a source of patient information for pterygium surgery. Ther Adv Ophthalmol. 2023;15:25158414231174143.
11. Kaptı HB, Erdem B. Evaluation of the Reliability and Quality of YouTube Videos on Congenital Nasolacrimal Duct Obstruction. Cureus. 2023;15:e36365.
12. İlhan HD, Çetinkaya Yaprak A, Erkan Pota Ç. Quality analysis of YouTube videos about thyroid orbitopathy. Int Ophthalmol. 6-43:2851;2023.
13. Young BK, Verter E, Howard MA. Quality analysis of publicly available videos for pediatric strabismus surgery. J AAPOS. 2020;24:102-4.
14. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health. 1999;53:105-11.
15. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the Internet: Caveant lector et viewor–Let the reader and viewer beware. JAMA. 1997;277:1244-5.
16. Bernard A, Langille M, Hughes S, Rose C, Leddin D, Veldhuyzen van Zanten S. A systematic review of patient inflammatory bowel disease information resources on the World Wide Web. Am J Gastroenterol. 2007;102:2070-7.
17. Palmour N, Vanderbyl BL, Zimmerman E, Gauthier S, Racine E. Alzheimer's disease dietary supplements in websites. HEC Forum. 2013;25:361-82.
18. Rapp AK, Healy MG, Charlton ME, Keith JN, Rosenbaum ME, Kapadia MR. YouTube is the Most Frequently Used Educational Video Source for Surgical Preparation. J Surg Educ. 2016;73:1072-6.
19. Lucatto LFA, Barbosa GCS, Prazeres JMB, et al. Use of surgical videos available on social media among retina surgeons: results of a vitreoretinal specialist survey. Arq Bras Oftalmol. 2023:S0004-27492023005002309.
20. Lucatto LFA, Prazeres JMB, Guerra RLL, et al. Evaluation of quality and utility of YouTube vitreoretinal surgical videos. Int J Retina Vitreous. 2022;8:9.
21. Yildiz MB, Yildiz E, Balci S, Özçelik Köse A. Evaluation of the Quality, Reliability, and Educational Content of YouTube Videos as an Information Source for Soft Contact Lenses. Eye Contact Lens. 2021;47:617-21.
22. Bora K, Das D, Barman B, Borah P. Are internet videos useful sources of information during global public health emergencies? A case study of YouTube videos during the 2015-16 Zika virus pandemic. Pathog Glob Health. 2018;112:320-8.
23. Fernandez-Alvarez MDM, Cachero-Rodríguez J, Fu MR, et al. YouTube videos on lymphedema as an information source for Spanish speaking breast cancer survivors. Support Care Cancer. 2024;32:540.
24. Engin O, Songür K. Evaluation of YouTube videos as a source of information in traumatic brain injury rehabilitation: A cross-sectional study. Medicine (Baltimore). 2024;103:e39254.
25. Gabarron E, Fernandez-Luque L, Armayones M, Lau AY. Identifying Measures Used for Assessing Quality of YouTube Videos with Patient Health Information: A Review of Current Literature. Interact J Med Res. 2013;2:e6.
26. Bae SS, Baxter S. YouTube videos in the English language as a patient education resource for cataract surgery. Int Ophthalmol. 2018;38:1941-5.
27. Altunel O, Sirakaya E. Evaluation of YouTube videos as sources of information about multifocal intraocular lens. Semin Ophthalmol. 2021;36:423-8.

Appendix A. DISCERN scoring system						
		Question rating				
		No	Partially			Yes
Section 1	**Is the publication reliable?**					
1	Are the aims clear?	1	2	3	4	5
2	Does it achieve its aims?	1	2	3	4	5
3	Is it relevant?	1	2	3	4	5
4	Is it clear what sources of information were used to compile the publication (other than the author or producer)?	1	2	3	4	5
5	Is it clear when the information used or reported in the publication was produced?	1	2	3	4	5
6	Is it balanced and unbiased?	1	2	3	4	5
7	Does it provide details of additional sources of support and information?	1	2	3	4	5
8	Does it refer to areas of uncertainty?	1	2	3	4	5
Section 2	**How good is the quality of information on treatment choices?**					
9	Does it describe how the treatment works?	1	2	3	4	5
10	Does it describe the benefits of each treatment?	1	2	3	4	5
11	Does it describe the risks of each treatment?	1	2	3	4	5
12	Does it describe what would happen if no treatment is used?	1	2	3	4	5
13	Does it describe how the treatment choices affect overall quality of life?	1	2	3	4	5
14	Is it clear that there may be more than one possible treatment choice?	1	2	3	4	5
15	Does it provide support for shared decision-making?	1	2	3	4	5
Section 3	**Overall rating of the publication**					
16	Based on the answers to the above questions, rate the overall quality of the publication as a source of information for patients about treatment choices.	**Low**	**Moderate**			**High**
		1	2	3	4	5

Appendix B. Journal of the American Medical Association (JAMA) Scoring System	
Authorship	The authors and contributors, the institutions to which they adhere, and their credentials should be provided.
Disclosure	Conflicts of interest, funding, sponsorship, advertising, support, and video ownership should be fully disclosed.
Attribution	All copyright data should be clearly listed, and references and sources for all content should be stated.
Currency	The initial date of posted content and dates of updates should be provided.

Appendix C. Global Quality Scoring (GQS) System	
(1)	Poor quality, very unlikely to be of any use to patients
(2)	Poor quality but some information present, of very limited use to patients
(3)	Suboptimal flow, some information covered but important topics missing, somewhat useful to patients
(4)	Good quality and flow, most important topics covered, useful to patients
(5)	Excellent quality and flow, highly useful to patients

Original Article

DOI: 10.4274/haseki.galenos.2024.9593
Med Bull Haseki 2024;62:201-207

Pediatricians' Knowledge of Screen Use and Identifying of Own Children's Screen Use Habits

Senay Guven Baysal*, Muhammet Tas**

*University of Health Sciences Turkey, Gazi Yasargil Training and Research Hospital, Clinic of Developmental Pediatrics, Diyarbakir, Turkey
**University of Health Sciences Turkey, Gazi Yasargil Training and Research Hospital, Clinic of Pediatrics, Diyarbakir, Turkey

Abstract

Aim: Guidelines for children have been developed to reduce the risks associated with screen use. We aimed to learn the knowledge of pediatricians with children aged 0-6 years about screen use and to identify the screen use habits of their own children.

Methods: This cross-sectional study was conducted among 212 pediatricians who had children aged 0-6 years between August and September 2023. Survey questions were created according to the American Academy of Pediatrics screen time recommendations. The questions were delivered to the pediatricians via social media.

Results: Of the pediatricians, 64.2% were mothers and 35.8% were fathers. The mean level of screen-use knowledge was 7.73 (maximum score =10). The pediatricians' average screening time during the day was 2.44 hours. There was a significant inverse correlation between the pediatricians' screen use knowledge level, their average screen time, and the duration of television use when the child was at home (p=0.035 and p=0.010, respectively). There was a difference between the institutions in which the pediatrician worked in terms of screen use knowledge levels (p=0.018).

Conclusion: Pediatricians were knowledgeable about screening guidelines. It is important for pediatricians to approach screening from a holistic perspective and integrate it into their practices.

Keywords: Child, media, screen time, pediatricians, physicians

Introduction

The number of screens in society is increasing day by day in our country, as in the world. One of the biggest reasons for this is that technology has become the building block of our lives, both in business and daily life. On the other hand, the ease of use, portability, and content capacity of screens (television, tablet, computer, etc.) are other factors that increase the use of screen-based media in society. However, recent studies have shown that screen media use starts at an early age (1,2). Young children and babies in this digital generation are exposed to more technology and use more devices than in previous years, thus increasing their screen time (3). Screen time definitions and guidelines for healthy screening in early childhood are still under debate (4).

Very early-onset screen exposure carries significant risks for children's health and development. Most studies have shown that children start using screen media before the age of two (2,5). Previous studies have shown that screen exposure in young children causes problems such as speech-language delay, sleep disturbance, obesity, school failure, depression, and anxiety (6-10). Although being healthy is possible with physical, psychological, and social well-being, the role of pediatricians is as effective in providing counseling to parents as it is in protecting the health of children. Pediatricians play a significant role in both protecting the health of children and counseling families about screen-based media guidelines.

To reduce the risks associated with screen use, guidelines have been developed that recommend

Address for Correspondence: Senay Guven Baysal, University of Health Sciences Turkey, Gazi Yasargil Training and Research Hospital, Clinic of Developmental Pediatrics, Diyarbakir, Turkey
Phone: +90 505 365 72 60 **E-mail:** senay177@yahoo.com **ORCID:** orcid.org/0000-0002-5454-923X
Received: 07.11.2023 **Accepted:** 09.09.2024

appropriate time limits and considerations for screen time for children. The Canadian guidelines recommend no screen time for children under 2 years old; screen time should be 1 hour or less per day for children aged 2-5 years; and screen time should not be a routine part of childcare for children under 5 years old (2). The American Academy of Pediatrics (AAP) updated its media use recommendations in 2016. Although it does not provide a specific screen time limit, it emphasizes the importance of considering not only the quantity or duration of interactions but also the quality of interactions with digital media. According to the AAP guidelines, screen time should be avoided in children aged below 18 months, except for video chat. Between 18 and 24 months of age, parental involvement is very important and can be brief, provided that a parent finds high-quality programs and watches them with them. For children aged 2-5 years, screen time should be less than 1 hour a day, and children should watch with their parents to interpret and discuss what they watch. A family media plan with consistent rules is recommended. Six years and older should limit activities involving screens. All screens should be turned off during family meals and outings, and parents should avoid using screens as pacifiers, babysitters, or to stop tantrums. Screens should be turned off 30-60 minutes before bedtime and removed from the bedroom (11).

Our hypothesis is that pediatricians should have up-to-date knowledge and awareness about screening to be more sensitive when counseling families on these issues. In this study, we aimed to learn the knowledge of pediatricians who have children aged 0-6 years in our country about screen media use and to define the screen media use habits of their own children. With the results we obtained in our study, we expect to increase awareness about screen use and to create new suggestions and discussions on this subject by conducting research on a subject that has not yet been studied with pediatricians.

Methods

Compliance with Ethical Standards

Approval for the study was obtained from the Institutional Ethics Committee of the University of Health Sciences Turkey, Gazi Yasargil Training and Research Hospital (dated: 04.08.2023, approval no.: 491). This study was conducted in accordance with the Helsinki Declaration.

Study Design

A cross-sectional study was planned based on previous studies on screen and AAP screen time recommendations, and a cross-sectional study was planned by creating questions with yes/no, multiple choice, and multiple answers (12). A total of 217 pediatricians completed our questionnaire between August 2023 and September 2023, but the responses of 5 were not evaluated because of inconsistencies in their responses, and the study was conducted with 212 pediatricians with children aged 0-6 years who responded to our questionnaire between August and September 2023 (Figure 1).

The questions consisted of 3 parts. The first section provides demographic information about pediatricians. Age, gender, marital status, number of children, titles, geographical region in which they live, institution in which they work, years of employment, age, job, and educational status of their spouses were questioned. In the second section, pediatricians were questioned about their screen-use habits at home, their attitudes toward their children's screen use, the technological devices their children own, the technological devices and programs they allow their children to use, the amount of screen time they and their children spend in front of the screen, the age at which their children were introduced to media tools, and whether screen time is restricted or why. "Screen time" includes any time spent in front of a television or other screen device such as video games, computers, smartphones,

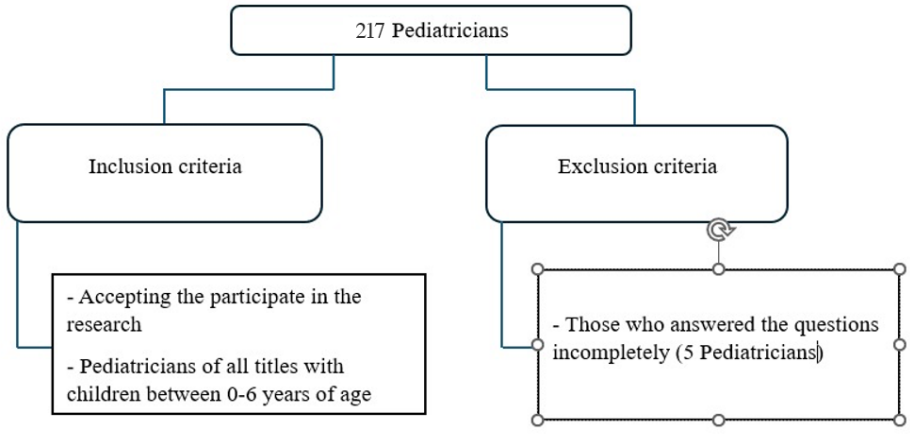


Figure 1. Flow diagram of the study

and tablets (13). "Screen media use" is defined as the use of these devices. In the third section, pediatricians' knowledge on screen use was questioned. There is a table with 10 questions and one multiple-choice question. The aim of this section was to determine the knowledge levels of pediatricians and their requests for screen use training. The questions about knowledge about screen use were scored, and the total score was calculated as "10" when all questions were answered correctly.

The questions were converted into questionnaires via Google Forms and delivered to the pediatricians via social media (e-mail, WhatsApp, telegram). Before starting the survey, brief information and contact information were given, and all questions were considered answered. Physicians were reminded 1 week after the first mailing, and the study was terminated after 1 month. Some questions were evaluated as ratios, and others were compared according to the demographic information of the physician.

Statistical Analysis

The conformity of the data to the normal distribution was examined using the Shapiro-Wilk test. The Mann-Whitney U test was used for comparisons between two independent and non-normally distributed groups, and the Kruskal-Wallis H test was used for comparisons between three groups. The Spearman correlation coefficient was used to determine the relationship between quantitative variables. Statistical analysis was performed using IBM SPSS Statistics for Windows. Version 25.0. Armonk, NY: IBM Corp., and $p<0.05$ was considered statistically significant.

Results

Pediatricians of all titles participated in the study, and the majority of them were pediatric specialists and residents. Physicians from all regions participated in our study, mostly from the Aegean region and southeast Anatolia. State hospitals were the majority of the institutions where the participating pediatricians worked.

Table 1 presents additional sociodemographic data. The average screen time of pediatricians during the day was determined as 2.44 hours. Children's screen time on weekdays and weekends was categorized as intervals, and it was observed that as the quantitative value increased, the time allocated to the screen increased. Accordingly, on weekdays, the average screen time was 2.41 hours, and on weekends, it was 2.46 hours. When asked about the programs they allowed their children to use, more than half of them stated that they frequently watched cartoons and educational videos created for children. When the age at which children first watched electronic devices was considered, it was learned that television and telephone

Table 1. Socio-demographic characteristics

Variables	n	%
Parental status		
Mother	136	64.2
Father	76	35.8
Marital status		
Married	200	94.3
Single	12	5.7
Number of children aged 0-6		
1	180	84.9
2	32	15.1
Child gender		
Female	104	49.1
Male	108	50.9
Spouse working status		
Yes	195	92
No	17	8
Spouse being a pediatrician		
Yes	35	16.5
No	177	83.5
Child's caregiver		
Spouse	26	12.3
Relative	72	34.0
Babysitter	51	24.1
Nursery	63	29.6
Title		
Pediatric Resident	45	21.2
Pediatric Specialist	94	44.3
Subspecialist	53	25.1
Assistant Professor	9	4.2
Assoc. and Prof.	11	4.2
Institution		
State hospital	114	53.8
University hospital	75	35.4
Private hospital	14	6.6
Clinic	9	4.2
Region		
Mediterranean	19	9.0
Marmara	37	17.5
Central Anatolia	27	12.7
Aegean	48	22.6
Black Sea	12	5.7
Southeast Anatolia	47	22.2
Eastern Anatolia	22	10.4

were frequently used between the ages of 1-2 years (33% and 25%, respectively), while tablets were used between the ages of 2-3 years. More than half of them stated that they had accompanied their child or explained the content of the screen. Almost all of the pediatricians (99.29%) reported that they restricted their children's screen use (Table 2).

Pediatricians were asked 10-point knowledge questions in line with the APA screen time recommendations, and the mean knowledge level of screen use was 7.73. Approximately 91% of the pediatricians answered more than 6 of the questions correctly. It was observed that 26.42% of the pediatricians answered 90% of the screen use knowledge level questions correctly (Table 3). There was no difference in the knowledge level of screen media use among pediatricians according to their gender (p=0.668). There was no difference between knowledge levels of screen use according to marital status and the number of children they had (p=0.880 and p=0.491, respectively). It was also determined that knowledge levels did not differ according to their titles (p=0.231).

There was a difference in the level of knowledge on screen use according to whether the spouse was a pediatrician or not (p=0.001). There was a difference between the institutions of employment in terms of screen use knowledge levels (p=0.018). The mean level of knowledge about media use among pediatricians working in state hospitals was 7.43, and the mean level of knowledge about media use among pediatricians working in university hospitals was 8.05, and the difference between these values was significant (p=0.017). At the same time, we determined that there was a significant difference between the mean knowledge level of screen use among pediatricians working in state hospitals and pediatricians working in their own clinics (p=0.015) (Table 4). There was a significant reverse correlation between pediatricians' knowledge of screen use and the average screen time of pediatricians and the duration of television programming when the child was at home (p=0.035 and p=0.010, respectively). There was a significant inverse correlation between the average screening time of pediatricians and the year of employment of the pediatrician (p=0.007). There was a significant relationship between the average screen time of the spouse, the time the television was on when the child was at home and the average screen time of the pediatrician (p<0.001 and p=0.016, respectively). A significant correlation was also found between the average screen time of the spouse and the time the television was on while the child was at home (p=0.018).

There was a significant relationship between the age of the pediatrician, the age of the spouse, the working years of the pediatrician and the duration of television

Table 2. Evaluation of screen-based media use

Variables	n	%
What are your screen use habits at home?		
Social media	160	32.92
Film, series, competition program, etc.	147	30.25
Education	108	22.22
Game	55	11.32
I don't have except for my work life	16	3.29
Which screen-based technological devices do you have at home?		
Smartphone	210	27.60
Computer	198	26.02
Smart TV	181	23.78
Tablet	127	16.69
Game tool (PlayStation, Xbox, etc.)	45	5.91
What programs do you allow your child to watch on screen?		
Watching cartoons	163	41.79
Educational videos for children	110	28.21
Playing games	69	17.69
Adult programs	13	3.33
I do not allow my child to watch	35	8.97
What is your reason for allowing your child to spend time with screens?		
Making time for my own personal affairs	122	24.16
Entertainment	118	23.37
Calming down	64	12.67
Feeding food	63	12.48
Education	62	12.28
Accompanying other family members	41	8.12
I do not allow my child to spend time with screens	35	6.93
What is your habit of accompanying your child during screen media use?		
I will accompany him/her and be with him/her	81	30.45
I'll tell him/her the contents of the screen	68	25.56
I'll be with him/her as I watch him/her silently	61	22.93
I do not accompany my child	34	12.78
Other	22	8.27
What is your reason for limiting your child's screen use?		
Concern about delaying development	166	29.54
The young age of my child	136	24.20
Harmful content	135	24.02
Eye disease concern	103	18.33
Other	18	3.20
I do not restrict screen usage	4	0.71
What are the negative effects of screen use on children?		
Speech and language delay	210	22.29
Cognitive and sensory delay	200	21.23
Sleep disorders	200	21.23
Obesity	170	18.05
Skeletal and muscular system diseases	162	17.20

programming while the child was at home (p<0.05) (Table 5).

It was observed that 81.6% of pediatricians did not receive any training on screen use guidelines. However, 80.7% of the physicians wished to receive training on screen usage. In addition, 89.6% of the pediatricians were found to make suggestions to their patients regarding screening guidelines.

Table 3. Pediatricians' level of knowledge about screen use

Variables	n	%
Children between 2 and 5 years of age can be allowed more than 1 hour of screen time per day, accompanied by their parents.		
Correct	69	32.5
Wrong	143	67.5
Children 6 years and older can use screens without parental supervision.		
Correct	24	11.3
Wrong	188	88.7
Children between 6 and 8 years old may be allowed more than 2 hours of screen time per day.		
Correct	55	25.9
Wrong	157	74.1
Children should not be given phones at mealtimes and before bedtime to calm them down.		
Correct	183	86.3
Wrong	29	13.7
Babies between 0-18 months may be allowed to video-chat with family members under parental supervision.		
Correct	143	86.3
Wrong	69	13.7
Children can be given mobile phones in the car or in crowded places.		
Correct	21	9.9
Wrong	191	90.1
Children aged 2-5 can watch educational cartoons.		
Correct	162	76.4
Wrong	50	23.6
Screen may be allowed when feeding children between 18-24 months.		
Correct	16	7.5
Wrong	196	92.5
Children between the ages of 4-5 can use screen time with their parents for a maximum of 1 hour per day.		
Correct	186	87.7
Wrong	26	12.3
The screen-based media is not recommended during the first 24 months.		
Correct	202	95.3
Wrong	10	4.7

Discussion

Screen media usage has become a common and serious problem among children. Screening time is associated with less energy expenditure, passive receptivity to the surrounding environment, and the possibility of encountering low-quality and inappropriate content (14). The health problems caused by screen exposure highlight the need for a guide on this issue, and more studies should be conducted (8,9). The place of screens in our lives has become widespread with the use of screens by adults both at work and on social media. According to data from the Turkish Statistical Institute for 2023, 95.5% of households had Internet access in Turkey, and the internet usage rate was 87.1% among individuals aged 16-74 (15). According to the European Children Online Project data for Turkey, the internet usage rate by parents was 23.5% for women and 49% for men (16).

Table 4. Comparison of screen use knowledge levels and variables

Variables	Knowledge level				p-value
n=212	Mean	SD	Median	Min.-Max.	
Parental status					
Mother	7.77	1.46	8	3:10	0.668[a]
Father	7.66	1.71	8	4:10	
Marital status					
Married	7.74	1.53	8	3:10	0.880[a]
Single	7.58	1.93	8	4:10	
Number of children					
1	7.73	1.35	8	3:10	0.491[b]
2	7.53	1.73	8	4:10	
>3	7.88	1.86	8	4:10	
Title					
Pediatric Resident	8.09	1.43	9	5:10	0.231[b]
Pediatric Specialist	7.49	1.64	8	3:10	
Subspecialist	7.81	1.53	8	4:10	
Assistant Professor	8.22	1.39	8	6:10	
Assoc. and Prof.	7.55	1.29	8	6:10	
Spouse working status					
Yes	7.77	1.54	8	3:10	0.272[a]
No	7.29	1.72	7	4:10	
Spouse being a pediatrician					
Yes	8.49	1.50	9	5:10	**0.001[a]**
No	7.58	1.52	8	3:10	
Institution					
State hospital	7.43	1.59	8	3:10	**0.018[b]**
University hospital	8.05	1.36	8	5:10	
Private hospital	7.78	1.76	8	4:10	
Clinic	8.78	1.39	9	6:10	

p<0.05 significance level, [a]Mann-Whitney U test, [b]Kruskal-Wallis H test
SD: Standard deviation, Min.-Max.: Minimum-maximum

Table 5. Comparison of screen use knowledge level of pediatricians according to demographic data				
Variables (n=212)	**Screen use knowledge level**	**Average screen time**	**Spouse's average screen time**	**Screen time when child is at home**
Pediatrician's age	r_s=-0.098 p=0.156	r_s=-0.119 p=0.085	r_s=-0.160 p=0.020	r_s=0.295 p≤0.001
Age of spouse	r_s=-0.081 p=0.240	r_s=-0.107 p=0.121	r_s=-0.059 p=0.392	r_s=0.228 p=0.001
Pediatrician's year of practice	r_s=-0.074 p=0.286	r_s=-0.185 p=0.007	r_s=-0.110 p=0.112	r_s=0.246 p≤0.001
Average screen time	r_s=-0.145 p=0.035	-	-	-
Spouse's average screen time	r_s=-0.113 p=0.101	r_s=0.604 p≤0.001	-	-
TV time with child at home	r_s=-0.176 p=0.010	r_s=0.166 p=0.016	r_s=0.162 p=0.018	-
p<0.05 significance level, r_sSpearman correlation coefficient				

In addition, successive disasters in our country (Coronavirus disease-2019, Kahramanmaras earthquake) unfortunately revealed the reality of virtual education and necessarily increased children's screen time at home (17). Inappropriate or uncontrolled screen use can disrupt many important activities, such as spending interactive time with other family members, reading books, playing creative outdoor games, or engaging in physical activities, and constitutes time spent without activities that are important for early childhood development (18,19). This raises many questions about the negative consequences on the cognitive, social, and emotional development of children during early childhood (20). Considering the risks that occur before and the harms that occur after exposure, it is important for pediatricians to advise patients and their families regarding reliable and accurate information by adhering to screening guidelines. In the 2016 APA report, it was revealed that only 16% of pediatricians interviewed families about screen-related questions, and in the same report, 29% of parents reported that they trusted pediatricians for advice on media use (12). In our study, most pediatricians advised their patients about screening. However, when the international literature was reviewed, it was learned that in previous studies, pediatricians were not sufficiently supported in guiding and discussing these issues with the family about screen use in young children (21,22). Regardless of the age of the child, pediatricians should ask parents how much screen time their child spends, advise accordingly, and warn them about the consequences of excessive use (19).

In the study by Amos et al. (23), all 53 participants reported that they were aware of the existing screen time guidelines, and the majority reported moderate to excellent knowledge on these issues. A total of 43.4% of respondents reported that they had made screen time recommendations in the past 12 months and/or that they almost always planned and made these recommendations during routine clinical visits (23). When we determined their level of knowledge with the questions we designed based on the guidelines, approximately 91% of pediatricians responded correctly to more than six out of ten questions, and 89.6% of pediatricians made recommendations to their patients about screening. In the same study, Canadian pediatricians mentioned behavioral problems, mental health concerns, obesity, and sedentary lifestyle concerns with increasing screen time. In a study by Costa et al. (24), television viewing in early childhood was found to be associated with overweight and the onset of obesity. In our study, pediatricians also reported that excessive screen use may have negative effects on language-speech, cognitive and sensory delays, sleep disorders, obesity, and skeletal-muscular system diseases (7). They had the same concerns as the other study.

Study Limitations

The fact that pediatricians who are aware of this issue participated in our study can be interpreted as an optimistic but weak aspect of our study. It cannot be generalized to the average knowledge level and average screening time of all pediatricians. Furthermore, we did not involve family physicians, general practitioners, and health professionals working in pediatrics, such as nurses, whose target patient population included children. We restricted our findings to pediatricians' screen knowledge level and time. Because we did not include another group, we were unable to make comparisons between groups.

Despite these limitations, this study has some strengths. We tried to raise awareness regarding this issue in pediatrics using data we obtained from pediatricians from every region, title, and institution. This is also the first study on this issue in the national literature that provides insight into the awareness of pediatricians in Turkey regarding current screening time guidelines and their practices.

Conclusion

This study found that pediatricians with preschool-aged children were knowledgeable about screen use. It is important for pediatricians to approach screening from a holistic perspective and integrate it into their practices.

Footnote

Ethics Committee Approval: Ethical approval was obtained from the Institutional Ethics Committee of the University of Health Sciences Turkey, Gazi Yasargil Training and Research Hospital (dated: 04.08.2023, approval no.: 491).

Informed Consent: Informed consent was obtained from all participants.

Authorship Contributions

Concept: S.G.B., M.T., Design: S.G.B., M.T., Data Collection or Processing: S.G.B., M.T., Analysis or Interpretation: S.G.B., M.T., Literature Search: S.G.B., M.T., Writing: S.G.B.

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

Financial Disclosure: This study received no financial support.

References

1. Ponti M. Screen time and preschool children: Promoting health and development in a digital world. Paediatr Child Health. 2023;28:184-92.
2. Canadian Paediatric Society, Digital Health Task Force, Ottawa, Ontario. Screen time and young children: Promoting health and development in a digital world. Paediatr Child Health. 2017;22:461-77. Erratum in: Paediatr Child Health. 2018;23:83.
3. Savina E, Mills JL, Atwood K, Cha J. Digital media and youth: A primer for school psychologists. Contemp Sch Psychol. 2017;21:80-91.
4. Beatty C, Egan SM. Screen time in early childhood: A review of prevalence, evidence and guidelines. An Leanbh Ó. 2020;13:17-31.
5. Bedford R, Saez de Urabain IR, Cheung CH, Karmiloff-Smith A, Smith TJ. Toddlers' Fine Motor Milestone Achievement Is Associated with Early Touchscreen Scrolling. Front Psychol. 2016;7:1108.
6. Collet M, Gagnière B, Rousseau C, Chapron A, Fiquet L, Certain C. Case-control study found that primary language disorders were associated with screen exposure. Acta Paediatr. 2019;108:1103-9.
7. Radesky JS, Silverstein M, Zuckerman B, Christakis DA. Infant self-regulation and early childhood media exposure. Pediatrics. 2014;133:1172-8.
8. Muppalla SK, Vuppalapati S, Reddy Pulliahgaru A, Sreenivasulu H. Effects of Excessive Screen Time on Child Development: An Updated Review and Strategies for Management. Cureus. 2023;18:e40608.
9. Qu G, Hu W, Meng J, et al. Association between screen time and developmental and behavioral problems among children in the United States: evidence from 2018 to 2020 NSCH. J Psychiatr Res. 2023;161:140-9.
10. Barr R. Memory constraints on infant learning from picture books, television, and touchscreens. Child Dev Perspect. 2013;7:205-10.
11. Livingstone S. New 'screen time' rules from theAmerican Academy of Pediatrics. LSE. 2016.
12. Council on Communications and Media. Media and Young Minds. Pediatrics. 2016;138:e20162591.
13. Kabali HK, Irigoyen MM, Nunez-Davis R, et al. Exposure and use of mobile media devices by young children. 2015;136:1044-50.
14. Brooker L, Blaise M, Edwards S. The SAGE handbook of play and learning in early childhood. SAGE Publications Ltd. 2014.
15. TÜİK DA. Hanehalkı Bilişim Teknolojileri (BT) Kullanım Araştırması. 2023. Erişim adresi: https://data.tuik.gov.tr/Bulten/Index?p=Hanehalki-Bilisim-Teknolojileri-(BT)-Kullanim-Arastirmasi-2023-49407
16. Kaşıkçı DN, Kürşat Çağıltay, Karakuş T, Kurşun E, Ogan C. Internet habits and safe Internet use of children in Turkey and Europe. Education and Science. 2014.
17. Ayer Ç, Ekinoğlu E, Kaner Tohtak G. Evaluation of Changing Lifestyle and Nutrition Habits of Preschool Children During COVID-19 Process: Cross Sectional Study. Turkiye Klinikleri J Health Sci. 2024;9:256-65.
18. Posjećeno AAoPJ. Children and media tips from the American Academy of Pediatrics. 2018.
19. Assathiany R, Guery E, Caron FM, et al. Children and screens: A survey by French pediatricians. Arch Pediatr. 2018;25:84-8.
20. Priftis N, Panagiotakos D. Screen Time and Its Health Consequences in Children and Adolescents. Children (Basel). 2023;10:1665.
21. Straker L, Zabatiero J, Danby S, Thorpe K, Edwards S. Conflicting Guidelines on Young Children's Screen Time and Use of Digital Technology Create Policy and Practice Dilemmas. J Pediatr. 2018;202:300-3.
22. Banker SL, Fenick AM, Qin L, Talwalkar JS. Perceptions of Pediatric Residents Regarding Counseling About Use of Social Networking Sites. Clin Pediatr (Phila). 2018;57:1275-80.
23. Amos KA, Ogilvie JD, Ponti M, Miller MR, Yang F, Ens AR. Paediatricians' awareness of Canadian screen time guidelines, perception of screen time use, and counselling during the COVID-19 pandemic. Paediatr Child Health. 2023;28:357-61.
24. Costa S, Johnson W, Viner RM. Early childhood TV viewing and subsequent BMI trajectories to mid-adulthood in the 1970 British Cohort Study. Presented at: ISBNPA 2014 Annual Meeting, San Diego, CA, USA.

Original Article

DOI: 10.4274/haseki.galenos.2024.9981
Med Bull Haseki 2024;62:208-215

Comparison of Intracranial Hemorrhage and Clinical Outcomes Among Patients Undergoing Mechanical Thrombectomy with and without Thrombolytic Therapy

Zulfikar Memis, Irem Kuru

University of Health Sciences Turkey, Istanbul Haseki Training and Research Hospital, Clinic of Neurology, Istanbul, Turkey

Abstract

Aim: It remains controversial whether intravenous tissue plasminogen activator (IV tPA) increases complications in patients undergoing mechanical thrombolysis. In this study, we aimed to show the effect of IV tPA administration on complications.

Methods: In this cross-sectional study, the records of patients who were followed up at the stroke center and underwent mechanical thrombectomy (MT) between 2022 and 2023 were retrospectively reviewed. Demographic data, cerebral angiography data, neuroimaging time, medical history, and medication use; neurologic examination findings at baseline, 24th hour, and 3rd month were obtained from the patient files. Patients were divided into two groups according to whether intracranial tPA was administered before MT and two groups according to the presence of intracranial hemorrhage at 24 hours.

Results: A total of 172 patients [94 women (54.7%) and 78 men (45.3%)] were included in the study. Mean age was 67.6±14.7 years. At 24 h, the rate of symptomatic intracranial hemorrhage (sICH) was significantly (p=0.004) higher in the thrombectomy group than in the tPA plus thrombectomy group. The admission (p=0.033) and 24-hour National Institutes of Health Stroke Scale (p=0.001) scores were significantly higher in the sICH group than in the non-sICH group. Third-month modified Rankin scale score (p=0.003), diastolic embolism rate (p=0.009), and Tan score (p=0.007) were significantly higher in the sICH group than in the non-sICH group.

Conclusion: Intravenous tissue plasminogen activator did not increase sICH or distal embolism in patients undergoing MT, and there was no difference in terms of favorable clinical outcomes. Symptomatic intracranial hemorrhage was associated with increased mortality and poor clinical outcome.

Keywords: Thrombectomy, thrombolytic therapy, intracranial hemorrhage

Introduction

For acute ischemic stroke, clinicians use intravenous tissue plasminogen activator (IV tPA) alone or in combination with mechanical thrombectomy (MT). Intravenous tissue plasminogen activator has been shown to reduce the rate of disability in the first 3 hours after ischemic stroke (1). In 2008, IV tPA was shown to be effective for functional independence in patients with ischemic stroke who received IV tPA in the first 4.5 hours, but intracerebral bleeding rates were higher in these patients (2). Recent studies have shown that in addition to IV tPA, MT in large vessel occlusion reduces the rate of disability after ischemic stroke. Since IV tPA and MT were administered together in almost all of these studies, it remains a matter of debate whether IV tPA is beneficial for large vessel occlusions or whether it increases the complication rate. Some of the topics of these debates include whether it increases potential bleeding rates, whether it increases distal embolism by lysing the thrombus, or whether it increases door-puncture times. Along with these studies, the American Heart Association/American Stroke Association has recommended that patients with ischemic stroke receive IV tPA in the first 4.5 hours if there are no contraindications and MT if there is large vessel occlusion (3-7).

We hypothesized that administration of IV tPA before thrombectomy would not cause intracranial hemorrhage

Address for Correspondence: Zulfikar Memis, University of Health Sciences Turkey, Istanbul Haseki Training and Research Hospital, Clinic of Neurology, Istanbul, Turkey
Phone: +90 555 667 22 98 **E-mail:** zulfikarmemis1@gmail.com **ORCID:** orcid.org/0000-0002-5985-900X
Received: 09.07.2024 **Accepted:** 09.09.2024

or distal embolism after the procedure. The present study aimed to investigate the effect of IV tPA administration on complications in patients undergoing mechanical thrombectomy.

Materials and Methods

Compliance with Ethical Standards

This study was approved by the Clinical Research Ethics Committee of the University of Health Sciences Turkey, Istanbul Haseki Training and Research Hospital (dated: 27.12.2023, approval no.: 260-2023) of the Declaration of Helsinki. Informed consent was obtained from all patients.

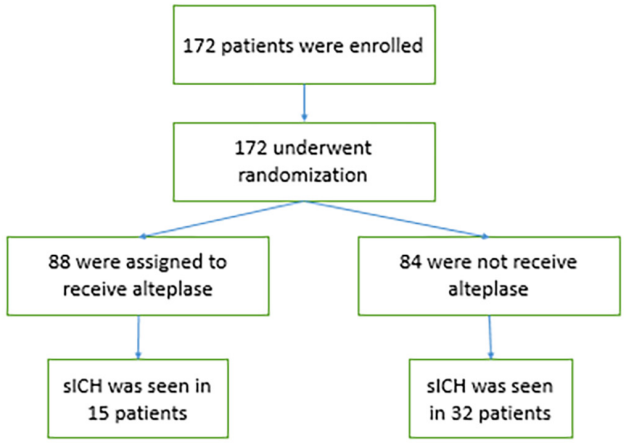
Study Design

This study had a cross-sectional study design. The records of patients who underwent IV tPA and/or MT at stroke centers between 2022 and 2023 were retrospectively reviewed. Demographic data, angiography data, imaging times, medical history, medication use, and neurological examination findings at baseline, 24th hour, and 3rd month were obtained from the patient files. Patients were divided into 2 groups according to whether intracranial tPA was administered before MT and 2 groups according to the presence of intracranial hemorrhage at 24 hours (Graph 1).

Patient Selection and Endovascular Treatment

Basilar artery, posterior cerebral artery P1 segment, internal carotid artery (ICA) (tandem, T, L) or middle cerebral artery (MCA) M1 segment occlusion, no bleeding on brain computed tomography (CT), Alberta Stroke Program Early CT (ASPECT) score >6, patients with a National Institutes of Health Stroke Scale (NIHSS) score of ≥6, a pre-stroke modified Rankin Scale (mRS) score of 0-1, a symptom-door time of less than 4.5 hours, a symptom-door time of 4.5-16 hours with diffusion weighted imaging/fluid-attenuated inversion recovery (FLAIR) discordance, and patients who underwent MT were included in the study. Table 1 presents the inclusion criteria.

Patients were evaluated by a neurologist at the emergency department presentation. All patients underwent noncontrast brain CT and CT angiography (CTA). Patients with no contraindications admitted in the first 4.5 hours after symptoms and those with no contraindications admitted for less than 16 hours with no ischemic lesion reflected on magnetic resonance imaging FLAIR sequence were started on alteplase at a dose of 0.9 mg/kg in the emergency department and then taken to the angio unit for MT. Mechanical thrombectomy was performed under conscious sedation or general anesthesia using a monoplane angiography device. The femoral artery was used as the access site for the procedure. A 6 French (F) guiding catheter (Destination, Terumo, Tokyo, Japan) was then placed in the subclavian artery, vertebral artery, common carotid artery, or cervical segment of the ICA. A distal access catheter (Catalyst 5F-6F Stryker, Kalamazoo, Michigan), a microcatheter (Rebar, Medtronic, Minneapolis, USA), and a 0.014-inch microwire (synchroo stryker) were used. Mechanical thrombectomy was performed using either stent retriever thrombectomy (isolated stent retriever, ARTS, SAVE, solumbra) or aspiration (ADAPT) techniques. The stent retriever was deployed in the occluded segment in the MT with the appropriate size (Trevo, Stryker, Kalamazoo, Michigan, USA; Thrombite, Zylox-Tonbridge, Hangzhou, China; Solitaire X, Medtronic, Minneapolis, USA). If the procedure failed after two attempts, the technique was changed. If the procedure was not successful despite seven thrombectomy procedures, the procedure was terminated. In tandem occlusions, direct aspiration was performed first, and if unsuccessful,



Graph 1. Numbers of patients
sICH: Symptomatic intracranial hemorrhage

Table 1. Inclusion criteria

18-90 years old
Large vessel occlusion
Symptom-door time of less than 4.5 hours
Symptom-door time between 4.5 and 16 hours and no ischemic lesion reflected in the MRI image on the FLAIR sequence
NIHSS >6
Pre-admission mRS 0-1
ASPECT >6

MRI: Magnetic resonance imaging, mRS: Modified Rankin Scale, NIHSS: National Institutes of Health Stroke Scale, ASPECT: Alberta Stroke Program Early Computed Tomography Score (CT-0 score), FLAIR: Fluid-attenuated inversion recovery

balloon angioplasty was performed at the internal carotid origin. In cases of reocclusion at the ICA origin despite balloon angioplasty, carotid artery stenting was performed after the administration of 300 mg of acetylsalicylic acid and 300 mg of clopidogrel.

We followed up patients who underwent MT in the intensive care unit after the procedure. We monitored the blood pressure and neurologic examination every 30 minutes for the first 2 hours, and then every hour thereafter. We performed a brain CT scan 24 hours after MT. The brain CT scan revealed no bleeding; we started antiaggregant or anticoagulant therapy to address the etiology.

Clinical Evaluation Scales

Alberta Stroke Program Early CT scores were rated on cranial CT (8). Modified Tan scoring was used to obtain a collateral score on the CTA (9). The recanalization level was evaluated according to the modified treatment in cerebral ischemia (mTICI) classification. Accordingly, mTICI 0 was defined as no flow, mTICI 1 as filling of the distal MCA but no blood supply to the cortical branches, mTICI 2a as less than half of the MCA irrigation area, mTICI 2b as more than half of the MCA irrigation area, mTICI 2c as the entire MCA irrigation area but more slowly than the normal side, and mTICI 3 as complete recanalization. After the first thrombectomy attempt, mTICI 2c-3 recanalization was considered first-pass recanalization (10). Bleeding causing an increase of ≥4 points according to admission NIHSS score was defined as symptomatic ICH. Disability status on day 90 was evaluated with mRS.

Statistical Analysis

SPSS 28.0 software was used for the analyses. Mean, standard deviation, median minimum, maximum, frequency, and ratio values were used in the descriptive statistics of the data. The distribution of variables was measured by Kolmogorov-Smirnov. Kruskal-Wallis, Mann-Whitney U tests were used to analyze quantitative, independent data. The chi-square test was used in the analysis of qualitative independent data, and the Fisher's exact test was used when chi-square test conditions were not met. The results were evaluated at a significance level of $p<0.05$.

Results

A total of 172 patients [94 women (54.7%) and 78 men (45.3%)] who met the inclusion criteria (Table 1) between 2022 and 2023 were included in the study. The mean age was 67.6±14.7 years, the mean ASPECT at presentation was 9.1±1.1, and history of previous stroke was 33 patients (19.2%). IV tPA was administered in 88 (51.2%) patients. The mean symptom-puncture time was 250.6±174.3 minutes, symptom-canalization time was 308.3±181.2 minutes, puncture-canalization time was 56.8±24.0 minutes, door-imaging time was 19±15.8 minutes, door-IV tPA time was 51.2±20.9 minutes, and door-puncture time was 56.5±17.5 minutes. The successful recanalization rate (TICI 2B, 2C, 3) was 86.1%. At 24 hours, intracranial hemorrhage was seen in 47 (27.3%) patients (Table 2).

Table 2. General information about patients undergoing thrombectomy

		Min.-Max.	Mean ± SD/n%
Age		21.0-90.0	67.6±14.7
Sex	Woman		94/54.7%
	Man		78/45.3%
Cigarette smoking	(-)		99/57.6%
	(+)		73/42.4%
Comorbid disease			
Hypertension			100/58.1%
Atrial fibrillation			72/41.9%
Coronary artery disease			71/41.3%
Diabetes mellitus			48/27.9%
Obesity			44/25.6%
Stroke history	(-)		139/80.8%
	(+)		33/19.2%
Admission aspect		5.0-10.0	9.1±1.1
Warfarin use	(-)		162/94.2%
	(+)		10/5.8%

Table 2. Continued			
		Min.-Max.	**Mean ± SD/n%**
NOAC use	(-)		153/89.0%
	(+)		19/11.0%
IV tPA	(-)		84/48.8%
	(+)		88/51.2%
Symptom puncture time		15.0-795.0	250.6±174.3
Symptom recanalisation time		75.0-815.0	308.3±181.2
Puncture recanalization time		10.0-190.0	56.8±24.0
The door imaging time		2.0-120.0	19.0±15.8
Door tPA time		20.0-120.0	51.2±20.9
Door puncture time		15.0-110.0	56.5±17.5
NIHSS score			
Admission		3.0-22.0	9.7±3.5
24_{th} hour		0.0-25.0	6.2±4.9
3_{rd} month mRS score		0.0-6.0	2.4±2.0
Glucose		76.0-608.0	148.8±75.7
sICH in 24 hours		(-)	125/72.7%
(+)		47/27.3%	

SD: Standard deviation, ASPECT: Alberta Stroke Program Early Computed Tomography Score (CT-0 score), NOAC: new oral anticoagulant, IV tPA: Intravenous tissue plasminogen activator, mRS: Modified rankin scale, NIHSS: National institutes of health stroke scale, sICH: Symptomatic intracranial hemorrhage

Symptom puncture (p=0.000), symptom recanalization (p=0.000), and puncture recanalization (p=0.021) time were significantly higher in the thrombectomy group than in the tPA plus thrombectomy group. In our study, the use of stent retrievers did not differ significantly between the IV tPA plus thrombectomy and thrombectomy groups. The rate of symptomatic intracranial hemorrhage (sICH) in the thrombectomy group at admission 24 hours was significantly (p=0.004) higher than that in the tPA plus thrombectomy group, and there was no superiority between the two groups in the 3rd month mRS scores (0.971^{m}) (Table 3).

The smoking rate was significantly lower (p=0.045) in the group with sICH than in the group without sICH. History of stroke was significantly (p=0.039) higher in the group with sICH than in the group without sICH. The IV door-IV tPA rate was significantly (p=0.004) lower in the sICH group than in the non-sICH group. Admission (p=0.033), 24th hour (p=0.001) the NIHSS score was significantly (p<0.05) higher in the group with sICH than in the group without sICH. The 3rd month mRS score was significantly (p=0.003) higher in the sICH group than in the non-sICH group. The distal embolism rate was significantly (p=0.009) higher in the sICH group than in the non-sICH group. The Tan score was significantly (p=0.007) lower in the sICH group than in the non-sICH group (Table 4).

Discussion

In this study, the effects of intravenous tissue plasminogen activator (IV tPA) administration on mechanical thrombectomy (MT) time, complications during the MT procedure, and intracerebral hemorrhage (ICH) in patients undergoing MT were analyzed. In a single-center retrospective study of 250 patients (105 with IV tPA plus MT and 145 with MT alone), symptom and door puncture times were found to be shorter in the MT group (11). It was found that IV tPA administration did not affect the thrombectomy door-puncture time; on the contrary, the puncture recanalization time was longer in the thrombectomy-only group than in the thrombectomy plus IV tPA group (p<0.05). In our study, first-pass recanalization, use of a stent retriever, and distal embolism rates did not differ significantly (p>0.05) between the IV tPA plus thrombectomy and thrombectomy groups. Therefore, administration of IV tPA may not be a disadvantage in terms of the initiation of MT therapy or development of procedural complications.

Several recent studies have evaluated the efficacy of MT in patients with acute stroke by comparing it with combined IV tPA plus MT therapy (11-14). A meta-analysis showed that combined IV tPA plus MT was associated with a higher probability of functional independence than MT alone. However, these results were derived from retrospective cohort studies in which MT alone was

Table 3. Comparison between thrombectomy and combined tPA plus thrombectomy				
Mean ± SD/n%		tPA plus thrombusectomy Mean ± SD/n%	Thrombectomy	p-value
Age		66.8±15.5	68.5±13.8	0.484[m]
Sex	Woman	49/55.7%	45/53.6%	0.781[X²]
	Man	39/44.3%	39/46.4%	
Cigarette smoking	(-)	49/55.7%	50/59.5%	0.610[X²]
	(+)	39/44.3%	34/40.5%	
Comorbid disease				
Hypertension		47/53.4%	53/63.1%	0.198[X²]
Atrial fibrillation		34/38.6%	38/45.2%	0.380[X²]
Coronary artery disease		39/44.3%	32/38.1%	0.407[X²]
Diabetes mellitus		25/28.4%	23/27.4%	0.881[X²]
Obesity		22/25.0%	22/26.2%	0.858[X²]
Stroke history	(-)	72/81.8%	67/79.8%	0.732[X²]
	(+)	16/18.2%	17/20.2%	
Admission aspect		9.3±1.1	9.0±1.1	0.127[m]
Symptom puncture time		156.4±96.3	311.1±186.4	**0.000**[m]
Symptom recanalisation time		207.1±97.6	373.3±192.7	**0.000**[m]
Puncture recanalization time		51.2±19.5	60.4±26.0	**0.021**[m]
The door imaging time		16.4±10.0	21.7±19.8	0.117[m]
Door tPA time		51.7±21.1	36.7±5.8	0.177[m]
Door puncture time		58.9±17.2	54.9±17.6	0.233[m]
NIHSS score				
Admission		9.1±3.1	10.3±3.8	**0.032**[m]
24th hour		5.8±4.7	6.6±5.0	0.246[m]
3rd month mRS score		2.40±2.06	2.32±1.96	0.971[m]

[m]Mann-Whitney U test, [X²]Chi-square test
Bold values: p<0.05
SD: Standard deviation, ASPECT: Alberta Stroke Program Early Computed Tomography Score (CT-0 score), NOAC: new oral anticoagulant, IV tPA: Intravenous tissue plasminogen activator, mRS: Modified Rankin Scale, NIHSS: National Institutes of Health Stroke Scale

Table 4. Comparison of patients with and without hemorrhage				
		sICH (-) Mean ± SD/n%	sICH (+) Mean ± SD/n%	p-value
Age		67.7±14.6	68.0±15.0	0.847[m]
Sex	Woman	63/50.4%	31/65.9%	0.094[X²]
	Man	62/49.6%	16/34.1%	
Cigarette smoking	(-)	66/52.8%	33/70.2%	**0.045**[X²]
	(+)	59/47.2%	14/29.8%	
Hypertension		77/61.6%	23/48.9%	0.155[X²]
Atrial fibrillation		51/40.8%	21/44.6%	0.569[X²]
Coronary artery disease		57/45.6%	14/29.7%	0.068[X²]
Diabetes mellitus		36/20.8%	12/25.5%	0.574[X²]
Obesity		33/26.4%	11/23.4%	0.516[X²]
Stroke history	(-)	106/84.8%	33/71.7%	**0.039**[X²]
	(+)	19/15.2%	14/28.3%	

Table 4. Continued					
			sICH (-)	sICH (+)	p-value
			Mean ± SD/n%	Mean ± SD/n%	
Admission aspect			9.2±1.0	8.9±1.3	0.134[m]
Warfarin use	(-)		119/95.2%	43/91.3%	0.842[X²]
	(+)		6/4.8%	4/8.7%	
NOAC use	(-)		113/91.1%	38/82.6%	0.117[X²]
	(+)		11/8.9%	8/17.4%	
IV tPA	(-)		52/42.4%	32/68.1%	**0.004**[X²]
	(+)		73/57.6%	15/32.9%	
Symptom puncture time			248.1±176.3	256.3±171.5	0.609[m]
Symptom recanalisation time			306.0±183.0	313.6±178.9	0.644[m]
Puncture recanalization time			57.6±24.3	54.9±23.7	0.398[m]
The door imaging time			17.6±13.5	22.9±20.5	0.127[m]
Door tPA time			50.4±19.3	53.8±28.3	0.961[m]
Door puncture time			56.3±18.0	57.1±16.4	0.877[m]
NIHSS score					
Admission			9.3±3.3	10.8±4.0	**0.033**[m]
24th hour			5.4±4.6	8.3±5.2	**0.001**[m]
3rd month mRS score			2.11±1.96	3.04±2.03	**0.003**[m]
Glucose			146.3±72.6	149.2±72.8	0.610[m]
HB			12.5±2.3	12.1±2.2	0.339[m]
PLT			250.5±113.0	244.0±79.4	0.798[m]
Stent retriever usage (+)		(-)	29/29.9%	10/24.4%	0.511[X²]
		68/70.1%	31/75.6%		
Distal embolism (+)		(-)	59/60.8%	15/36.6%	**0.009**[X²]
		38/39.2%	26/63.4%		
Antiaggregant therapy (+)		(-)	70/56.0%	27/57.4%	0.993[X²]
		55/44.0%	20/42.6%		
Tan score <50% 50-99% 100%		(-)	17/19.1%	18/45.0%	**0.007** [X²]
		25/28.1%	6/15.0%		
		31/34.8%	14/35.0%		
		16/18.0%	2/5.0%		

[m]Mann-Whitney U test, [X²]Chi-square test
Bold values: p<0.05
SD: Standard deviation, ASPECT: Alberta Stroke Program Early Computed Tomography Score (CT-0 score), NOAC: new oral anticoagulant, IV tPA: Intravenous tissue plasminogen activator, mRS: Modified Rankin Scale, NIHSS: National Institutes of Health Stroke Scale, sICH: Symptomatic intracranial hemorrhage

administered to many patients who were not eligible for IV tPA (15). Kaesmacher et al. (16) showed that MT alone was not superior to combined IVT-MT in terms of good clinical outcomes in a meta-analysis of patients admitted only during the IV tPA window. In our study, no superiority was found between the two groups in 3rd month mRS scores (p<0.05).

ICH is associated with high morbidity and mortality after MT (17). It is known that IV tPA administration alone increases the rate of ICH 3 to 10-fold compared to placebo (18,19). Yang et al. (20) showed that the symptomatic ICH rates were similar between patients treated with MT alone and combined therapy. The DEVT study revealed a higher incidence of ICH in patients receiving IV tPA (21). In a prospectively designed study, Quang Anh et al. (22) reported that combined treatment did not increase the risk of ICH. A 2021 meta-analysis found that combination therapy did not increase the risk of ICH (23). This study showed that the occurrence of ICH within 24 hours of symptom onset was significantly reduced in the combined group compared with the MT alone group (p<0.05). This may be explained by the fact

that the MT group's symptom-puncture and symptom recanalization times were higher than those of the tPA plus MT group ($p<0.05$). In our study, we also compared patients with and without ICH at 24 hours. In the group with ICH, admission and 24-hour NIHSS score were significantly ($p<0.05$) higher than the group without ICH. The 3rd month mRS score was significantly ($p<0.05$) higher in the ICH group than in the non-ICH group. These findings showed that ICH increased mortality and decreased functional independence, consistent with the literature. Further studies are needed to clarify the relationship between the incidence of ICH, factors causing ICH, and patient outcomes.

In support of the view that the technique used during the MT procedure causes variability in the risk of ICH, the use of a stent retriever has been shown to cause more damage to the vessel wall, especially the endothelium, than the direct aspiration technique (24). Contrary to this argument, Yıldırım (25) showed that the ICH rates were similar between the two techniques. In our study, no difference was observed in the rates of stent retrievers in patients with and without ICH. In our study, the rate of distal embolism was significantly ($p<0.05$) higher in the group with bleeding than in the group without bleeding. The fact that distal embolism is known to increase the volume of necrosis and cerebral edema and to increase the risk of hemorrhage is consistent with our study results.

Study Limitations

Our study's first limitation was that it was retrospective in nature. Another limitation was that the number of participants in the sample group was small due to the first year of activity at the stroke center. Different results can be obtained with larger study groups. All patients did not receive full-dose IV tPA. Patients who underwent TICI 3 recanalization during the MT procedure had their iv tPA treatment discontinued before completion. The IV tPA doses were not evaluated in these patients. Despite these limitations, 250 patients were included in our study, and their neuroimaging findings were examined in detail. We believe that the patient outcomes in our stroke center, which works intensively and is in close communication with other departments, will make an important contribution to the literature.

Conclusion

In patients undergoing MT, IV tPA administration did not increase ICH, or distal embolism, and there was no difference in terms of good clinical outcomes. ICH was found to be associated with increased mortality and poor clinical outcome. Further studies with larger patient groups are needed on this subject.

Footnote

Ethics Committee Approval: Ethical approval was obtained from the Clinical Research Ethics Committee of the University of Health Sciences Turkey, Istanbul Haseki Training and Research Hospital (dated: 27.12.2023, approval no.: 260-2023).

Informed Consent: Informed consent was obtained from all patients.

Authorship Contributions

Surgical and Medical Practies: Z.M., Concept: Z.M., I.K., Design: Z.M., I.K., Data Collection or Processing: Z.M., I.K., Analysis or Interpretation: Z.M., I.K., Literature Search: Z.M., I.K., Writing: Z.M., I.K.

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

Financial Disclosure: This study received no financial support.

References

1. National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. N Engl J Med. 1995;333:1581-7.
2. Hacke W, Kaste M, Bluhmki E, et al. Thrombolysis with alteplase 3 to 4,5 hours after acute ischemic stroke. N Engl J Med. 2008;359:1317-29.
3. Saver JL, Goyal M, Bonafe A, et al. Stent-Retriever Thrombectomy after Intravenous t-PA vs. t-PA Alone in Stroke. N Engl J Med. 2015;372:2285-95.
4. Yaghi S, Eisenberger A, Willey JZ. Symptomatic intracerebral hemorrhage in acute ischemic stroke after thrombolysis with intravenous recombinant tissue plasminogen activator: a review of natural history and treatment. JAMA Neurol. 2014;71:1181-5.
5. Wardlaw JM, Murray V, Berge E, del Zoppo GJ. Thrombolysis for acute ischaemic stroke. Cochrane Database Syst Rev. 2014:CD000213.
6. Berkhemer OA, Fransen PS, Beumer D, et al. A randomized trial of intraarterial treatment for acute ischemic stroke. N Engl J Med. 2015;372:11-20.
7. Powers WJ, Rabinstein AA, Ackerson T, et al. Guidelines for the Early Management of Patients With Acute Ischemic Stroke: 2019 Update to the 2018 Guidelines for the Early Management of Acute Ischemic Stroke: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. Stroke. 2019;50:344-418.
8. Barber PA, Demchuk AM, Zhang J, Buchan AM. Validity and reliability of a quantitative computed tomography score in predicting outcome of hyperacute stroke before thrombolytic therapy. ASPECTS Study Group. Alberta Stroke Programme Early CT Score. Lancet. 2000;335:1670-4.

9. Zhang S, Chen W, Tang H, et al. The prognostic value of a four-dimensional ct angiography-based collateral grading scale for reperfusion therapy in acute ischemic stroke patients. PLoS One. 2016;11:e0160502.
10. Zaidat OO, Castonguay AC, Linfante I, et al. First pass effect: A new measure for stroke thrombectomy devices. Stroke. 2018;49:660-6.
11. van Kranendonk KR, Kappelhof M, Bruggeman AAE, et al. Hemorrhage rates in patients with acute ischemic stroke treated with intravenous alteplase and thrombectomy versus thrombectomy alone. J Neurointerv Surg. 2023;15:262-9.
12. Hassan AE, Ringheanu VM, Preston L, Tekle W, Qureshi AI. IV tPA is associated with increase in rates of intracerebral hemorrhage and length of stay in patients with acute stroke treated with endovascular treatment within 4.5 hours: should we bypass IV tPA in large vessel occlusion? J Neurointerv Surg. 2021;13:114-8.
13. Rinkel LA, Treurniet KM, Kappelhof M, et al. Influence of time metrics on the treatment effect of intravenous alteplase prior to endovascular treatment in MR CLEAN-NO IV. J Neurointerv Surg. 2023;15:54-9.
14. Kobeissi H, Adusumilli G, Ghozy S, et al. Mechanical thrombectomy alone versus with thrombolysis for ischemic stroke: A meta-analysis of randomized trials. Interv Neuroradiol. 2023:15910199231154331.
15. Phan K, Dmytriw AA, Lloyd D, et al. Direct endovascular thrombectomy and bridging strategies for acute ischemic stroke: a network meta-analysis. J Neurointerv Surg. 2019;11:443-449.
16. Kaesmacher J, Mordasini P, Arnold M, et al. Direct mechanical thrombectomy in tPA-ineligible and -eligible patients versus the bridging approach: a meta-analysis. J Neurointerv Surg. 2019;11:20-7.
17. Fiorelli M, Bastianello S, von Kummer R, et al. Hemorrhagic transformation within 36 hours of a cerebral infarct: relationships with early clinical deterioration and 3-month outcome in the European Cooperative Acute Stroke Study I (ECASS I) cohort. Stroke. 1999;30:2280-4.
18. National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. N Engl J Med. 1995;333:1581-7.
19. Hacke W, Donnan G, Fieschi C, et al. Association of outcome with early stroke treatment: pooled analysis of ATLANTIS, ECASS, and NINDS rt-PA stroke trials. Lancet. 2004;363:768-74.
20. Yang P, Zhang Y, Zhang L, et al. Endovascular thrombectomy with or without intravenous alteplase in acute stroke. N Engl J Med. 2020;382:1981-93.
21. Zi W, Qiu Z, Li F, et al. Effect of Endovascular Treatment Alone vs Intravenous Alteplase Plus Endovascular Treatment on Functional Independence in Patients With Acute Ischemic Stroke: The DEVT Randomized Clinical Trial. JAMA. 2021;325:234-43.
22. Quang Anh N, Dang Luu V, Hoang Khoe L, et al. Effect of mechanical thrombectomy with vs. without intravenous thrombolysis in acute ischemic stroke. Clin Ter. 2022;173:257-64.
23. Du H, Lei H, Ambler G, et al. Intravenous thrombolysis before mechanical thrombectomy for acute ischemic stroke: A meta-analysis. J Am Heart Assoc. 2021;10:e022303.
24. Peschillo S, Diana F, Berge J, Missori P. A comparison of acute vascular damage caused by ADAPT versus a stent retriever device after thrombectomy in acute ischemic stroke: a histological and ultrastructural study in an animal model. J Neurointerv Surg. 2017;9:743-9.
25. Yıldırım S. Effect of intravenous thrombolytic treatment on prognosis and intracerebral hemorrhage in stroke patients treated with mechanical thrombectomy. Turk J Cereb Vasc Dis. 2023;29:77-83.

Original Article

DOI: 10.4274/haseki.galenos.2024.9993
Med Bull Haseki 2024;62:216-222

Factors Affecting the Complete Response in Breast and Axillary Regions Following Neoadjuvant Chemotherapy for Breast Cancer

Halit Ozgul*, Remzi Can Cakir**, Omer Celik*, Turan Can Yildiz*, Erhan Aydemir*, Betul Dagoglu Hark***, Azmi Lale****

*University of Health Sciences Turkey, Antalya Training and Research Hospital, Clinic of General Surgery, Antalya, Turkey
**Hatay Training and Research Hospital, Clinic of General Surgery, Hatay, Turkey
***Firat University Faculty of Medicine, Department of Biostatistics and Medical Informatics, Elazig, Turkey
****Antalya City Hospital, Clinic of Surgical Oncology, Antalya, Turkey

Abstract

Aim: Neoadjuvant chemotherapy (NAC) has transitioned from a treatment modality used solely for inoperable and locally advanced breast cancer to a therapeutic approach for early-stage breast cancer. High-risk patients, such as those with HER2-positive and triple-negative breast cancer, particularly benefit from NAC. This study aimed to evaluate the factors affecting pathological complete response (pCR) in primary breast tumors and axillary lymph nodes in patients with breast cancer.

Methods: The study included female patients with breast cancer who received NAC at a training and research hospital between 2020 and 2024. Patients were categorized based on age, tumor stage, and tumor biology: luminal A, HER2-positive luminal B, HER2-negative luminal B, HER2-positive alone, or triple-negative. The presence or absence of E-cadherin in tumor cells and Ki-67 levels were also examined. Data were obtained from medical records to assess the impact of these factors on complete response in patients with breast cancer and axillary metastatic lymph nodes following NAC.

Results: Univariate analysis revealed that histopathological subtypes, estrogen receptor and progesterone receptor (PR) status, HER2 status, perineural invasion, lymphovascular invasion (LVI), Ki-67 index, and carcinoma *in situ* (CIS) component significantly influenced pCR. Multivariate analysis confirmed that PR status [Odds ratio (OR): 3.33, 95% confidence interval (CI): 1.57-7.08, p=0.002], HER2 status (OR: 3.56, 95% CI: 1.71-7.44, p=0.001), LVI (OR: 3.91, 95% CI: 1.84-8.30, p<0.001), Ki-67 index (OR: 1.03, 95% CI: 1.01-1.05, p<0.001), and CIS component (OR: 7.01, 95% CI: 2.44-20.11, p<0.001) were independent predictors of complete response.

Conclusion: Our findings underscore the multifaceted nature of NAC response in breast cancer, which is influenced by histopathological and molecular characteristics.

Keywords: Neoadjuvant chemotherapy, pathological complete response, breast cancer, axillary lymph nodes

Introduction

Breast cancer is the most prevalent malignant tumor among women and a leading cause of cancer-related morbidity and mortality worldwide (1). One in eight women will develop breast cancer during their lifetimes (2). Neoadjuvant chemotherapy (NAC) is the standard therapeutic approach for patients with locally advanced breast cancer. It is particularly indicated in patients with large tumor volumes, the presence of lymph node metastases, HER2-positive breast cancer, or triple-negative breast cancer (TNBC) (3). The primary clinical benefit of NAC is reduced tumor staging, which increases the likelihood of breast-conserving surgery (4,5). Breast cancer is currently classified into distinct molecular subtypes based on the presence or absence of immunohistochemical markers.

Address for Correspondence: Halit Ozgul
University of Health Sciences Turkey, Antalya Training and Research Hospital, Clinic of General Surgery, Antalya, Turkey
E-mail: halitozgul38@gmail.com **ORCID:** orcid.org/0009-0006-6457-9738
Received: 19.07.2024 **Accepted:** 14.10.2024

These subtypes have different risk profiles and therapeutic strategies (6). Human epidermal growth factor receptor 2 (HER2 or HER2/neu) is a critical molecular target, and it is characterized by a transmembrane receptor with tyrosine kinase activity. HER2 is overexpressed or amplified in approximately 20% of breast cancers, contributing to aggressive tumor behavior and poor clinical outcomes (7).

We hypothesized that specific demographic, clinical, and molecular factors may predict the likelihood of achieving pathological complete response (pCR) in patients undergoing NAC for breast cancer. Post-NAC pathological response data reflect tumor chemosensitivity (8). Several studies have demonstrated specific correlations between morphological changes and the use of chemotherapeutic agents using pCR (9,10). Trastuzumab and pertuzumab, humanized monoclonal antibodies targeting different epitopes of HER2, have been reported to significantly improve pCR rates with dual anti-HER2 blockade in HER2-positive subtypes (11). However, previous studies on the predictive factors of NAC in breast cancer suggested that no single factor adequately predicted pCR (12).

Therefore, the current study aimed to evaluate the factors affecting pCR in primary breast tumors and axillary lymph nodes in patients with breast cancer receiving NAC. These predictive factors. Despite these factors, this study aimed to contribute to clinical decision-making and potentially improve treatment personalization. This may improve patient outcomes by optimizing NAC regimens and surgical planning, ultimately contributing to effective and personalized approaches.

Methods

Ethical Standards

This study received approval from the Institutional Review Board (IRB) of University of Health Sciences Turkey, Antalya Training and Research Hospital (IRB number: 5/15) on April 25, 2024. Informed consent was obtained from all participants.

Patient Selection

This cross-sectional observational study included 222 female patients with breast cancer who received NAC at a training and research hospital between 2020 and 2024. Patients were categorized into five groups based on age, tumor stage, and tumor biology: luminal A, HER2-positive luminal B, HER2-negative luminal B, HER2-positive alone, and triple-negative. The presence or absence of E-cadherin in tumor cells and Ki-67 levels were analyzed. Data were obtained from medical records. The effects of these factors on complete response in malignant breast masses and metastatic axillary lymph nodes following NAC were evaluated. Patients exhibiting complete response in both regions were considered to have achieved pCR (Figure 1).

Neoadjuvant chemotherapy was performed based on the molecular subtypes and biomarkers of breast cancer. For patients with TNBC, we selected chemotherapy regimens that included platinum-based therapies. Specifically, the most frequently used regimens were cisplatin combined with gemcitabine or carboplatin. In cases in which the tumor expressed PD-L1 positivity or exhibited high microsatellite instability (MSI), pembrolizumab was added to these regimens, forming a triple-combination therapy. If PD-L1 or MSI was not detected, a dual regimen was used without immunotherapy. For HER2-positive tumors, treatment was consistent with the current guidelines, involving a three-drug regimen of pertuzumab, trastuzumab, and docetaxel.

Patients with hormone receptor-positive (HR+) and HER2-negative tumors were divided into luminal A and luminal B subtypes based on Ki-67 expression levels. For luminal B tumors (Ki-67 >14%), a more aggressive approach was used, typically involving four cycles of dose-dense doxorubicin and cyclophosphamide (AC), followed by four cycles of dose-dense paclitaxel. This regimen was selected based on the updated ESMO guidelines and was particularly beneficial for younger patients and those with high-risk clinical features.

Exclusion Criteria

Patients who did not complete NAC, had incomplete records, and had distant metastasis were excluded from the study.

Statistical Analysis

All analyses were performed using the IBM SPSS Statistics Version 22.0 statistical software package. Categorical variables were expressed as numbers

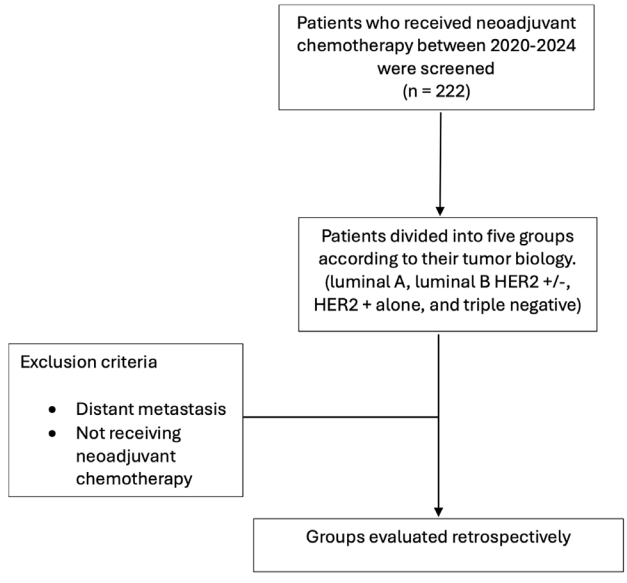


Figure 1. Flow chart of patient selection

and percentages, whereas continuous variables were summarized as mean and standard deviation or median (minimum-maximum) values, where appropriate. The normality of data distribution for continuous variables was confirmed using the Kolmogorov-Smirnov test. For the comparison of continuous variables between the two groups, Student's t-test or the Mann-Whitney U test was used, depending on whether the statistical hypotheses were fulfilled. Univariate and multiple logistic regression analyses were performed to identify factors predictive of complete response following NAC. Each variable was modeled as a univariate without considering other variables, and the common effect was revealed by multiple logistic regression. Odds ratios (ORs) and confidence intervals (CIs) were reported for variables with statistically significant effects on response.

Results

Table 1 presents the descriptive statistics of the 222 patients included in the study according to their demographic and histopathologic characteristics. The median age of the patients was 53.4 years (range, 29-85 years), and the mean body mass index (BMI) was 26.03±1.68. Among the included patients, 201 (90.5%) had ductal carcinoma histopathology. Estrogen receptor (ER) positivity was observed in 164 patients (73.9%), progesterone receptor (PR) positivity in 139 (62.6%), and HER2 positivity in 79 (35.6%). A total of 140 patients (63.1%) were negative for E-cadherin, 198 patients (89.2%) were negative for perineural invasion (PNI), and 121 patients (54.5%) were negative for lymphovascular invasion (LVI). The mean Ki-67 index was calculated at 35.76±25.68. Regarding the pathological response to NAC, 21 patients (9.5%) exhibited no response, 115 patients (51.8%) had a partial response, and 86 patients (38.7%) achieved a complete response.

The descriptive statistics and p-values related to variables affecting pCR following NAC are presented in Table 2. Accordingly, the mean age and BMI distributions were comparable between patients with partial or no response and those with complete response (p=0.564 and 0.725, respectively). Concerning histopathological subtypes, mixed carcinoma (n=4, 2.90%), mucinous carcinoma (n=7, 5.10%), and micropapillary carcinoma (n=4 cases, 2.90%) were more common in patients with partial or no response compared with those with a complete response (p=0.010). Among patients with partial or no response, 117 (86.0%) were ER-positive and 107 (78.70%) were PR-positive. Among patients with a complete response, 47 (54.70%) were ER-positive and 32 (37.20%) were PR-positive, indicating a significant effect of the ER and PR status on the pathological response (p<0.001 for both). Furthermore, the number of PNI-positive patients was

Table 1. Demographic and histopathological characteristics of the patients

		n (%)
Age (years)		53.4 (29-85)
BMI (kg/m^2)		26.03±1.68
Localization	Right	115 (51.8)
	Left	107 (48.2)
Histopathology	Ductal carcinoma	201 (90.5)
	Lobular carcinoma	4 (1.8)
	Medullary	1 (0.5)
	Mix carcinoma	4 (1.8)
	Mucinous carcinoma	7 (3.2)
	Micropapillary	4 (0.5)
ER	Negative	58 (26.1)
	Positive	164 (73.9)
PR	Negative	83 (37.4)
	Positive	139 (62.6)
HER2	Negative	143 (64.4)
	Positive	79 (35.6)
E-cadherin	Negative	140 (63.1)
	Positive	82 (36.9)
PNI	Negative	198 (89.2)
	Positive	24 (10.8)
LVI	Negative	121 (54.5)
	Positive	101 (45.5)
Ki-67 index (%)		35.76
CIS component	Negative	166 (74.8)
	Positive	56 (25.2)
Molecular classification	Luminal A	43 (19.4)
	Luminal B (Ki-67+)	70 (31.5)
	Luminal B (HER2+)	51 (23.0)
	Tripple-	32 (14.4)
	HER2	26 (11.7)
Neoadjuvant chemotherapy protocol	PTD	52 (23.4)
	PT	26 (11.7)
	AC	103 (46.4)
	CISG	27 (12.2)
	DC	7 (3.2)
	DAC	2 (0.9)
	CARPAK	5 (2.3)
cN stage	cN0	4 (1.8)
	cN1	130 (58.6)
	cN2	88 (39.6)
Menopause	No	84 (37.8)
	Yes	138 (62.2)
Pathological response to NAC	No response	21 (9.5)
	Partial	115 (51.8)
	Complete	86 (38.7)

PNI: Perineural invasion, LVI: Lymphovascular invasion, CIS: Carcinoma *in situ*, NAC: Neoadjuvant chemotherapy

Table 2. Predictors affecting pathological response to NAC

		Response to NAC		p value
		Partial and no response n (%)	Complete response n (%)	
Age		53.2±11.1	53.8±11.3	0.564
BMI		26.0±1.7	26.1±1.6	0.725
Localization	Right	65 (47.80)	50 (58.10)	0.168
	Left	71 (52.20)	36 (41.90)	
Histopathology	Ductal carcinoma	116 (85.30)	85 (98.80)	**0.010**
	Lobular carcinoma	4 (2.90)	0	
	Medullary	0	1 (1.2)	
	Mix carcinoma	4 (2.90)	0	
	Mucinous carcinoma	7 (5.10)	0	
	Micropapillary	4 (2.90)	0	
ER	Negative	19 (14.0)	39 (45.30)	**<0.001**
	Positive	117 (86.0)	47 (54.70)	
PR	Negative	29 (21.30)	54 (62.80)	**<0.001**
	Positive	107 (78.70)	32 (37.20)	
HER2	Negative	103 (75.70)	40 (46.50)	**<0.001**
	Positive	33 (24.30)	46 (53.50)	
E-Cadherin	Negative	65 (47.80)	50 (58.10)	0.345
	Positive	71 (52.20)	36 (41.90)	
PNI	Negative	114 (83.80)	84 (97.70)	**0.001**
	Positive	22 (16.20)	2 (2.30)	
LVI	Negative	61 (44.90)	60 (69.80)	**<0.001**
	Positive	71 (55.10)	26 (30.20)	
Ki-67 index		26.9±21.8	49.8±25.2	**<0.001**
CIS component	Negative	88 (64.70)	78 (90.70)	**<0.001**
	Positive	48 (35.30)	8 (9.30)	
Molecular classification	Luminal A	41 (30.10)	2 (2.30)	**<0.001**
	Luminal B (Ki-67+)	51 (37.50)	19 (22.10)	
	Luminal B (HER2+)	25 (18.40)	26 (30.20)	
	Triple-	13 (9.60)	19 (22.10)	
	HER2	6 (4.40)	20 (23.30)	
NAC protocol	PTD	26 (19.10)	26 (30.2)	**0.005**
	PT	6 (4.40)	20 (23.30)	
	AC	82 (60.30)	21 (24.40)	
	CISG	10 (7.40)	17 (19.80)	
	DC	7 (5.10)	0	
	DAC	2 (1.50)	0	
	CARPAK	3 (1.40)	2 (2.30)	
cN stage	cN0	1 (0.70)	3 (3.50)	**0.684**
	cN1	81 (59.60)	49 (57.0)	
	cN2	54 (39.70)	34 (39.50)	
Menopause	No	52 (38.20)	32 (37.20)	**0.888**
	Yes	84 (61.80)	36 (62.80)	

Bold values indicate statistical significance at p<0.05
Statistics: χ^2 or Fisher's exact test for categorical variables and Student's t-test or Mann-Whitney U test for continuous variables
NAC: Neoadjuvant chemotherapy, BMI: Body mass index, ER: estrogen receptor, PR: Progesterone receptor, PNI: Perineural invasion, LVI: Lymphovascular invasion, CIS: Carcinoma in situ

higher in the partial or no response group (n=22, 16.20%) than in the complete response group (n=2, 2.30%) (p=0.001). Similarly, LVI positivity was more common in the partial or no response group (n=71, 55.10%) than in the complete response group (n=26, 30.20%) (p<0.001). The mean Ki-67 index was significantly higher in patients with complete response (49.8±25.2) than in those with partial or no response (p<0.001). There was a higher number of patients with luminal A and luminal B (Ki-67+) subtypes in the partial or no response group, whereas luminal B (HER2+), triple-negative, and HER2-enriched subtypes were more prevalent in the complete response group (p<0.001). Upon evaluating the NAC protocols used, AC (n=82, 60.30%), DC (n=7, 5.10%), and DAC (n=2, 1.50%) were more common in the partial or no response group than in the complete response group (p=0.005).

Table 3 presents the results of univariate and multivariate regression analyses conducted to identify predictive factors for resistance to pCR following NAC. HER2-negative patients were 3.59 (95% CI: 2.02-6.39) times more likely to achieve a complete response than HER2-positive patients. The ORs for complete response in PNI-positive patients versus PNI-negative patients were 8.11 (95% CI: 1.86-35.42) and 2.84 (95% CI: 1.60-5.02), respectively. The OR for a complete response in patients with a positive carcinoma *in situ* (CIS) component compared with those without this component was 3.51 (95% CI: 2.37-11.93). Multivariate logistic regression revealed that PR status, HER2 status, LVI, Ki-67 index, and CIS component were significant predictors, with the positive CIS component having the highest effect on complete response (OR: 7.01, 95% CI: 2.44-20.11).

Discussion

This study evaluated the demographic and histopathological characteristics of 222 patients with breast cancer who received NAC and identified factors affecting pCR. The findings provide critical insights into the predictors of response, potentially guiding personalized treatment strategies and improving outcomes for patients with breast cancer. The mean ages and BMIs of our cohort were 53.4 years and 26.03 kg/m², respectively, with no significant differences observed between patients with partial or no response and those with complete response. This is consistent with previous studies showing that age and BMI are not primary determinants of NAC response in breast cancer, emphasizing the greater importance of other biological factors.

A significant proportion of our patients (90.5%) had ductal carcinoma, consistent with the overall prevalence of breast cancer. Our results indicated a higher probability of pCR in patients with ductal carcinoma than in other histopathological subtypes (p=0.010), supporting the

Table 3. Univariate and multivariate analyses of predictors of complete response following neoadjuvant chemotherapy

	Univariate analyses			Multivariate analyses		
	OR	**95% CI**	**p value**	**OR**	**95% CI**	**p value**
ER	5.11	2.68-9.73	**<0.001**			
PR	6.23	3.42-11.34	**<0.001**	3.33	1.57-7.08	0.002
HER2	3.59	2.02-6.39	**<0.001**	3.56	1.71-7.44	**0.001**
PNI	8.11	1.86-35.42	**0.005**			
LVI	2.84	1.60-5.02	**<0.001**	3.91	1.84-8.30	**<0.001**
Ki67 (%)	1.039	1.026-1.052	**<0.001**	1.03	1.01-1.05	**<0.001**
CIS component	3.51	2.37-11.93	**<0.001**	7.01	2.44-20.11	**<0.001**
Molecular Classification						
HER2+	-	-				
Luminal A	68.33	12.64-369.31	**<0.001**			
Luminal B (HER2-)	29.96	6.14-146.19	**<0.001**			
Luminal B (HER2+)	21.32	4.66-97.65	**<0.001**			
Triple-	7.64	1.68-34.71	**0.008**			

Bold values indicate statistical significance at p<0.05
OR: Odds ratio, CI: Confidence interval, ER: Estrogen receptor, PR: Progesterone receptor, PNI: Perineural invasion, LVI: lymphovascular invasion, CIS: Carcinoma in situ
Statistics: Logistic regression analyses

existing evidence that ductal carcinoma may be more sensitive to NAC than lobular and mucinous carcinomas, which are often associated with weaker responses. Predicting pCR contributes to evaluating the benefits of NAC in patients with newly diagnosed breast cancer and assisting in selecting the optimal surgical approach preoperatively. However, there is no consensus on the imaging-based assessment of pCR following NAC, and it is not possible to reliably predict pCR (13,14). Over time, surgical trends have shifted toward the implementation of less-invasive procedures that minimize long-term morbidity without compromising oncological safety. The benefits of neoadjuvant therapies in facilitating breast conservation are well established (15).

In a study by Dou et al. (16), age, T stage, N stage, ER status, PR status, HER-2 status, Ki-67, histological grade, molecular subtype, clinical stage, and pathology type were strongly associated with pCR rates (p<0.05). However, no significant correlation was observed between pCR and chemotherapy regimen, surgical method, menopausal status, BMI, or lymphatic infiltration (p>0.05). A younger age, lower T and N stages, ER negativity, PR negativity, HER2-positivity, high Ki-67 expression, and lower histological grades were reported to be more likely to achieve pCR (16). Another study observed a higher frequency of pCR in patients with right breast cancer, with 63.5% of pCR occurring in the right breast and 36.5% occurring in the left breast (p=0.012) (17). It has also been suggested that the pre-NAC Ki-67 index reflects tumor cell proliferative capacity and is closely related to NAC sensitivity (18), thus being consistently recognized as an independent predictor of NAC response (19,20). Concerning HER2, the efficacy of NAC in HER2-positive patients has been significantly improved through the use of trastuzumab and/or pertuzumab. Consequently, HER2 status has emerged as an independent predictor of the efficacy of NAC (21,22). Another study highlighted that women with more advanced cancer stages and a Ki-67 index >20% were more likely to achieve pCR (23).

In a study conducted by Yan et al. (24), the group with a tumor size of 2 cm following NAC exhibited the highest rate of pCR. Patients with luminal A subtype had the lowest pCR rate, whereas those with TNBC had the highest pCR rate. Furthermore, the HER2-positive subtype showed a higher pCR percentage than the luminal B subtype (24). Qian et al. (25) categorized 325 patients into two groups based on whether they achieved pCR. Within this cohort, 126 patients achieved pCR (a rate of 38.8%). Overall, compared with the non-pCR group, patients in the pCR group had several significant characteristics: older age, smaller tumor size, lower stage, a higher Ki-67 index, a higher proportion of HER2-positive tumors, and a lower percentage of HR+ tumors (p<0.05) (25). Hormone receptor-positive breast cancer exhibits a better prognosis than HER2-positive breast cancer or TNBC. In contrast, HER2-positive breast cancer and TNBC exhibit better therapeutic response to chemotherapy. However, only a few studies have evaluated the oncological outcomes of NAC in patients with locally advanced HR-negative breast cancer (26,27).

In a study by Lan et al. (28), univariate analysis of predictive factors between the pCR and non-pCR groups revealed statistically significant differences in cT, cN, ER, PR, and Ki-67 status (p<0.05). However, there were no statistically significant differences between the two groups in terms of age, menopausal status, HER2 status,

or chemotherapy cycles (28). In another study, Guan et al. (29) reported that 57 patients (14.8%) achieved breast pCR. Univariate analysis indicated that tumor size, ER, PR, and Ki-67 were associated with breast pCR. Additionally, multivariate analysis identified tumor size, PR, and Ki-67 as statistically significant factors. Dou et al. (30) found that hormone receptor status was an independent predictor of the pCR rate in patients with breast cancer who received NAC. The authors reported that the ER+/PR and ER/PR phenotypes were more responsive to chemotherapy than the ER+/PR+ phenotypes (30).

In our study, the percentages of ER and PR positivity were significantly lower in patients achieving pCR (ER: 54.7% vs. 86.0%, $p<0.001$; PR: 37.2% vs. 78.7%, $p<0.001$), underscoring the known association between HR positivity and reduced chemotherapy sensitivity. HER2 positivity was also a significant determinant, with a higher prevalence in patients achieving pCR (HER2+: 34.9% vs. 38.5%, $p<0.001$). HER2-positive tumors, typically more aggressive, respond well to chemotherapy combined with HER2-targeted therapies, highlighting the importance of incorporating HER2-targeted agents into NAC regimens.

In this study, PNI and LVI were significantly associated with weaker NAC responses (PNI: $p=0.001$, LVI: $p<0.001$), indicating more aggressive tumor biology and higher metastatic potential, thus explaining the reduced chemotherapy efficacy. The Ki-67 proliferation index emerged as a critical determinant, with higher values significantly associated with pCR (49.8% vs. 26.9%, $p<0.001$). Ki-67 serves as a marker of cellular proliferation, with higher indices reflecting a larger fraction of actively dividing cells that are more susceptible to chemotherapy. The molecular classification revealed significant differences in pCR rates. Luminal A tumors, with an OR of 68.33 for treatment resistance, were most resistant to NAC ($p<0.001$). In contrast, HER2-positive and triple-negative subtypes exhibited better responses, reflecting their aggressive nature but higher chemosensitivity.

Lastly, multivariate analysis identified the presence of CIS, PR, HER2, LVI, and Ki-67 index as significant resistance determinants. In particular, CIS had the highest effect on treatment resistance (OR: 7.01, $p<0.001$), highlighting the complexity of treating these tumors.

Study Limitations

This study has several limitations. Its retrospective design may introduce selection bias, limiting the generalizability of the findings. Additionally, the single-center nature of the study may not fully capture the broader variability in treatment protocols and patient populations observed in multicenter studies. The relatively small sample size, particularly within certain molecular subtypes, may restrict the statistical power to detect subtle differences in response predictors. The reliance on pathologic assessment of pCR without standardized imaging protocols could have impacted the accuracy of response evaluation. Despite these limitations, this study provides valuable insights into the factors influencing pCR in breast cancer patients receiving NAC. The findings contribute to the growing body of evidence that can improve personalized treatment planning and optimize outcomes for breast cancer patients.

Conclusion

This study highlighted the multifaceted nature of NAC response in breast cancer, which is driven by histopathological and molecular characteristics. These findings demonstrate the need for personalized therapeutic approaches based on individual tumor biology to enhance treatment efficacy and patient outcomes. Further research should explore these predictive factors in larger, more diverse cohorts to validate and improve these insights.

Footnote

Ethics Committee Approval: This study received approval from the Institutional Review Board (IRB) of University of Health Sciences Turkey, Antalya Training and Research Hospital (IRB number: 5/15) on April 25, 2024.

Informed Consent: Informed consent was obtained from all participants.

Authorship Contributions

Surgical and Medical Practices: H.O., R.C.C., O.C., T.C.Y., E.A., B.D.H., A.L., Concept: H.O., Design: H.O., O.C., Data Collection or Processing: O.C., T.C.Y., E.A., Analysis or Interpretation: B.D.H., A.L., Literature Search: H.O., R.C.C., T.C.Y., Writing: H.O., R.C.C.

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

Financial Disclosure: This study received no financial support.

References

1 Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. 2020;70:7-30.

2. The Global Cancer Observatory. All Rights Reserved, March 2021.

3. Li Q, Liu J, Jiang Z, Liu Q. CSCO breast cancer guideline: precise, economical and oriental. Sci China Life Sci. 2020;63:1410-2.

4. Mistry KA, Thakur MH, Kembhavi SA. The effect of chemotherapy on the mammographic appearance of breast cancer and correlation with histopathology. Br J Radiol. 2016;89:20150479.

5. Tate SC, Andre V, Enas N, Ribba B, Gueorguieva I. Early change in tumour size predicts overall survival in patients with first-line metastatic breast cancer. Eur J Cancer. 2016;66:95-103.

6. Waks AG, Winer EP. Breast Cancer Treatment: A Review. JAMA. 2019;321:288-300.
7. Piccart-Gebhart MJ, Procter M, Leyland-Jones B, Goldhirsch A, Untch M, Smith I, et al. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. N Engl J Med. 2005;353:1659-72.
8. Mieog JS, van der Hage JA, van de Velde CJ. Preoperative chemotherapy for women with operable breast cancer. Cochrane Database Syst Rev. 2007;2007:CD005002.
9. Fukada I, Araki K, Kobayashi K, Shibayama T, Takahashi S, Gomi N, et al. Pattern of Tumor Shrinkage during Neoadjuvant Chemotherapy Is Associated with Prognosis in Low-Grade Luminal Early Breast Cancer. Radiology. 2018;286:49-57.
10. Xie F, Chen R, Zhang L, Yin Z, Zhu Q, You S, et al. Efficacy of two-weekly nanoparticle albumin-bound paclitaxel as neoadjuvant chemotherapy for breast cancer. Nanomedicine (Lond). 2019;14:1595-603.
11. Swain SM, Baselga J, Kim SB, Ro J, Semiglazov V, Campone M, et al. Pertuzumab, trastuzumab, and docetaxel in HER2-positive metastatic breast cancer. N Engl J Med. 2015;372:724-34.
12. Zhang F, Zheng W, Ying L, Wu J, Wu S, Ma S, et al. A Nomogram to Predict Brain Metastases of Resected Non-Small Cell Lung Cancer Patients. Ann Surg Oncol. 2016;23:3033-9.
13. Fowler AM, Mankoff DA, Joe BN. Imaging Neoadjuvant Therapy Response in Breast Cancer. Radiology. 2017;285:358-75.
14. Croshaw R, Shapiro-Wright H, Svensson E, Erb K, Julian T. Accuracy of clinical examination, digital mammogram, ultrasound, and MRI in determining postneoadjuvant pathologic tumor response in operable breast cancer patients. Ann Surg Oncol. 2011;18:3160-3.
15. Pilewskie M, Morrow M. Axillary Nodal Management Following Neoadjuvant Chemotherapy: A Review. JAMA Oncol. 2017;3:549-55.
16. Dou H, Jia S, Ba Y, Luo D, Yu P, Li F, et al. Clinical characteristics and pathologic complete response (pCR) rate after neoadjuvant chemotherapy in postpartum women with breast cancer. J Cancer Res Clin Oncol. 2023;149:14185-204.
17. Orrù S, Pascariello E, Pes B, Rallo V, Barbara R, Muntoni M, et al. Biomarker dynamics affecting neoadjuvant therapy response and outcome of HER2-positive breast cancer subtype. Sci Rep. 2023;13:12869.
18. Ellis MJ, Suman VJ, Hoog J, Goncalves R, Sanati S, Creighton CJ, et al. Ki67 Proliferation Index as a Tool for Chemotherapy Decisions During and After Neoadjuvant Aromatase Inhibitor Treatment of Breast Cancer: Results From the American College of Surgeons Oncology Group Z1031 Trial (Alliance). J Clin Oncol. 2017;35:1061-9.
19. Zhang J, Gao S, Zheng Q, Kang Y, Li J, Zhang S, et al. A Novel Model Incorporating Tumor Stiffness, Blood Flow Characteristics, and Ki-67 Expression to Predict Responses After Neoadjuvant Chemotherapy in Breast Cancer. Front Oncol. 2020;10:603574.
20. Ma Y, Zhang S, Zang L, Li J, Li J, Kang Y, et al. Combination of shear wave elastography and Ki-67 index as a novel predictive modality for the pathological response to neoadjuvant chemotherapy in patients with invasive breast cancer. Eur J Cancer. 2016;69:86-101.
21. Hamy AS, Belin L, Bonsang-Kitzis H, Paquet C, Pierga JY, Lerebours F, et al. Pathological complete response and prognosis after neoadjuvant chemotherapy for HER2-positive breast cancers before and after trastuzumab era: results from a real-life cohort. Br J Cancer. 2016;114:44-52.
22. Oh DY, Bang YJ. HER2-targeted therapies - a role beyond breast cancer. Nat Rev Clin Oncol. 2020;17:33-48.
23. Saffie Vega I, Sapunar Zenteno J, Buscaglia Fernandez F, Reyes Cosmelli F, Lagos Chavez R, Chahuán Manzur B. Predictors of pathologic complete response to neoadjuvant treatment in HER2-overexpressing breast cancer: a retrospective analysis using real-world data. Ecancermedicalscience. 2022;16:1338.
24. Yan S, Wang W, Zhu B, Pan X, Wu X, Tao W. Construction of Nomograms for Predicting Pathological Complete Response and Tumor Shrinkage Size in Breast Cancer. Cancer Manag Res. 2020;12:8313-23.
25. Qian B, Yang J, Zhou J, Hu L, Zhang S, Ren M, et al. Individualized model for predicting pathological complete response to neoadjuvant chemotherapy in patients with breast cancer: A multicenter study. Front Endocrinol (Lausanne). 2022;13:955250.
26. Uray IP, Brown PH. Chemoprevention of hormone receptor-negative breast cancer: new approaches needed. Recent Results Cancer Res. 2011;188:147-62.
27. Silva LCFF, Arruda LSM, David Filho WJ, Cruz FJSM, Trufelli DC, Del Giglio A. Hormone receptor-negative as a predictive factor for pathologic complete response to neoadjuvant therapy in breast cancer. Einstein (Sao Paulo). 2019;17:eAO3434.
28. Lan A, Chen J, Li C, Jin Y, Wu Y, Dai Y, et al. Development and Assessment of a Novel Core Biopsy-Based Prediction Model for Pathological Complete Response to Neoadjuvant Chemotherapy in Women with Breast Cancer. Int J Environ Res Public Health. 2023;20:1617.
29. Guan D, Jie Q, Wu Y, Xu Y, Hong W, Meng X. Real-world data on breast pathologic complete response and disease-free survival after neoadjuvant chemotherapy for hormone receptor-positive, human epidermal growth factor receptor-2-negative breast cancer: a multicenter, retrospective study in China. World J Surg Oncol. 2022;20:326.
30. Dou H, Li F, Wang Y, et al. Estrogen receptor as a predictive factor for tumor pathologic complete response after neoadjuvant chemotherapy in operable breast cancer. Oncol Lett. 2021;21:30.

Original Article

DOI: 10.4274/haseki.galenos.2024.9854
Med Bull Haseki 2024;62:223-228

Effect of Sarcopenia on Clinical Outcomes of Patients with Hairy Cell Leukemia

Memis Hilmi Atay*, Taner Tunc**, Oznur Aydin*, Fatih Uzunkaya***

*Ondokuz Mayıs University Faculty of Medicine, Department of Internal Medicine, Samsun, Turkey
**Ondokuz Mayıs University Faculty of Arts and Sciences, Department of Statistics, Samsun, Turkey
***Ondokuz Mayıs University Faculty of Medicine, Department of Radiology, Samsun, Turkey

Abstract

Aim: Sarcopenia may develop in patients with hairy cell leukemia (HCL). There is no study in the literature showing the prognostic importance of sarcopenia in patients with HCL. In this study, the effect of pretreatment sarcopenia on clinical outcomes in patients with HCL was investigated.

Methods: This study included 34 patients with pre-treatment abdominal computed tomography (CT) images who received cladribine (purine nucleoside analog) treatment between April 2006 and April 2022 at Ondokuz Mayıs University Hospital. To ensure measurement standardization, measurements were performed using abdominal CT sections showing the L3 vertebra. The optimal cut-off value for the skeletal mass index to be used for the prediction of sarcopenia was determined by receiver operating characteristic analysis. Patients were divided into two groups according to whether they were sarcopenic or not, and their clinical results were compared.

Results: Overall survival (OS) tended to be shorter in the sarcopenic group than in the non-sarcopenic group (p=0.046). Progression-free survival was significantly better in the non-sarcopenic group than in the sarcopenic group (p=0.009). In the multivariate analysis, sarcopenia (hazard ratio=0.154, p=0.043) was an effective variable for OS.

Conclusion: Sarcopenia is a prognostic factor for prognosis and treatment parameters in patients with HCL.

Keywords: Hairy cell leukemia, computed tomography, skeletal mass index, sarcopenia, prognostic factor

Introduction

Hairy cell (HC) leukemia (HCL) is a rare, mature B cell-derived chronic leukemia characterized by splenomegaly, pancytopenia, and peripheral HC (1-3). The annual incidence is 0.3 cases per 100,000 (4). The frequency is 4-5 times higher in men than in women. It is commonly observed between 55 and 60 years of age (5). Treatment for HCL initially uses purine analogs (cladribine or pentostatin). The full response rate during the first 5 years is 85-90% (6). At the 5-year follow-up, relapse was observed in 58% of patients who responded to treatment (7).

Sarcopenia is a progressive, generalized syndrome characterized by the loss of skeletal muscle mass and strength (8). The metabolic activity and systemic inflammation of cancer cells cause muscle loss, and sarcopenia develops (9). Computed tomography (CT) is a potential imaging biomarker for predicting survival outcomes in clinical practice because of its ability to provide objective quantitative and qualitative measurements of skeletal muscle and fat tissue. A meta-analysis showed the prognostic importance of sarcopenia in patients with hematological malignancies (10). Recent studies have demonstrated that early diagnosis of sarcopenia in patients with hematological malignancies can reverse the process of muscle loss and prevent the negative effects of sarcopenia syndrome on patient progression (11).

There is no study in the literature showing the prognostic importance of sarcopenia in patients with HCL. In this study, the effect of pretreatment sarcopenia on clinical outcomes in HCL patients receiving cladribine was investigated.

Address for Correspondence: Memis Hilmi Atay, Ondokuz Mayıs University Faculty of Medicine, Department of Internal Medicine, Samsun, Turkey
Phone: +90 530 468 15 59 **E-mail:** draatay@gmail.com **ORCID:** orcid.org/0000-0001-9666-9655
Received:14.03.2024 **Accepted:** 10.09.2024

Methods

Study Design and Compliance with Ethical Standards

This retrospective study included 50 patients aged >18 years attending Ondokuz Mayıs University Faculty of Medicine, Clinic of Hematology from April 2006 to April 2022 with HCL diagnosis. The study included 34 patients with pre-treatment abdominal CT images who received cladribine (purine nucleoside analog) treatment. Five patients receiving other treatments and 11 patients without pre-treatment abdominal CT imaging were excluded from the study (Figure 1). The retrospective study was approved by the Ondokuz Mayıs University Clinical Research Ethics Committee (approval no.: OMÜ KAEK 2022/576, date: December 28, 2022).

All demographic data were obtained from patient files and electronic medical records. Body mass index (BMI) was calculated using the formula kg/m^2 by measuring pre-treatment weight and height.

Patients received a 24-hour continuous cladribine infusion (0.1 mg/kg) for 7 days. Response to treatment was assessed using bone marrow biopsy and regression of splenomegaly based on physical examination in the third month. In peripheral blood (without transfusion), hemoglobin 11 g/dL, platelets 100,000/mL, and absolute neutrophil count 1500/mL, regression of splenomegaly according to physical examination, lack of HC with peripheral distribution, and in bone marrow were assessed as full response. Peripheral blood samples close to normalizing, close to 50% resolution of splenomegaly, and bone marrow HC were accepted as partial responses. Observation of HCs during follow-up of blood and bone marrow samples from patients with full or partial response was defined as morphologic relapse. Hematologic relapse was defined as the reappearance of cytopenia below the thresholds defined above for complete response (CR) and partial response (PR). Overall survival (OS) was defined as the duration from diagnosis until death. Progression-free survival (PFS) was defined as the duration from treatment until patient recurrence.

With the aim of pretreatment assessment by a radiology expert experienced in the field, patients in the study group were assessed for muscle and fat tissue quality using abdominal CT images. To ensure measurement standardization, measurements were performed using the ImageJ program (National Institutes of Health, Bethesda, Maryland, USA) on abdominal CT sections showing the L3 vertebra. Within this scope, the areas of total paravertebral muscle tissue, visceral adipose tissue (VAT), and subcutaneous adipose tissue (SAT) were separately calculated in mm^2. To normalize differences in body structure, values were divided by the square of the patient's height; thus, the skeletal mass index (SMI) was calculated as mm^2/m^2. Examples of CT assessments are shown in Figure 2. Patients were divided into two groups according to whether they were sarcopenic or not, and their clinical results were compared.

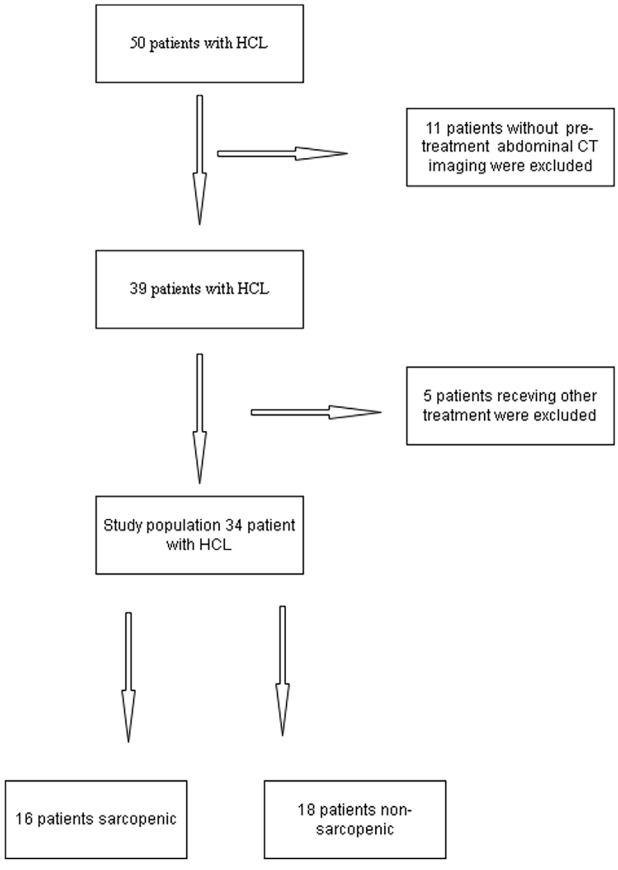


Figure 1. Study design and patient flowchart
HCL: Hairy cell leukemia, CT: Computed tomography

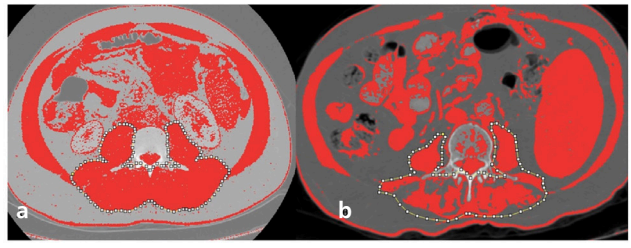


Figure 2. Sarcopenia assessment the areas of total paravertebral muscle at the mid-level of the third lumbar vertebrate by CT scan. (A) Example of a non-sarcopenic man with normal muscle mass, (B) Example of a sarcopenic man with reduced muscle mass
CT: Computed tomography

Statistical Analysis

Statistical analysis were performed using IBM SPSS Statistics for Windows, version 23.0 (IBM Corp., Armonk, NY). The assumption of normality was checked using the Shapiro-Wilk test. Descriptive analyses are presented as mean ± standard deviation [interquartile range (IQR)], median (IQR), or n (%), as appropriate. To determine the prognosis of HCL, categorical data were analyzed using the Pearson's chi-square test, and numerical data with non-normal and normal distribution were analyzed using the Mann-Whitney U test or Student's t-test, respectively. Receiver operating characteristic analysis was performed to determine the optimal cut-off point for SMI. The optimal cutoff value for SMI for predicting sarcopenia was determined by receiver operating characteristic analysis. Patients with SMI 3028.3 (mm^2/m^2) were assessed as being sarcopenic. Survival curves were created using the Kaplan-Meier method, and the Breslow, Tarone-Ware, and log-rank tests were used to assess differences between groups. Univariate logistic regression analysis was used to identify independent risk factors related to treatment response. Univariate and multivariate analyses for independent predictors of OS and PFS were performed using the Cox proportional risk regression model. Univariate analyses ($p<0.1$) included age, spleen size, response to treatment (RTT), SAT, total mass index (TMI), SMI, and sarcopenia, which were also tested in multivariate models. Hazard ratios equivalent to 95% confidence intervals or odds ratios are reported. Variables with p-values <0.05 were considered significant.

Results

Of the patients who participated in the study, 31 were men (91.2%) and 3 were women (8.8%). Among the patients, 16 (47%) were sarcopenic. Table 1 summarizes the characteristics of patients with and without sarcopenia. The sarcopenic group had significantly lower hemoglobin values than the non-sarcopenic group (9.15 vs. 11.26; p=0.030). The non-sarcopenic patients had lower TMI than the sarcopenic patients (4662.78 vs. 6180.67; p<0.01).

Table 1. The characteristics of patients according to the presence of sarcopenia

Variables	Overall (n=34)	S^+ (Sarcopenic) (nS^+=16)	S^- (Non-sarcopenic) (nS^-=18)	p-value
Age	53.38±12.59	56.75±14.48	50.38±10.13	0.144
Sex				
Female (%) Men (%)	3 (8.8) 31 (91.2)	3 (18.8) 13 (1.2)	0 (0) 18 (100)	0.094
Spleen size (mm)	183.61±44.21	181.5±51.17	185.5±38.43	0.797
Hemoglobin (gr/dL) WBC (mm^3) PLT (h/mL)	10.26±2.92 2940 (2137.5) 56.5 (43.25)	9.15±2.06 2695 (2182.5) 63 (78.75)	11.26±3.27 3285 (4392.5) 54.5 (25.25)	**0.030*** 0.164 0.932
BSA (m^2)	1.86±0.15	1.83±0.16	1,88±0,14	0.294
BMI (kg/m^2) TMI (mm^2/m^2) SMI (mm^2/m^2)	26.5 (3.77) 5466.37±1004.08 3002.73±541.84	24.47 (5.66) 4662.78±694.60 2540.66±328.40	27.24 (3.14) 6180.67±619.13 3413.46±308.53	**0.025*** **0.001*** **0.0001***
VAT	15153.52±7779.98	14171.68±8743.28	16026.27±6952.00	0.496
SAT	14903 (11026.5)	15250.5 (15073.7)	14704.5 (10162.7)	0.695
PFS	35.5 (50.75)	23.5 (32.25)	59 (44)	**0.009***
OS	46.5 (48.5)	26.5 (67)	63 (37)	**0.046***
RTT				
Yes (CR or PR) % No	31 (91.2) 3 (8.8)	13 (81.3) 3 (18.7)	18 (100) 0 (0)	0.094
Relapse				
Yes % No %	11 (2.4) 23 (67.6)	5 (31.3) 11 (68.7)	6 (33.3) 12 (66.7)	0.594
Survival				
Alive % Exitus %	29 (85.3) 5 (14.7)	12 (75) 4 (25)	17 (94.4) 1 (5.6)	0.133

Student's t test, Mann-Whitney U test, Pearson chi-square test. Data are presented with n (%), mean ± SD (IQR) or median (IQR)
*Bold values denote statistical significance at the p<0.05 level
S^+: Sarcopenic, S^-: Non-sarcopenic, WBC: White blood cell, PLT: Platelet, BMI: Body mass index, TMI: Total mass index, BSA: Body surface area, VAT: Visceral adipose tissue, SMI: Skeletal mass index, SAT: Subcutaneous adipose tissue, OS: Overall survival, PFS: Progression-free survival, RTT: Response to treatment, IQR: Interquartile range, SD: Standard deviation, CR: Complete response, PR: Partial response

Table 2 presents the results of a univariate logistic regression analysis to identify variables affecting treatment response in HCL patients and a Cox regression survival analysis to identify factors affecting PFS and OS. In the multivariate analysis, spleen size (HR=1.014, p=0.0041), RTT (HR=0.027, p=0.004), and sarcopenia (HR=0.154, p=0.043) were effective variables for OS. For PFS, spleen size (HR=1.012, p=0.037) and RTT (HR=0.033, p=0.003) were statistically significant variables (Table 3).

The sarcopenic group's OS was generally shorter than that of the non-sarcopenic group (Figure 3a). The 12-month mean OS rates for the sarcopenic and non-sarcopenic groups were 62.5% and 94.4%, respectively, whereas the median OS was 26 months in the sarcopenic group and 62 months in the non-sarcopenic group. There was a significant difference in PFS between the two groups. The PFS was significantly better in the non-sarcopenic group than in the sarcopenic group (Figure 3b). The 12-month mean PFS rates for the sarcopenic and non-sarcopenic groups were 62.5% and 94.4%, respectively, whereas the median PFS was 54 months in the non-sarcopenic group and 21 months in the sarcopenic group.

Table 2. Univariate logistic regression analysis for response to treatment and univariate Cox regression analysis for OS and PFS in patients with HCL

Variables	Response to treatment		OS		PFS	
	OR (95% CI)	p-value	HR (95% CI)	p-value	HR (95% CI)	p-value
Age	1.106 (0.99-1.23)	0.071	1.092 (1.013-1.18)	0.021*	1.04 (0.98-1.10)	0.226
Sex	0.998 (0.89-1.07)	0.990	1.17 (0.35-3.87)	0.802	23.45 (0-878518)	0.557
Spleen size	0.915 (0.83-1.01)	0.073	0.97 (0.95-1.01)	0.056	0.99 (0.98-1.01)	0.366
Hemoglobin	1.018 (0.676-1.53)	0.934	0.93 (0.69-1.25)	0.622	1.07 (0.86-1.32)	0.559
RTT	-	-	0.035 (0.01-0.22)	**0.001***	21.38 (2.93-155.7)	**0.003***
Relapse	0.952 (0.08-11.79)	0.990	0.448 (0.5-4.05)	0.475	-	-
BMI	0.552 (0.32-0.96)	**0.033***	0.78 (0.59-1.02)	0.280	0.89 (0.73-1.09)	0.280
BSA	0.001 (0-25)	0.182	0.01 (0-1.55)	0.073	0.01 (0.00005-1.5)	0.073
VAT	-	0.171	-	0.667	1.00 (0.99-1.00)	0.929
SAT	00.9996 (0.9993-0.9999)	**0.038***	-	0.061	1.00 (0.99-1.00)	0.509
TMI	0.99 (0.987-1.001)	0.080	0.998 (0.997-0.999)	**0.017***	1.00 (0.99-1.00)	0.964
SMI	0.998 (0.995-1.005)	0.093	0.998 (0.996-0.999)	**0.027***	0.998 (0.996-0.999)	**0.025***
Sarcopenia	-	0.998	0.161 (0.018-1.463)	0.105	0.379 (0.099-1.429)	0.150

*Bold values denote statistical significance at the p<0.05 level
OS: Overall survival, PFS: Progression-free survival, HCL: Hairy cell leukemia, HR: Hazard ratio, RTT: Response to treatment, BMI: Body mass index, BSA: Body surface area, VAT: Visceral adipose tissue, SAT: Subcutaneous adipose tissue, TMI: Total mass index, SMI: Skeletal mass index

Table 3. Multivariate Cox regression analysis for OS and PFS in patients with HCL

Variables	OS		PFS	
	HR (95% CI)	p-value	HR (95% CI)	p-value
Age	1.00 (0.96-1.06)	0.738	1.01 (0.96-1.1)	0.585
Spleen size	1.014 (1.001-1.027)	**0.041***	1.012 (0.97-1.33)	**0.037***
Hemoglobin	1.07 (0.92-1.25)	0.343	1.14 (1.02-1.42)	0.115
BMI	1.07 (0.91-1.26)	0.412	1.09 (0.94-1.25)	**0.021***
RTT	0.027 (0.02-0.31)	**0.004***	0.033 (0.003-0.32)	**0.003***
Relapse	0.311 (0.083-1.166)	0.083	-	-
VAT	1.000 (0.99-1.00)	0.263	1 (1-1)	0.418
SAT	1.000 (0.99-1.00)	0.210	1 (1-1)	0.274
TMI	0.999 (0.99-1.00)	0.549	0.999 (0.997-1.01)	0.358
SMI	1.002 (0.99-1.01)	0.383	1.002 (0.998-0.005)	0.325
Sarcopenia	0.154 (0.025-0.983)	**0.043***	0.17 (0.034-0.842)	0.137

*Bold values denote statistical significance at the p<0.05 level
OS: Overall survival, PFS: Progression-free survival, HCL: Hairy cell leukemia, RTT: Response to treatment, BMI: Body mass index, BSA: Body surface area, VAT: Visceral adipose tissue, SAT: Subcutaneous adipose tissue, TMI: Total mass index, SMI: Skeletal mass index

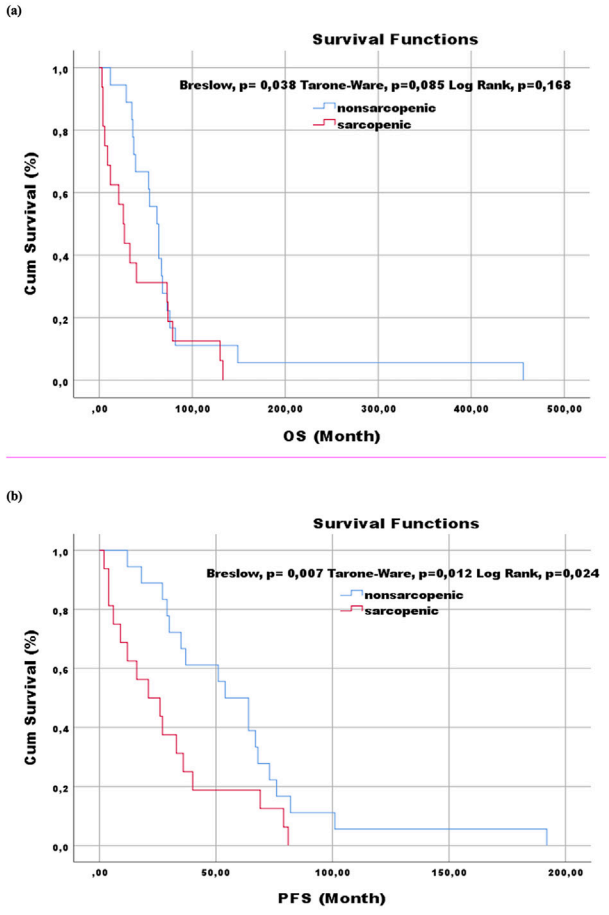


Figure 3. OS (a) and progression-free survival (b) according to SMI

OS: Overall survival, SMI: Skeletal mass index, PFS: Progression-free survival

Discussion

This study is significant because it is the first to demonstrate that sarcopenia is an independent factor affecting OS in HCL patients. Patients with diffuse large B cells, older patients, and those with early sarcopenia were reported to have shorter OS durations (12). Additionally, patients with Hodgkin lymphoma identified to have sarcopenia using SMI were reported to have shorter OS (13). Albano et al. (14) reported that sarcopenia did not affect the OS of patients with mantle cell lymphoma. In a study of leukemia patients, Nakamura et al. (15) showed that sarcopenic AML patients, especially aged over 60 years, had 0% 3-year OS. Patients with sarcopenic HCL had a significantly shorter OS duration in our study compared to those without sarcopenic HCL (26.5 vs. 63 months, p=0.046).

In our study, patients with sarcopenia had significantly lower BMI, TMI, and SMI. The prognostic importance of TMI and SMI for OS was shown. SMI was a statistically significant factor for PFS (p=0.025, HR=0.998). Inflammatory cytokines, tumor-derived factors, and growth factors released from tumors in patients with malignancy cause sarcopenia through both degradation and reduced protein synthesis (16). Linked to this, injury occurs in the muscles of the skeletal system. This leads to reductions in muscle strength, quality, amount, and yield. TMI and SMI values may be affected after treatment. Therefore, sarcopenia may be a factor affecting both treatment response and PFS, which is linked to the lack of full effect of the treatment protocol.

The most important feature distinguishing our study from other studies is the calculation of the sarcopenia cut-off value using a different method. Studies on sarcopenia trends do not have a standardized cutoff value because different patient populations and differences in muscle regions are used for calculations [SMI, psoas mass index (PMI)] (17). Several studies published in this field to reveal the sarcopenic status of patients used threshold values according to sex as a significant determinant of sarcopenia (18). Due to differences in PMI values between male and female patients, using sex-specific cut-off points obtained for PMI by receiver operating characteristic curve analysis, patients were divided into two groups: sarcopenic and non-sarcopenic. In our study, an index for all skeletal muscles was calculated, not just PMI, and no statistically significant difference was observed for SMI values between male and female patients. Accordingly, our study did not use sex-based threshold values. In our study, the presence of sarcopenia was determined using an effect size-based method using threshold values obtained from the ROC curve analysis for mean and median SMI.

Study Limitations

One of our study's most important limitations is the small number of patients. We believe that studies with more homogeneous and adequate patient numbers will allow for a better interpretation of the effects of sarcopenia on OS. Another limitation was that most patients did not undergo abdominal CT after treatment, making it impossible to assess sarcopenia after treatment. Finally, the Revised European Working Group on Sarcopenia in Older People-EWGSOP reported that muscle function is important for the assessment of sarcopenia (19). Muscle function could not be assessed in this retrospective study. Studies on HCL in previous years reported that splenomegaly (>3 cm), leukocytosis (>10^9/L), and high beta-2 microglobulin levels were poor prognostic factors in HCL patients (20). The major strength of this study is that our study differed from other studies in that the calculation of the sarcopenia cut-off value was performed using a different method.

Conclusion

Our study is significant because it is the first to show that sarcopenia is a prognostic factor for prognosis and treatment parameters. We believe that our study will guide future scoring studies related to sarcopenia during HCL patient diagnosis.

Footnote

Ethics Committee Approval: The study was approved by the Ondokuz Mayıs University Clinical Research Ethics Committee (approval no.: OMÜ KAEK 2022/576, date: December 28, 2022).

Informed Consent: Retrospective study.

Authorship Contributions

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

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

Financial Disclosure: This study received no financial support.

References

1. Paillassa J, Troussard X. Biology and Treatment of Hairy Cell Leukemia. Curr Treat Options Oncol. 2020;21:44.
2. Frassoldati A, Lamparelli T, Federico M, et al. Hairy cell leukemia: a clinical review based on 725 cases of the Italian Cooperative Group (ICGHCL). Leuk Lymphoma. 1994;13:307-16.
3. Grever MR, Abdel-Wahab O, Andritsos LA, et al. Consensus guidelines for the diagnosis and management of patients with classic hairy cell leukemia. Blood. 2017;129:553-60.
4. Teras LR, DeSantis CE, Cerhan J R, et al. 2016 US lymphoid malignancy statistics by World Health Organization subtypes. CA Cancer J Clin. 2016;66:443-59.
5. Tadmor T, Polliack A. Epidemiology and environmental risk in hairy cell leukemia. Best Pract Res Clin Haematol. 2015;28:175-9.
6. Saven A, Burian C, Koziol JA, Piro LD. Long-term follow-up of patients with hairy cell leukemia after cladribine treatment. Blood. 1992;92:1918-26.
7. Rosenberg JD, Burian C, Waalen J, Saven A. Clinical characteristics and long-term outcome of young hairy cell leukemia patients treated with cladribine: a single-institution series. Blood. 2014;123:177-83.
8. Pamoukdjian F, Bouillet T, Lévy V, Soussan M, Zelek L, Paillaud E. Prevalence and predictive value of pre-therapeutic sarcopenia in cancer patients: A systematic review. Clin Nutr. 2018;37:1101-13.
9. Argilés JM, Busquets S, Stemmler B, López-Soriano FJ. Cachexia and sarcopenia: mechanisms and potential targets for intervention. Curr Opin Pharmacol. 2015;22:100-6.
10. Surov A, Wienke A. Sarcopenia predicts overall survival in patients with malignant hematological diseases: A meta-analysis. Clin Nutr. 2021;40:1155-60.
11. Colloca GF, Bellieni A, Di Capua B, et al. Sarcopenia Diagnosis and Management in Hematological Malignancies and Differences with Cachexia and Frailty. Cancers (Basel). 2023;15:4600.
12. Nakamura N, Hara T, Shibata Y, et al. Sarcopenia is an independent prognostic factor in male patients with diffuse large B-cell lymphoma. Ann Hematol. 2015;94:2043-53.
13. Zilioli VR, Albano D, Arcari A, et al. Clinical and prognostic role of sarcopenia in elderly patients with classical Hodgkin lymphoma: a multicentre experience. J Cachexia Sarcopenia Muscle. 2021;12:1042-55.
14. Albano D, Pasinetti N, Dondi F, Giubbini R, Tucci A, Bertagna F. Prognostic Role of Pre-Treatment Metabolic Parameters and Sarcopenia Derived by 2-[18F]-FDG PET/CT in Elderly Mantle Cell Lymphoma. J Clin Med. 2022;11:1210.
15. Nakamura N, Ninomiya S, Matsumoto T, et al. Prognostic impact of skeletal muscle assessed by computed tomography in patients with acute myeloid leukemia. Ann Hematol. 2019;98:351-9.
16. Fukushima H, Fujii Y, Koga F. Metabolic and Molecular Basis of Sarcopenia: Implications in the Management of Urothelial Carcinoma. Int J Mol Sci. 2019;20:760.
17. Xu XT, He DL, Tian MX, Wu HJ, Jin X. Prognostic Value of Sarcopenia in Patients With Diffuse Large B-Cell Lymphoma Treated With R-CHOP: A Systematic Review and Meta-Analysis. Front Nutr. 2022;9:816-83.
18. Iltar U, Sözel H, Sözel YK, et al. Prognostic impact of the psoas muscle index, a parameter of sarcopenia, in patients with diffuse large B-cell lymphoma treated with rituximab-based chemoimmunotherapy. Leuk Lymphoma. 2021;62:1098-106.
19. Cruz-Jentoft AJ, Bahat G, Bauer J, et al. Sarcopenia: revised European consensus on definition and diagnosis. Age Ageing. 2019;48:16-31.
20. Forconi F, Sozzi E, Cencini E, et al. Hairy cell leukemias with unmutated IGHV genes define the minor subset refractory to single-agent cladribine and with more aggressive behavior. Blood. 2009;114:4696-702.

Original Article

DOI: 10.4274/haseki.galenos.2024.9916
Med Bull Haseki 2024;62:229-234

Nivolumab Efficacy and Safety in Cancer Patients with Renal Dysfunction

Ezgi Degerli, Mesut Yilmaz, Gulcin Sahingoz Erdal, Rumeysa Colak, Caner Kapar, Ayse Gul Satman Durmaz, Esra Turan Canbaz, Deniz Tural

University of Health Sciences Turkey, Bakirkoy Dr. Sadi Konuk Training and Research Hospital, Department of Medical Oncology, Istanbul, Turkey

Abstract

Aim: Immune checkpoint inhibitors have transformed cancer treatment; however, clinical trials often exclude patients with renal dysfunction. This study aimed to evaluate the efficacy and safety of nivolumab in this population, addressing the limited available data.

Methods: A retrospective cross-sectional study was conducted on patients with baseline renal dysfunction who received nivolumab between 2018 and 2023. The safety and efficacy endpoints, including immune-related adverse events (irAEs), treatment response, and progression-free survival.

Results: Fifty patients with various malignancies were included, with 30% experiencing manageable worsening of renal function. Approximately 51% of patients experienced no irAEs, whereas 8% experienced grade 3 or 4 adverse events. The treatment discontinuation rate due to adverse effects was 2%. Significantly, 68% of patients showed treatment benefits, with a median progression-free survival of 450 days.

Conclusion: Nivolumab is effective and safe for patients with renal dysfunction, with comparable outcomes to those without renal impairment. Despite the occurrence of IrAEs, they were manageable, and we observed benefits in long-term progression-free survival.

Keywords: Nivolumab, immune checkpoint inhibitors, kidney diseases, drug-related side effects and adverse reactions, progression-free survival

Introduction

In recent years, immune checkpoint inhibitors have greatly improved the outcomes of patients with malignancies. Many patients receive immunotherapy or a combination of chemotherapy and immunotherapy (1). However, most clinical trials on immunotherapies do not include patients with kidney problems, even though managing this specific group of patients presents practical challenges. Conventional chemotherapy often does not work well for patients with kidney problems, raising concerns about the effectiveness and safety of immunotherapies in this population.

Nivolumab is an approved anti-programmed cell death protein 1 monoclonal antibody used to treat various cancers, including non-small-cell lung cancer (NSCLC), renal cell carcinoma (RCC), melanoma, urothelial carcinoma, squamous cell carcinoma of the head and neck, colorectal cancer, and esophageal and gastric cancer (2). The effects of nivolumab in patients with existing kidney problems have not been thoroughly evaluated. Like other immunotherapies, nivolumab may cause immune-related adverse events (irAEs), such as joint pain, colitis, hepatitis, pneumonitis, rash, vitiligo, nephritis, and endocrinopathies (3,4). Notably, dose adjustments for nivolumab are not recommended for patients with kidney problems.

In this study, we hypothesized that, based on the mechanisms of immunotherapy, nivolumab treatment would not worsen kidney problems or lead to a higher incidence of irAEs in patients with kidney impairment. We aimed to retrospectively analyze the safety and efficacy of nivolumab in this specific patient population, focusing on relevant clinical endpoints.

Address for Correspondence: Ezgi Degerli, University of Health Sciences Turkey, Bakirkoy Dr. Sadi Konuk Training and Research Hospital, Department of Medical Oncology, Istanbul, Turkey
Phone: +90 505 556 58 88 **E-mail:** ezgitastan.19@gmail.com **ORCID:** orcid.org/0000-0002-8664-5701
Received: 10.05.2024 **Accepted:** 16.09.2024

Methods

Compliance with Ethical Standards

The study protocol and subject matter were reviewed and approved by the University of Health Sciences Turkey, Bakirkoy Dr. Sadi Konuk Training and Research Hospital Clinical Research Ethics Committee (approval no.: 2024/47, date: 04.03.2024). The research design was retrospective cross-sectional. The ethics committee anonymized and approved the database information without obtaining consent.

Study Design

Patients with baseline renal dysfunction treated with nivolumab between 2018 and 2023 were retrospectively screened (Figure 1). All patients had chronic renal failure. The additional inclusion criteria consisted of receiving at least one dose of either nivolumab or baseline renal dysfunction. Renal dysfunction was defined as an estimated glomerular filtration rate (eGFR) ≤60 mL/min/1.73 m^2. The eGFR was calculated using the creatinine equation published by The Chronic Kidney Disease Epidemiology Collaboration (5). Patients who received combination chemotherapy or immunotherapy with multiple agents (ipilimumab + nivolumab or chemotherapy + nivolumab) were excluded.

Statistical Analysis

The statistical analyses were conducted using SPSS (Statistical Package for the Social Sciences) version 22.0 for Windows. Intergroup comparisons of normally distributed data were performed using Student's t-test, while non-normally distributed data were analyzed using the Mann-Whitney U test. We used these tests to determine demographic and clinical characteristics, laboratory findings, and renal dysfunction in patients. Additionally, the Kaplan-Meier estimator was used to determine the progression-free survival (PFS) and overall survival (OS) functions.

Results

Patients

Table 1 presents patient demographic data and characteristics. Fifty patients with advanced malignancies and baseline eGFR 60 mL/min/1.73 m^2 received nivolumab. The median age was 67.5 years [interquartile range (IQR): 60.7-71.2]. Most had renal cell carcinoma (70%), were male (72%), and had an Eastern Cooperative Oncology Group performance score of 0-2. All patients received prior treatment, and those with RCC received prior tyrosine kinase inhibitor therapy. The median nivolumab dose was 10 (IQR: 6-28) cycles. Baseline creatinine values ranged between 1.2-4.7 mg/dL (median 1.34, IQR: 1.2-1.56), and baseline eGFR ranged between 58-21 mL/min/1.73 m^2 (median 54 IQR: 44-57, excluding patients on dialysis). Three patients had an eGFR of 30 mL/min/1.73 m^2, two of whom stage had 5 end-stage renal failure and were undergoing regular hemodialysis.

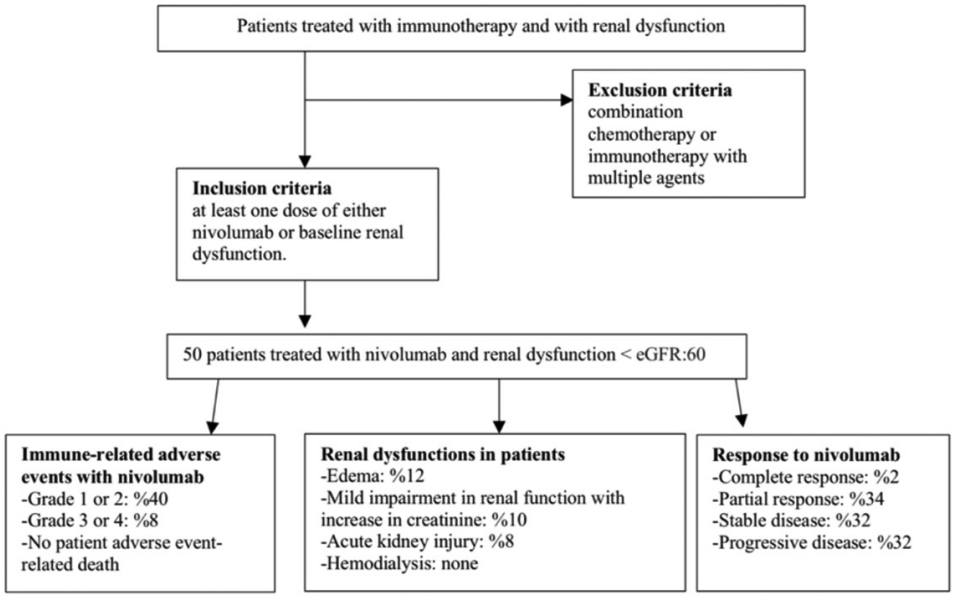


Figure 1. Flow chart of patient treatment

eGFR: Estimated glomerular filtration rate

Safety and Efficacy

Approximately half of the patients did not present with any grade (n=26, 51%). Grade 3 or 4 adverse events occurred in 8% of the patients who received nivolumab; the most common grade 3 or 4 adverse events were pneumonitis, hipofizitis, myocarditis, and nephrite. Grade 1 or 2 adverse events occurred in 40% of the patients; the most common grade 1 or 2 adverse events were thyroiditis and fatigue (Table 2).

One patient had metastatic RCC with a baseline eGFR of 40 mL/min/1.73 m^2 and concomitant brucellosis that occurred during treatment and experienced grade 3 nephritis that resolved with corticosteroids (prednisone 1 mg/kg). The other patient had NSCLC and cardiac dysfunction (ejection fraction 25-30%) and experienced grade 3 myocarditis that emerged concurrently and rapidly resolved with high-dose (1 mg/kg) prednisone. Another patient had grade 4 pneumonitis after nivolumab therapy and responded to high-dose prednisolone, but his imaging findings were consistent with hyper-progressive disease. Treatment was discontinued in these three patients.

We examined the effects of nivolumab on baseline renal dysfunction in patients. We examined renal worsening in 4 groups: development of edema with proteinuria, mild impairment in renal function with increased creatinine, development of acute kidney injury (AKI), and need for hemodialysis (Table 3). Worsening renal function occurred in 15 (30%) patients.

Thirty-four patients (68%) experienced treatment benefits (complete response, partial response, or stable disease). One (2%) patient had a complete response, 16 (32%) had stable disease as the best response, 17 (34%) had a partial response, and 16 patients (32%) had primary progressive disease. Sixteen patients (32%) had primary progressive disease (Table 4). The patient with a complete response received nivolumab for RCC and had a baseline eGFR: 42 mL/min/1.73 m^2. The patient is alive, and treatment is ongoing. Of the patients with partial responses, four had NSCLC and 13 had RCC. Among two patients on dialysis, one responded to therapy, and the other responded, although none of these patients experienced significant toxicities.

One patient experienced hyperprogression and death within 60 days of treatment initiation. This patient had grade 4 immune-related pneumonitis and was evaluated as having hyperprogression. The median PFS from initial treatment was 450 days with nivolumab (Figure 2), and the median OS was not reached.

Table 1. Baseline characteristics of the patients

Baseline characteristic (n=50)	n (%)
Median age (range)	67.5 (43-83)
Gender	
Male	36 (72%)
Female	14 (28%)
Disease state	
Renal cell carcinoma	35 (70%)
Nonsmall cell lung cancer	9 (18%)
Melanoma	4 (8%)
Urothelial cell carcinoma	1 (2%)
Mesothelioma	1 (2%)
ECOG	
0	14 (28%)
1	30 (60%)
2	6 (12%)
Comorbities	
Diabetes mellitus	14 (28%)
Hypertension	24 (48%)
Ischemic heart disease	13 (26%)
Chronic obstructive pulmonary disease	4 (8%)
Median nivolumab doses	10.5
Median creatinine level (range) mg/dL	1.34 (1-4.77)
Median eGFR mL/min/1.73 m^2	54

ECOG: Eastern Cooperative Oncology Group, eGFR: Estimated glomerular filtration rate

Table 2. Immune-releated adverse events with nivolumab

irAEs	Grade 1 or 2 (n)	Grade 3 or 4 (n)
Fatique	5	0
Hypophysitis	2	1
Thyroiditis	9	0
Myocarditis	0	1
Nephritis	1	1
Pneumonitis	3	1
Rash	3	0
Vitiligo	2	0

irAEs: Immune-related adverse events

Discussion

In this study, immunotherapy use in patients with cancer and baseline renal dysfunction resulted in irAE rates similar to those in previous clinical trials, including those in patients without renal dysfunction. Encouraging results were demonstrated in this heavily pretreated population with multiple comorbid illnesses. Clinical trials on nivolumab have included patients without renal dysfunction. In our study, the use of nivolumab therapy in patients with baseline renal dysfunction resulted in rates of irAEs and effectivity similar to those of clinical trials, including patients without renal dysfunction (6-8). Our patient group consisted of patients with multiple comorbid illnesses who received pretreatment. The 15-month PFS for the second treatment series and beyond was similar to that of other studies (9-11). This finding increased our confidence in using nivolumab for fragile patients with renal dysfunction.

Renal dysfunction was divided into four parameters and examined: development of edema with proteinuria, mild impairment of renal function with increased creatinine, development of AKI, and need for dialysis. None of the patients experienced worsening renal dysfunction until week 12. After the 12th week, peripheral edema was observed in six patients with proteinuria, but there was no increase in creatinine levels. Two of them were already on hemodialysis. This condition may be associated with inadequate hemodialysis. The volume of fluid administered with nivolumab may not have contributed to the progression of edema (100 cc). The patients responded to diuretic treatment, and they did not recur with the adjustment of oral diuretic dose. There was a moderate creatinine increase in 5 patients related to volume depletion. The oral intake of these patients needed to be improved. Response to intravenous isotonic liquid replacement. Four patients were diagnosed with acute renal injury according to the AKI criteria (12). Two of these cases involved immune-related nephritis, which was confirmed by renal biopsy.

Table 3. Renal dysfunction in patients	
No renal or clinic dysfunction	35 (70%)
Edema	6 (12%)
Increased creatinine in patients without AKI	5 (10%)
AKI	4 (8%)
Need for dialysis	0
AKI: Acute kidney injury	

Table 4. Response to nivolumab	
Response	**n (%)**
Complete response	1 (2%)
Partial response	17 (34%)
Stable disease	16 (32%)
Progressive disease	16 (32%)

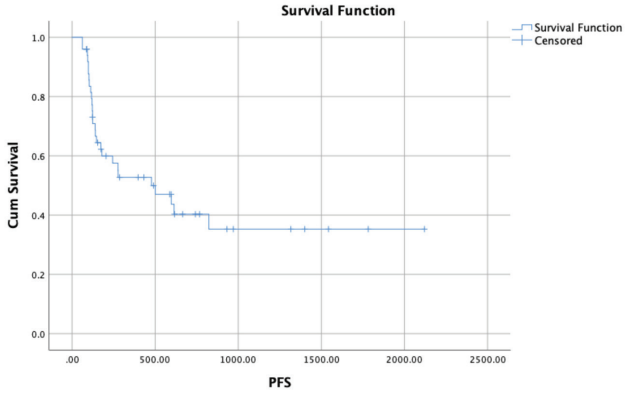


Figure 2. Progression-free survival
PFS: Progression-free survival

According to the American Society of Clinical Oncology guidelines (13), kidney biopsy should be considered in patients undergoing immunotherapy if they have proteinuria >3 g, oliguria, dysmorphic hematuria, or who do not respond well to initial treatment with corticosteroids. However, recent studies and case reviews have suggested that renal biopsy is important and should be performed even in patients who do not meet these criteria. In a recent case review, Rashidi et al. (14) recommended kidney biopsy when a clinical workup does not provide a clear explanation for AKI in the context of immune checkpoint inhibitor therapy, even in the absence of the traditional criteria for kidney biopsy, such as heavy proteinuria, hematuria, and pyuria (14).

Patients with immune-related nephritis responded to high-dose immunosuppressive therapy. One patient with grade 3 nephritis had concomitant brucellosis. Nivolumab could not be continued in this patient. However, nivolumab was continued in another patient after completion of immunosuppressive therapy. Compared with the literature, the risk of developing grade 3-4 immune side effects was the same in patients without renal dysfunction. No patient required additional hemodialysis with nivolumab therapy. Our two patients were already undergoing hemodialysis, and one patient on dialysis had previously undergone failed kidney transplantation.

Most clinical trials for cancer therapies usually exclude patients because they primarily focus on studying the effects and characteristics of these drugs, which is not feasible in dialysis patients (15). Although dose adjustment is not typically required in these patients, the risk of developing irAEs appears to be similar to that of the general population. This may be attributed to the fact that ICIs are not excreted through the kidneys, so it's logical that the frequency of adverse reactions is similar in both populations (16). Unfortunately, data are limited to the use of immunotherapy in patients undergoing kidney transplantation (17,18). The old allograft can be rejected upon initiating immunotherapy in patients undergoing hemodialysis (19,20). To prevent this, the dose regulation of immunosuppressive treatment can be performed before immunotherapy, and additional steroids can be administered if necessary (21). Hirsch et al. (22) described a case of acute rejection after immunotherapy in a patient undergoing dialysis who had previously undergone kidney transplantation. Mejia et al. (23) reported a case of a failed kidney allograft in a patient undergoing hemodialysis who received nivolumab and ipilimumab for metastatic papillary renal cancer. We can explain this situation using findings consistent with cell-mediated rejection induced by the blockade of the PD-1

pathway. This case series by Strohbehn et al. (24) discussed the effectiveness and safety of immunotherapy in 19 patients undergoing hemodialysis, with a high success rate. Four of these patients had previously undergone failed kidney transplantation. None of the patients showed any clinical signs or symptoms of rejection of the failed allograft when treated with immunotherapy, and none experienced abdominal pain (24). Similarly, the patient who underwent renal transplantation also did not experience renal rejection. They did not require additional immunosuppressive therapy. However, progressive disease was detected in scans obtained 3 months later.

Our study divided cancer outcomes in patients into four categories: complete remission, partial remission, stable disease, and progressive disease. Of all the patients studied, 65 had evidence of remission or stable disease. The response rate to treatment was similar to that reported in the literature for patients who did not experience renal function loss (25,26).

Study Limitations

The main limitations of our study are its retrospective and cross-sectional. Patient numbers overall were limited, limiting the ability to draw definitive conclusions. Most patients had stage 3 chronic kidney disease (eGFR between 30 and 60 mL/min/1.73 m^2). The number of patients with an eGFR 30 mL/min/1.73 m^2 was limited. This led to inadequate assessment in patients with significant renal damage. As a result, they might depict a relatively healthier cohort of patients with underlying renal dysfunction. However, conclusions regarding patients with end-stage renal failure cannot be reached. Despite these limitations, we believe this single-center experience is valuable because it will contribute to the literature by describing the side effects and efficacy in this fragile patient group. It is well known that data regarding patients with renal dysfunction are limited in the literature (27), highlighting the need for new research and treatments.

Conclusion

Although these patients can tolerate immunotherapy, which offers expectancy and moderate survival benefits, monitoring the side effects is crucial for safety. Nivolumab was shown to be efficacious and safe in these patients, with no increase in adverse events. Long-term studies are needed to confirm these findings.

Footnote

Ethics Committee Approval: The study protocol and subject matter were reviewed and approved by the University of Health Sciences Turkey, Bakirkoy Dr. Sadi Konuk Training and Research Hospital Clinical Research Ethics Committee (approval no.: 2024/47, date: 04.03.2024).

Informed Consent: The research design was retrospective cross-sectional. The ethics committee anonymized and approved the database information without obtaining consent.

Authorship Contributions

Concept: E.D., Design: E.D., M.Y., D.T., Data Collection or Processing: G.S.E., A.G.S.D., Analysis or Interpretation: R.C., C.K., Literature Search: E.D., C.K., E.C.T., Writing: E.D.

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

Financial Disclosure: This study received no financial support.

References

1. Franzin R, Netti GS, Spadaccino F, et al. The Use of Immune Checkpoint Inhibitors in Oncology and the AKI Outbreak: Where Do We Stand? Front Immunol. 2020;11:574271.
2. https://www.drugs.com/history/opdivo.html./O4.WEN.2024
3. Les I, Martínez M, Pérez-Francisco I, et al. Predictive Biomarkers for Checkpoint Inhibitor Immune-Related Adverse Events. Cancers (Basel). 2023;15:1629.
4. Amoroso V, Gallo F, Alberti A, et al. Immune-related adverse events as potential surrogates of immune checkpoint inhibitors' efficacy: a systematic review and meta-analysis of randomized studies. ESMO Open. 2023;8:100787.
5. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate [Erratum in Ann Intern Med. 2011 Sep 20;155(6):408]. Ann Intern Med. 2009;150:604-12.
6. Tykodi SS, Gordan LN, Alter RS, et al. Safety and efficacy of nivolumab plus ipilimumab in patients with advanced non-clear cell renal cell carcinoma: results from the phase 3b/4 CheckMate 920 trial. J Immunother Cancer. 2022;10:e003844.
7. Naumann RW, Hollebecque A, Meyer T, et al. Safety and Efficacy of Nivolumab Monotherapy in Recurrent or Metastatic Cervical, Vaginal, or Vulvar Carcinoma: Results of the Phase I/II CheckMate 358 Trial. J Clin Oncol. 2019;37:2825-34.
8. Zhang H, Zhang L, Chen K, et al. Efficacy and safety of nivolumab for advanced/recurrent non-small-cell lung cancer: an up-to-date meta-analysis of large-scale phase III randomized controlled trials. Future Oncol. 2022;18:3667-75.
9. Borghaei H, Gettinger S, Vokes EE, et al. Five-Year Outcomes from Randomized Phase III Trials CheckMate 017 and 057: Nivolumab Versus Docetaxel in Patients with Previously Treated Non-Small-Cell Lung Cancer. J Clin Oncol. 2021 Mar 1;39:723-33. Epub 2021 Jan 15. Erratum in J Clin Oncol. 2021;39:1190.
10. Blas L, Shiota M, Miyake H, et al. Clinical factors associated with tumor response, progression, and survival in patients with advanced renal cell carcinoma treated with nivolumab in the SNiP-RCC study. Int J Urol. 2023;30:788-96.

11. Motzer RJ, Escudier B, McDermott DF, et al. CheckMate 025 Investigators. Nivolumab versus Everolimus in Advanced Renal-Cell Carcinoma. N Engl J Med. 2015;373:1803-13.
12. KDIGO Clinical Practice Guidelines for Acute Kidney Injury. Kidney International Suppl (2012)2,8-12.)
13. Schneider BJ, Naidoo J, Santomasso BD, et al. Management of Immune-Related Adverse Events in Patients Treated With Immune Checkpoint Inhibitor Therapy: ASCO Guideline Update [Erratum in J Clin Oncol. 2022 Jan 20;40(3):315]. J Clin Oncol. 2021;39(36):4073-126.
14. Rashidi A, Shah C, Sekulic M. The Role of Kidney Biopsy in Immune Checkpoint Inhibitor-Associated AKI. Kidney360. 2022;3:530-3.
15. Bermejo S, Bolufer M, Riveiro-Barciela M, Soler MJ. Immunotherapy and the Spectrum of Kidney Disease: Should We Individualize the Treatment?. Front Med (Lausanne). 2022;9:906565.
16. Kitchlu A, Jhaveri KD, Sprangers B, Yanagita M, Wanchoo R. Immune checkpoint inhibitor use in patients with end-stage kidney disease: an analysis of reported cases and literature review. Clin Kidney J. 2021;14:2012-22.
17. Carroll RP, Boyer M, Gebski V, et al. Immune checkpoint inhibitors in kidney transplant recipients: a multicenter, single-arm, phase 1 study. Lancet Oncol. 2022;23:1078-86.
18. Lipson EJ, Bodell MA, Kraus ES, et al. Successful administration of ipilimumab to two kidney transplantation patients with metastatic melanoma. J Clin Oncol. 2014;32:e69-71.
19. G. Ishikawa, T. Sugiyama, T. Ito, et al. Renal allograft rejection after nivolumab therapy in patients with metastatic renal cell carcinoma. Int Cancer Conf J. 2021;10:116-118.
20. Lipson EJ, Bagnasco SM, Moore J Jr, et al. Tumor Regression and Allograft Rejection after Administration of Anti-PD-1. N Engl J Med. 2016;374:896-8.
21. Murakami N, Mulvaney P, Danesh M, et al; Immune Checkpoint Inhibitors in Solid Organ Transplant Consortium. A multicenter study on the safety and efficacy of immune checkpoint inhibitors in cancer patients with kidney transplant. Kidney Int. 2021;100:196-205.
22. Hirsch JS, Wanchoo R, Ng JH, et al. Use of Immune Checkpoint Inhibitors in Patients with End-Stage Kidney Disease, Single Center Experience and Review of the Literature. Kidney360. 2020;1:399-402.
23. Mejia CD, Frank AM, Singh P, et al. Immune checkpoint inhibitor therapy-associated graft intolerance syndrome in a failed kidney transplant recipient. Am J Transplant. 2021;21:1322-5.
24. Strohbehn IA, Lee M, Seethapathy H, et al. Safety and Efficacy of Immune Checkpoint Inhibitors in Patients on Dialysis: A Retrospective Case Series. Am J Kidney Dis. 2020;76:299-302.
25. Medina López RA, Rivero Belenchon I, Mazuecos-Quirós J, et al. Update on the treatment of metastatic renal cell carcinoma. World J Clin Oncol. 2022;13:1-8.
26. Borghaei H, Paz-Ares L, Horn L, et al. Nivolumab vs. docetaxel in patients with advanced nonsquamous nonsmall-cell lung cancer N Engl J Med. 2015;373:1627-39.
27. Kanz BA, Pollack MH, Johnpulle R, et al. Safety and efficacy of anti-PD-1 in patients with baseline cardiac, renal, or hepatic dysfunction. J Immunother Cancer. 2016;4:60.

Original Article

DOI: 10.4274/haseki.galenos.2024.10028
Med Bull Haseki 2024;62:235-242

The Impact of Immunoglobulin Replacement Therapy on Antibiotic Need in Adult Patients with Inborn Errors of Immunity

Esra Karabiber

Marmara University, Faculty of Medicine Turkey, Pendik Training and Research Hospital, Department of Chest Disease of Chest Disease, Division of Allergy and Immunology, Istanbul, Turkey

Abstract

Aim: Patients with inborn errors of immunity (IEI) have a higher frequency of infections and long-term antibiotic usage. We aimed to assess the effects of immunoglobulin replacement therapy (IgRT) on infection rates, antibiotic usage, and treatment outcomes in patients with IEI.

Methods: We retrospectively analyzed demographic data, infection frequency, antibiotic prescriptions, and IgRT in 122 IEI patients between March 2014 and September 2023. Specific IEI diagnoses were made following the European Society for Immunodeficiencies criteria.

Results: The median age of patients was 29 years [interquartile range (IQR): 23-40], with 54.1% being male. The median age at diagnosis was 25 years (IQR: 13-36), with a diagnostic delay of 96 months (IQR: 24-180). IgRT was administered to 74.5% of patients, with a median treatment duration of 20 years (IQR: 10-33.5). Antibiotic use was higher in patients receiving IgRT (median: 27, IQR: 16-42) compared to those not on IgRT (median: 14, IQR: 8-22; $p<0.001$). Patients with bronchiectasis had lower baseline immunoglobulin G, $CD19^+$, and natural killer cell counts, with more frequent antibiotic use, though hospitalization rates were similar to those without bronchiectasis. Immunoglobulin replacement therapy use was higher in the bronchiectasis group (61.5%, $p<0.001$). No significant differences in antibiotic use or hospitalization rates were observed between intravenous and subcutaneous IgRT groups.

Conclusion: Patients with IEI face significant respiratory infections despite IgRT and prophylactic antibiotics. Bronchiectasis is a key risk factor for increased antibiotic use. Early diagnosis and personalized treatment are crucial in reducing infection burden and improving outcomes in this population.

Keywords: Bronchiectasis, immunoglobulin replacement therapy, inborn errors of immunity, prophylactic antibiotics, respiratory infection, primary immunodeficiency

Introduction

Inborn errors of immunity (IEI), previously referred to as primary immunodeficiencies, are a diverse group of over 450 genetically determined disorders characterized by defects in various immune system components (1). These defects can compromise the immune system's ability to respond appropriately to pathogens, increasing susceptibility to infections, autoimmune diseases, and malignancies.

Patients with IEI often experience recurrent, severe, or unusual infections, which can significantly impact their quality of life and may lead to life-threatening complications if not managed effectively (2). IEI lead to specific infection susceptibilities. For example, individuals with humoral immune deficiencies are more prone to infections by encapsulated bacteria due to a lack of antibody defense, though they can still combat intracellular infections. In contrast, those with combined immunodeficiency (CID) are highly susceptible to opportunistic pathogens, including viruses, mycobacteria, protozoa, and fungi, because of T-cell deficiencies. Patients with chronic granulomatous disease have impaired phagocyte function, making them particularly vulnerable to infections from mycobacteria,

Address for Correspondence: Esra Karabiber, Marmara University, Faculty of Medicine Turkey, Pendik Training and Research Hospital, Department of Chest Disease of Chest Disease, Division of Allergy and Immunology, Istanbul, Turkey
Phone: +90 505 687 02 11 **E-mail:** dresrabulut@hotmail.com **ORCID:** orcid.org/0000-0002-8377-7637
Received: 19.08.2024 **Accepted:** 01.10.2024

fungi, and certain bacteria like *Staphylococcus aureus* and *Escherichia coli*.

Respiratory infections are common and often the first sign of IEI, leading to significant hospitalizations and fatalities in affected individuals. Reducing the infection burden is crucial for improving life expectancy. Encapsulated bacteria are frequently the cause, though viral infections are also common (3,4). Preventive strategies typically include immunoglobulin replacement therapy (IgRT), facial masks, social distancing, and prophylactic antibiotics (5-7). The use of intravenous immunoglobulin (IVIG) in these patients has provided clear benefits by significantly reducing acute and chronic infections. Immunoglobulin replacement therapy should be initiated in all phenotypes of severe CID (8). However, despite these interventions, many IEI patients continue to experience respiratory infections, increasing their risk of developing bronchiectasis (9-11).

The study hypothesized that IgRT would impact the frequency of prescribed antibiotics, the rate of infections, and the hospitalization rates due to infections among patients with IEI. By examining a cohort of IEI patients, we seek to provide a comprehensive overview of current clinical practices and their outcomes.

Methods

This retrospective observational cohort involved reviewing medical records of IEI patients who were under the care of the allergy immunology clinic at a tertiary hospital in Istanbul.

The inclusion criteria for this study were as follows: i) a confirmed diagnosis of IEI and ii) age over 18 years. Patients with secondary immunodeficiency were excluded from the study. Specific IEI diagnoses were made following the European Society for Immunodeficiencies criteria (12). The classification of patients with IEI was made based on the International Union of Immunological Societies (IUIS) and Middle East, North Africa, and Turkey (6,13).

Data Collection

Data were collected retrospectively from the medical records of the included patients between 2014 and 2023. The following variables were extracted: Demographic characteristics, including age, gender, and relevant family history; annual antibiotic prescriptions, including prophylactic and therapeutic use, were recorded. Over the study period, the number of hospitalizations for infection and without infection was documented. Comorbidities, such as autoimmune diseases, chronic lung disease, or malignancies, were gathered. A comprehensive list of medications, including those used for managing IEI, and details on IVIG and subcutaneous immunoglobulin (SCIG) therapies. Laboratory results, including complete blood counts, baseline immunoglobulin levels, and other relevant immunological parameters, such as lymphocyte subsets, were collected. Radiological findings, including Thorax computerized tomography performed during the study period, were reviewed to assess bronchiectasis or structural abnormalities related to immunodeficiency.

Ethical Considerations

The study was approved by the Ethics Committee of Marmara University, Faculty of Medicine (date: 10/7/22, approval no.: 1354). It was conducted in accordance with the Declaration of Helsinki. Informed consent was obtained from all patients or their legal guardians, where applicable, for using their medical records for research purposes.

Statistical Analysis

All data was analyzed using the SPSS statistical software package version 22 (IBM Corp., USA) and GraphPad Prism 8 (GraphPad software California, USA). Median and interquartile range (IQR) values for continuous variables and the frequency and percentage for the categorical variables were calculated. Differences between ordinal data were evaluated with the Mann-Whitney U test and the Kruskal-Wallis test. Categorical variables were evaluated with the 2-tailed chi-square or Fisher's exact tests. A p-value of less than 0.05 was considered the significance level for differences.

Results

The study included 122 patients with IEI at a tertiary care hospital in Istanbul. The median age of all patients was 29 years (IQR: 23-40), and 54.1% (n=66) were male. The median age at diagnosis was 25 years (IQR: 13-36). The median diagnostic delay was 96 months (IQR: 24-180), and the median duration of IgRT was 20 years (IQR: 10-33.5). Table 1 summarizes the patients' demographic characteristics and laboratory results. Figure 1 displays the classification of diagnosis based on IUIS criteria.

A history of pneumonia was present in 84 patients (68.8%), sinusitis in 64 patients (52.4%), and otitis in 46 patients (37.7%). Immunoglobulin replacement therapy was administered to 91 patients (74.5%), and prophylactic antibiotic therapy was given to 54 patients (44.2%). A total of 64 (70.3%) patients received IVIG therapy, while 27 (29.6%) patients were on subcutaneous IgRT. The median frequency of outpatient prescribed antibiotics was 22 (IQR: 12-40). The median number of hospitalizations due to infections was 0 (IQR: 0-1). Figure 2 shows the treatment algorithm of all study groups.

The frequency of antibiotic use was higher in patients receiving IgRT, with a median of 27 (IQR: 16-42), compared to those not receiving IgRT [14 (IQR: 8-22), $p<0.001$]. There was no significant difference

between the two groups regarding hospitalization due to infections, non-infection-related hospitalizations, or total hospitalizations. $CD19^+$ and natural killer (NK) cell counts were significantly lower in the group receiving IgRT [median for $CD19^+$: 102 (IQR: 35.5-234) vs. 194 (IQR: 155-303), p=0.004; median for NK: 81 (IQR: 37.5-160) vs. 195 (IQR: 69.7-297), p=0.006]. Table 2 summarizes the demographic and laboratory characteristics of the groups receiving and not receiving IgRT. When comparing the IVIG and SCIG groups, there was no difference in hospitalizations due to infections, non-infection-related hospitalizations, or total hospitalizations. The frequency of antibiotic use did not differ significantly between the IVIG and SCIG groups [median: 24, (IQR: 13.2-42.7) vs. 33 (IQR: 21-42); p=0.12].

The age of hospital admission for patients with bronchiectasis was significantly lower than those without bronchiectasis [median years: 18 (IQR: 7-32.5) vs. 27 (IQR: 18.7-38.2), p=0.004]. The presence of bronchiectasis was statistically significant regarding the frequency of antibiotic use. The median frequency of antibiotic use was 33.5 (IQR: 21-57.2) in the bronchiectasis group, compared to 15.5 (IQR: 9-24) in the non-bronchiectasis group (p<0.001). There was no significant difference in

Table 1. Demographic features of all study populations (n=122)	
Gender; male (%)	66 (54.09)
Current age, year (median, IQR)	29 (23-40)
Age at diagnosis (median year, IQR)	25 (13-36)
Age at admission to immunology (median year, IQR)	24 (10-37)
Symptoms onset-age (median year, IQR)	9.5 (3-25)
Diagnostic delay (median months, IQR)	96 (24-180)
Treatment	**n (%)**
IgRT -Intravenous route -Subcutaneous route No IgRT -Prophylactic antibiotic	91 (74.5) 64 (70.3) 27 (29.7) 31 (25.4) 54 (44.2)
Number of prescribed antibiotics (median, IQR)	22 (12-40)
Hospitalization, n, IQR -Infection-related -Non-infection related -Total	 0 (0-1) 0 (0-0.25) 0 (0-1)
Lung screening	**n (%)**
-Bronchiectasis, n (%) -Number of affected lobes, mean±SD	60 (49.1) 1.47±1.67
Types of bronchiectasis	
-Tubular -Cystic	36 (60) 24 (40)
Complete blood count	**Median (IQR)**
Leucocytes ($x10^3$/mL) Lymphocytes ($x10^3$/mL) Hemoglobin ($x10^3$/mL) Granulocytes ($x10^3$/mL) Eosinophils ($x10^3$/mL) Monocytes ($x10^3$/mL) Platelets ($x10^3$/mL)	6550 (5200-8625) 1800 (1200-2600) 13.2 (11.5-14.5) 3970 (2900-5225) 30 (10-72.5) 500 (400-700) 242 (180-299)
Serum immunoglobulins	**Median (IQR)**
IgG (baseline) (mg/dL) IgG (trough) (mg/dL) IgA (baseline) (mg/dL) IgM (baseline) (mg/dL) IgE (baseline)	486 (277-1007) 989 (837-1331) 18 (5-80) 40 (13-130) 1 (0.2-14)
Lymphocyte subsets (absolute count)	**Median (IQR)**
$CD3^+$ $CD4^+$ $CD8^+$ $CD19^+$ NK	1324 (862-1915) 642 (413-906) 599 (409-867) 141 (51-266) 91 (41-189)

Values are presented as median [IQR], and categorical variables are presented as n (%). Differences between groups were evaluated using chi-square or Fisher's exact test, as appropriate.
IQR: Interquartile range, NK: Natural killer, IgRT: Immunoglobulin replacement therapy, SD: Standard deviation, Ig: Immunoglobulin

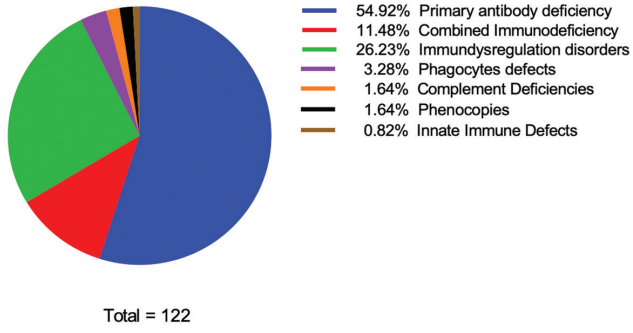


Figure 1. Classification of diagnoses according to IUIS criteria
IUIS: International Union of Immunological Societies

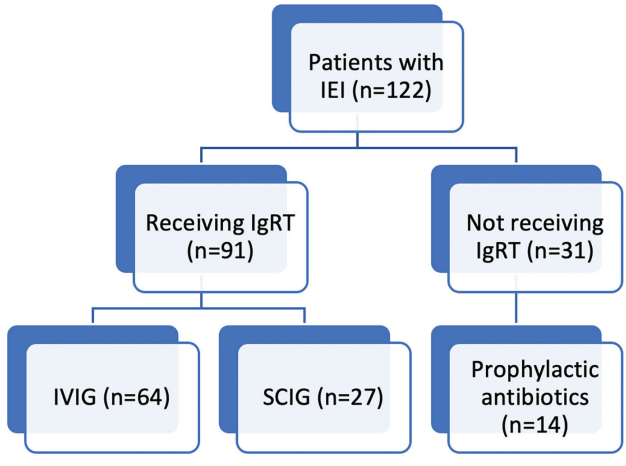


Figure 2. Treatment algorithm for patients with inborn errors of immunity
IEI: Inborn errors of immunity, IgRT: Immunoglobulin replacement therapy, IVIG: Intravenous immunoglobulin, SCIG: Subcutaneous immunoglobulin

the frequency of hospitalizations due to infections, non-infection-related hospitalizations, or total hospitalizations between those with and without bronchiectasis. Baseline immunoglobulin G (IgG) levels were significantly lower in patients with bronchiectasis compared to those without [median: 440 (IQR: 145-788) vs. 604 (IQR: 347-1090), p=0.03]. $CD19^+$ and NK cell counts were also significantly lower in the bronchiectasis group compared to the non-bronchiectasis group [median $CD19^+$: 102 (IQR: 33-218) vs. 168 (IQR: 63.7-294), p=0.04; median NK: 72.5 (IQR: 34.3-159) vs. 119.5 (IQR:58-235), p=0.02]. The rate of IgRT was higher in the bronchiectasis group (56 patients, 61.5%; p<0.001). Table 3 summarizes the demographic and laboratory characteristics of patients with and without bronchiectasis.

When comparing patients on prophylactic antibiotics to those not receiving them, the frequency of antibiotic use was notably higher among those on prophylactic therapy [median: 33.5 (IQR: 21-46) vs median: 14 (IQR: 8-29), p<0.001]. However, the two groups observed no significant differences in other clinical parameters.

Over a 9-year follow-up of 122 patients, a total of 3,606 infection episodes were documented, resulting in an antibiotic treatment rate of 3.28 infections per patient-year.

In the past five years, 12 (9.8%) patients experienced 18 infection episodes requiring hospitalization while on IgRT. Among these, 12 episodes were due to pneumonia, and 2 were caused by pyelonephritis. Additionally, 4 patients received outpatient treatment for pneumonia. None of the patients had low IgG trough levels.

Table 2. Clinical and immunological features of patients based on the IgRT

	With IgRT, n=91	**Without IgRT, n=31**	**p-value**
Gender, male (%)	48 (52.7)	19 (61.2)	0.67
Current age (years, median, IQR)	29 (23-42)	28 (23-38)	0.59
Age at admission (years, median, IQR)	20 (7-37)	25 (21-37)	0.18
Age at symptom onset (years, median, IQR)	7 (3-24)	12 (7-29)	0.08
Age at diagnosis (years, median, IQR)	24 (11-36)	25 (22-36)	0.1
Age at IgRT (years, median, IQR)	20 (10-34.5)	23 (10-30)	0.8
Treatment n, (%) **Prophylactic antibiotic**	40 (74)	14 (26)	1
Number of prescribed antibiotics (median, IQR)	27 (16-42)	14 (8-22)	<0.001
Hospitalization, n, IQR			
Infection-related Non-infection related Total	0 (0-1) 0 (0-1) 0 (0-2)	0 (0-0) 0 (0-0) 0 (0-1)	0.09 0.7 0.15
Complete blood count			
Leucocytes ($x10^3$/mL) Lymphocytes ($x10^3$/mL) Hemoglobin ($x10^3$/mL) Granulocytes ($x10^3$/mL) Eosinophils ($x10^3$/mL) Monocytes ($x10^3$/mL) Platelets ($x10^3$/mL)	6500 (5000-8600) 1700 (1000-2700) 13 (10.8-14) 3800 (2720-5300) 23 (60-70) 500 (400-700) 234 (164-295)	6700 (5900-8700) 2000 (1500-2400) 13.9 (12-15.2) 4200 (3200-5100) 20 (50-93) 500 (400-600) 254 (214-309)	0.4 0.28 0.008 0.4 0.01 0.84 0.17
Serum immunoglobulins, median (IQR)			
IgG (baseline) (mg/dL) IgG (trough) (mg/dL) IgA (baseline) (mg/dL) IgM (baseline) (mg/dL) IgE (baseline) (mg/dL)	395 (219-635) 989 (847-1371) 10 (4-45) 25 (10-80.5) 0.5 (0.2-5.2)	981 (484-1176) 1000 (615-1322) 39 (5-253) 92 (26-174) 6.5 (0.4-70)	<0.001 0.4 0.01 0.002 0.01
Lymphocyte subsets (absolute count, median, IQR)			
$CD3^+$ $CD4^+$ $CD8^+$ $CD19^+$ NK	1287 (800-1946) 619 (355-900) 599 (413-914) 102 (35-234) 81 (37-160)	1458 (997-1897) 683 (479-913) 592 (401-802) 194 (155-303) 195 (69-297)	0.6 0.5 0.6 0.004 0.006

Comparisons between groups were performed using the Mann-Whitney U test for non-normally distributed data. Differences between groups were evaluated using chi-square or Fisher's exact test, as appropriate.
IQR: Interquartile range, NK: Natural killer, IgRT: Immunoglobulin replacement therapy, Ig: Immunoglobulin

Table 3. Clinical and immunological features of patients based on the presence of bronchiectasis			
	With bronchiectasis	**Without bronchiectasis**	**p-value**
Gender, male (%)	36 (60)	30 (48.3)	0.2
Current age (years, median, IQR)	28.5 (22.2-40.7)	29 (23-40)	0.88
Age at admission (years, median, IQR)	18 (7-32.5)	27 (18.7-38.2)	0.004
Age at symptom onset (years, median, IQR)	6 (2.2-17.7)	15.5 (7-29)	<0.001
Age at diagnosis (years, median, IQR)	21 (9-32)	27.5 (19-38)	0.01
Age at IgRT (years, median, IQR)	20 (8-33)	27.5 (17-36)	0.06
Treatment n, (%) IgRT -Intravenous route -Subcutaneous route No IgRT -Prophylactic antibiotic	 56 (61.5) 34 (60.7) 22 (39.2) 4 (6.6) 30 (50)	 35 (38.4) 30 (85.7) 5 (14.2) 27 (43.5) 24 (38.7)	 <0.001 0.2
Number of prescribed antibiotics (median, IQR)	33.5 (21-57.2)	15.5 (9-24)	<0.001
Hospitalization, n, (IQR) Infection-related Non-infection related Total	 0 (0-1) 0 (0-1) 0 (0-2)	 0 (0-0) 0 (0-0) 0 (0-1)	 0.6 0.6 0.3
Complete blood count			
Leucocytes (x10^3/mL) Lymphocytes (x10^3/mL) Hemoglobin (x10^3/mL) Granulocytes (x10^3/mL) Eosinophils (x10^3/mL) Monocytes (x10^3/mL) Platelets (x10^3/mL)	6435 (5125-8975) 1950 (1025-2600) 12.8 (10.6-14.4) 3870 (2785-5475) 22 (4-70) 500 (400-700) 224 (152-305)	6650 (5350-8425) 1700 (1300-2550) 13.3 (11.8-14.7) 4050 (2975-4950) 42 (16-90) 500 (400-600) 251 (210-299)	0.8 0.9 0.1 0.8 0.03 1 0.2
Serum immunoglobulins, median (IQR)			
IgG (baseline) (mg/dL) IgG (trough) (mg/dL) IgA (baseline) (mg/dL) IgM (baseline) (mg/dL) IgE (baseline) (mg/dL)	440 (145-788) 982 (840-1305) 5.5 (4-43) 25 (10-124) 0.2 (0.2-4.2)	604 (347-1090) 1026 (803-1429) 27 (7-124) 47 (18-132) 2.9 (0.2-27)	0.03 0.8 0.01 0.2 0.01
Lymphocyte subsets (absolute count, median, IQR)			
$CD3^+$ $CD4^+$ $CD8^+$ $CD19^+$ NK	1354 (847-2111) 589 (333-906) 647(440-972) 102 (33-218) 72 (34-159)	1299 (918-1897) 713 (490-903) 560 (387-801) 168 (64-294) 119 (58-235)	0.6 0.2 0.09 0.04 0.02
Comparisons between groups were performed using the Mann-Whitney U test for non-normally distributed data. IQR: Interquartile range, NK: Natural killer, IgRT: Immunoglobulin replacement therapy, Ig: Immunoglobulin			

Discussion

Based on our findings among IEI patients, the frequency of prescribed antibiotics was higher in the IgRT group compared to the non-IgRT group. Additionally, the frequency of antibiotic use was higher in patients with bronchiectasis than those without bronchiectasis. Within the IgRT group, no significant differences were observed regarding hospitalization rates and antibiotic usage frequency between patients receiving IVIG and those receiving SCIG.

One of the key observations from our study is the substantial diagnostic delay, with a median of 8 years from the onset of symptoms to diagnosis. This delay is concerning, as earlier diagnosis and intervention could potentially mitigate some of the severe complications associated with IEI, such as recurrent infections and the development of bronchiectasis. In a study conducted with common variable immune deficiency (CVID) cases, patients with bronchiectasis had significantly lower trough IgG levels and efficacy and required a longer time to achieve target IgG levels than those without bronchiectasis. This delay was significantly associated with an increased frequency of infections. The presence of bronchiectasis was significantly associated with a prolonged time to reach target IgG levels. These long-term differences between

patients with and without bronchiectasis have important clinical implications (14). According to our findings, patients with bronchiectasis required more frequent antibiotic use despite maintaining adequate trough IgG levels. This result aligns with the existing literature.

The median age at diagnosis (25 years) suggests that many patients endure years of unmanaged symptoms before receiving appropriate care. This underscores the need for increased awareness and early screening, particularly in high-risk populations.

One of the mainstays of treatment for many IEI patients is IgRT. This therapy helps to restore some of the immune functions by supplying patients with the necessary antibodies to fight infections and modulate immune responses (5,15-17). Immunoglobulin replacement therapy has been shown to reduce the frequency and severity of infections in patients with antibody deficiencies and other forms of IEI (18).

Respiratory infections emerged as a predominant issue, with a high prevalence of pneumonia, sinusitis, and otitis among the patients. The fact that 68.8% of patients had a history of pneumonia indicates the vulnerability of IEI patients to severe respiratory infections, which is consistent with previous studies. Despite the widespread use of IgRT and prophylactic antibiotics, many patients continue to experience frequent infections, suggesting that current preventive strategies may not be entirely adequate for all individuals. This persistent risk of infections emphasizes the importance of personalized treatment approaches, which may include more aggressive prophylactic measures or adjustments in IgRT dosing.

Despite adequate IgRT, recurrent respiratory tract infections remain the most common clinical manifestation of CVID (19), often leading to the development of progressive bronchiectasis (9,10,20). While these infections were traditionally attributed mainly to encapsulated bacteria, recent studies also suggest a significant role for viral infections (4,21,22). Although respiratory tract infections are prevalent and severely impact the quality of life in primary antibody deficiency syndromes, the specific nature of the symptoms during these episodes is not well understood (23). Patients frequently receive antibiotics to manage respiratory infections, both as "rescue" treatments for acute episodes and as prophylactic measures to reduce infection frequency (24). However, the specific symptomatic triggers for initiating antibiotics and the clinical outcomes of these interventions remain unclear.

In a study involving 278 participants, despite receiving adequate IgRT, 6.9% continued to experience severe or very severe infections. Additionally, a substantial proportion of participants (84.9%) reported that infections imposed significant limitations on their daily lives. Notably, 18.3% of the participants who were dissatisfied with their treatment demonstrated a higher disease burden, characterized by more frequent non-routine healthcare visits, increased antibiotic use, and more days missed from school, work, or other responsibilities (25). In this study, the severe infection rate among patients with IgRT was 9.8%, similar to other studies. Our data also show that patients receiving IgRT had a higher frequency of antibiotic use than those not on IgRT. This finding could reflect a more severe clinical phenotype in the IgRT group, necessitating increased antibiotic prophylaxis to prevent infections. Interestingly, while IgRT was associated with lower $CD19^+$ and NK cell counts, it did not significantly impact hospitalization rates for infections or other causes. This suggests that while IgRT effectively reduces the severity of infections, it may not completely eliminate the need for hospital care, especially in patients with more severe immune deficiencies.

Despite these interventions, a significant number of IEI patients continue to experience respiratory infections, which heightens the risk of developing bronchiectasis (11). In a cohort study, patients with CVID observed over an average follow-up period of 11 years, 34.2% had chronic lung disease at the time of diagnosis, and this percentage increased to 46.3% during the follow-up, even with the administration of IgRT (26). Bronchiectasis, a common complication in IEI patients, was associated with more frequent antibiotic use and lower baseline IgG, $CD19^+$, and NK cell counts (5). These findings are in line with the understanding that bronchiectasis often results from chronic and recurrent infections, which may further impair immune function. The lower IgG levels in patients with bronchiectasis underscore the need for careful monitoring and potential adjustments in IgRT to ensure adequate immune protection.

The study also highlights the significant impact of bronchiectasis on the clinical course of IEI. Patients with bronchiectasis had an earlier age of hospital admission and a higher rate of IgRT use, indicating a more severe disease trajectory. This underscores the importance of early detection and management of bronchiectasis in IEI patients, as it can profoundly affect their quality of life and long-term prognosis.

Prophylactic antibiotic therapy is often employed to prevent bacterial infections, which are a common complication in these patients due to their impaired immune systems (20). In our study, patients on prophylactic antibiotics have more frequent antibiotic use despite receiving IgRT.

Study Limitations

Our study has several limitations. First, due to its retrospective design, we were unable to obtain detailed information about the patient's presenting symptoms.

Second, some of the antibiotics prescribed to these patients may have been unnecessary, as they could have been given during viral infections. However, the frequency of antibiotic prescriptions in these patients indicates frequent hospital visits. Despite these limitations, this study provides valuable insights into the impact of IgRT on infection frequency, antibiotic use, and hospitalization rates in severe IEI patients, offering a comprehensive view of clinical practices and outcomes.

Conclusion

Our study reinforces the critical role of early diagnosis, personalized treatment strategies, and vigilant monitoring in managing IEI patients. The high prevalence of respiratory infections and the development of bronchiectasis among these patients indicate the need for ongoing research to optimize preventive and therapeutic interventions. Future studies should focus on identifying biomarkers to predict which patients are at the highest risk for complications like bronchiectasis and tailoring treatment protocols accordingly to improve outcomes. This study provides valuable insights into the clinical efficacy of IgRT profiles in patients with IEI. The findings highlight the significant challenges these patients face, particularly in terms of diagnostic delays, infection-related complications, and the long-term management of their condition. This research will contribute to a better understanding of the challenges in managing these complex conditions and may offer insights into optimizing treatment strategies to improve patient outcomes.

Footnote

Ethics Committee Approval: The study was approved by the Ethics Committee of Marmara University, Faculty of Medicine (date: 10/7/22, approval no.: 1354). It was conducted in accordance with the Declaration of Helsinki.

Informed Consent: Informed consent was obtained from the patients participating in the study.

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

References

1. Tangye SG, Al-Herz W, Bousfiha A, et al. Human Inborn Errors of Immunity: 2022 Update on the Classification from the International Union of Immunological Societies Expert Committee. J Clin Immunol. 2022;42:1473-507.
2. Picard C, Bobby Gaspar H, Al-Herz W, et al. International Union of Immunological Societies: 2017 Primary Immunodeficiency Diseases Committee Report on Inborn Errors of Immunity. J Clin Immunol. 2018;38:96-128.
3. Kainulainen L, Vuorinen T, Rantakokko-Jalava K, Osterback R, Ruuskanen O. Recurrent and persistent respiratory tract viral infections in patients with primary hypogammaglobulinemia. J Allergy Clin Immunol. 2010;126:120-6.
4. Oksenhendler E, Gérard L, Fieschi C, et al. Infections in 252 Patients with Common Variable Immunodeficiency. Clin Infect Dis. 2008;46:1547-54.
5. Baris S, Ercan H, Cagan HH, et al. Efficacy of intravenous immunoglobulin treatment in children with common variable immunodeficiency. J Investig Allergol Clin Immunol. 2011;21:514-21.
6. Baris S, Abolhassani H, Massaad MJ, et al. The Middle East and North Africa Diagnosis and Management Guidelines for Inborn Errors of Immunity. J Allergy Clin Immunol Pract. 2023;11:158-80.
7. Bonilla FA, Bernstein IL, Khan DA, et al. Practice parameter for the diagnosis and management of primary immunodeficiency. Ann Allergy Asthma Immunol. 2005;94(5 Suppl 1):1-63.
8. Aydiner E. Primer İmmün Yetmezliklerde İmmünoglobulin Replasman Tedavisi:Güncel Durum Raporu 2019. Ankara, Bulut Yayınevi 2019.
9. Yong PF, Thaventhiran JE, Grimbacher B. "A rose is a rose is a rose," but CVID is Not a common variable immune deficiency (CVID); what do we know in 2011? Adv Immunol. 2011;111:47-107.
10. Verma N, Grimbacher B, Hurst JR. Lung disease in primary antibody deficiency. Lancet Respir Med. 2015;3:651-60.
11. Ozen A, Baris S, Karakoc-Aydiner E, Ozdemir C, Bahceciler NN, Barlan IB. Outcome of hypogammaglobulinemia in children: immunoglobulin levels as predictors. Clin Immunol. 2010;137:374-83.
12. Seidel MG, Kindle G, Gathmann B, et al. The European Society for Immunodeficiencies (ESID) Registry Working Definitions for the Clinical Diagnosis of Inborn Errors of Immunity. J Allergy Clin Immunol Pract. 2019;7:1763-70.
13. Tangye SG, Al-Herz W, Bousfiha A, et al. Human Inborn Errors of Immunity: 2019 Update on the Classification from the International Union of Immunological Societies Expert Committee. J Clin Immunol. 2020;40:24-64.
14. Onalan T, Colkesen F, Kilinc M, et al. Relationships between bronchiectasis and time to achieving target trough immunoglobulin G levels in patients with common variable immunodeficiency. Allergy Asthma Proc. 2024;45:180-5.
15. Yalcin Gungoren E, Yorgun Altunbas M, Dikici U, et al. Insights into Patient Experiences with Facilitated Subcutaneous Immunoglobulin Therapy in Primary Immune Deficiency: A Prospective Observational Cohort. J Clin Immunol. 2024;44:169.
16. Karakoç Aydıner E, Kıykım A, Barış S, Özen A, Barlan I. Use of subcutaneous immunoglobulin in primary immune deficiencies. Turk Pediatri Ars. 2016;51:8-14.
17. Orange JS, Hossny EM, Weiler CR, et al. Use of intravenous immunoglobulin in human disease: a review of evidence by members of the Primary Immunodeficiency Committee of the American Academy of Allergy, Asthma and Immunology. J Allergy Clin Immunol. 2006;117(4 Suppl). :S525-53.

18. Perez EE, Orange JS, Bonilla F, et al. Update on the use of immunoglobulin in human disease: A review of evidence. J Allergy Clin Immunol. 2017;139(3 Suppl):1-46.
19. Gathmann B, Mahlaoui N, CEREDIH et al. Clinical picture and treatment of 2212 patients with common variable immunodeficiency. J Allergy Clin Immunol. 2014;134:116-26.
20. Quinti I, Soresina A, Guerra A, et al. Effectiveness of immunoglobulin replacement therapy on clinical outcome in patients with primary antibody deficiencies: results from a multicenter prospective cohort study. J Clin Immunol. 2011;31:315-22.
21. Kainulainen L, Vuorinen T, Rantakokko-Jalava K, Osterback R, Ruuskanen O. Recurrent and persistent respiratory tract viral infections in patients with primary hypogammaglobulinemia. J Allergy Clin Immunol 2010;126:120-6.
22. Duraisingham SS, Manson A, Grigoriadou S, Buckland M, Tong CY, Longhurst HJ. Immune deficiency: changing spectrum of pathogens. Clin Exp Immunol. 2015;181:267-74.
23. Hurst JR, Workman S, Garcha DS, Seneviratne SL, Haddock JA, Grimbacher B. Activity, severity and impact of respiratory disease in primary antibody deficiency syndromes. J Clin Immunol. 2014;34:68-75.
24. Sperlich JM, Grimbacher B, Workman S, et al. Respiratory Infections and Antibiotic Usage in Common Variable Immunodeficiency. J Allergy Clin Immunol Pract. 2018;6:159-68.
25. Hodkinson JP, Griffiths PR, Narme AMI, Staiger C. UK PID Patients: Overview of perception on IgRT and infections using short survey. Pharmazie. 2023;78:231-7.
26. Quinti I, Soresina A, Spadaro G, et al. Long-term follow-up and outcome of a large cohort of patients with common variable immunodeficiency. J Clin Immunol. 2007;27:308-16.

Original Article

DOI: 10.4274/haseki.galenos.2024.10038
Med Bull Haseki 2024;62:243-249

Evaluation of ChatGPT's Performance in the Turkish Board of Orthopaedic Surgery Examination

Ahmet Yigitbay

Siverek State Hospital, Clinic of Orthopedics and Traumatology, Sanliurfa, Turkey

Abstract

Aim: Technological advances lead to significant changes in education and evaluation processes in medicine. In particular, artificial intelligence and natural language processing developments offer new opportunities in the health sector. This article evaluates Chat Generative Pre-Trained Transformer's (ChatGPT) performance in the Turkish Orthopaedics and Traumatology Education Council (TOTEK) Qualifying Written Examination and its applicability.

Methods: To evaluate ChatGPT's performance, TOTEK Qualifying Written Examination questions from the last five years were entered as data. The results of ChatGPT were assessed under four parameters and compared with the actual exam results. The results were analyzed statistically.

Results: Of the 500 questions, 458 were used as data in this study. Chat Generative Pre-Trained Transformer scored 40.2%, 26.3%, 37.3%, 32.9%, and 35.8% in the 2019, 2020, 2021, 2022, and 2023 TOTEK Qualifying Written Examination, respectively. When the correct answer percentages of ChatGPT according to years and the simple linear regression model applied to these data were analyzed, it was determined that there was a slightly decreasing trend in the correct answer rates as the years progressed. ChatGPT's TOTEK Qualifying Written Examination performance showed a statistically significant difference from the actual exam results. It was observed that the correct answer percentage of ChatGPT was below the general average success scores of the exam for each year.

Conclusions: This analysis of artificial intelligence's applicability in the field and its role in training processes is essential to assess ChatGPT's potential uses and limitations. Chat Generative Pre-Trained Transformer can be a training tool, especially for knowledge-based and logical questions on specific topics. Still, its current performance is not at a level that can replace human decision-making in specialized medical fields.

Keywords: Artificial intelligence, humans, orthopedics, specialty boards

Introduction

Artificial intelligence (AI) and natural language processing (NLP) technologies drive significant transformations across many fields. The use of these technologies in the healthcare sector has been impactful across a broad spectrum, from medical education to patient care. The primary purpose of AI is to improve patient experience, enhance the reliability of clinicians, and provide more information for the clinical decision-making process. Instead of replacing healthcare workers, these goals aim to enhance their experience (1-4). Language modeling systems like Chat Generative Pre-Trained Transformer (ChatGPT) support health professionals in various areas, from education to clinical applications. Chat Generative Pre-Trained Transformer, developed by OpenAI, is an AI model incorporating several language modeling and comprehension techniques, allowing users to communicate in their native language (5,6).

Technological advancements cause significant changes in education and assessment processes within medicine. Developments in AI and NLP, in particular, are introducing new possibilities in the healthcare sector (7-9). In this context, large language models like ChatGPT can play a significant role in medical education and exam evaluations.

This article aims to assess the performance of ChatGPT in the Turkish Orthopaedics and Traumatology Education Council (TOTEK) Qualification Written Exam and its applicability in the field. The role of ChatGPT in evaluating knowledge and skills in this area and its advantages and

Address for Correspondence: Ahmet Yigitbay, Siverek State Hospital, Clinic of Orthopedics and Traumatology, Sanliurfa, Turkey
Phone: +90 543 639 61 62 **E-mail:** ahmetyigitbay@gmail.com **ORCID:** orcid.org/0000-0002-7845-1974
Received: 28.08.2024 **Accepted:** 30.10.2024

disadvantages compared to traditional exam formats were examined. Additionally, based on practical field experience and feedback from medical experts, the usability of ChatGPT in orthopedics and traumatology education and practice was evaluated. The qualification written exam has been conducted by TOTEK under the Turkish Orthopaedics and Traumatology Association (TOTBID) since 2003, in two stages. The first stage consists of a written exam with "objective structured multiple-choice questions", and the second stage consists of an oral exam that includes an objective structured clinical examination (10,11).

This study is conducted to understand the potential of AI-supported language models in medical education and evaluation processes and to provide a framework for future directions. In the following sections of the article, the performance and applicability of ChatGPT will be analyzed in depth.

Methods

The primary objective of this study is to evaluate ChatGPT's performance in the TOTEK Qualification Written Exam. To this end, ChatGPT's ability to solve exam questions has been compared with the exam performance of physicians who have previously taken the exam.

Data Collection

Data from the last five years of TOTEK exam questions was used. Each year's questions were asked individually to ChatGPT, and the answers provided were recorded. To evaluate the performance of real physicians, the average exam results of physicians who had previously taken the exam over the past five years (2019-2023) were used. These data were obtained from the "period books" published by TOTBID, which contain past exam results (12-14).

Performance Evaluation

Chat Generative Pre-Trained Transformer's performance was measured by its ability to solve questions in the dataset. Chat Generative Pre-Trained Transformer's accuracy rate was used to determine how correctly it answered the exam questions. Only multiple-choice questions containing text were included in the assessment. Due to limitations, questions containing images, tables, figures, and canceled questions were excluded from the evaluation. Only first-stage exam questions were included in the assessment. Questions were asked to ChatGPT only once, and the responses were recorded.

The physicians' performance was determined by taking the average results of physicians who had previously taken the exam.

Comparison

The performance of ChatGPT and the performance of physicians who have taken the exam have been compared. Differences between the two groups were statistically analyzed, and the results were compared.

Assessment

The results obtained have been used to compare ChatGPT's performance in the TOTEK Qualification Written Exam with physicians' performance. These inferences have been evaluated to provide information about ChatGPT's applicability and potential in the field. Chat Generative Pre-Trained Transformer-4's version was used in all parts of this project.

The responses given by ChatGPT were evaluated under four categories, and these variables were statistically analyzed.

1: Correct answer, consistent logic
2: Correct answer, inconsistent logic
3: Incorrect answer, consistent logic
4: Incorrect answer, inconsistent logic

Ethical Considerations

All data used in the study were anonymized to avoid including personal information. The research was designed and conducted according to ethical rules. Chat Generative Pre-Trained Transformer was included in the educational process under licenses suitable for open-source and commercial use. Written permission was obtained from the TOTBID board of directors for this study (document no.: 159, dated: 26.04.2024).

Results

Of the 500 questions, 458 were used as data in this study. Since two questions were canceled according to the answer keys, these questions were excluded from the study. Chat Generative Pre-Trained Transformer scored 40.2%, 26.3%, 37.3%, 32.9%, and 35.8% in the 2019, 2020, 2021, 2022, and 2023 TOTEK Qualifying Written Examination, respectively (Table 1). 47.2% of the candidates were successful in the TOTEK qualifying exam held in 2023, 37% of the candidates in the exam held in 2022, 55.3% of the candidates in the exam held in 2021, 46.4% of the candidates in the exam held in 2020, and

Table 1. ChatGPT exam results by year

Year	Correct answer percentage %	Wrong answer percentage %
2019	41.05	58.95
2020	26.32	73.68
2021	37.35	62.65
2020	32.9	67.03
2023	36.96	63.04

ChatGPT: Chat Generative Pre-Trained Transformer

70.5% of the candidates in the exam held in 2019. In the TOTEK qualifying exam, the average success score for 2019 was 60, the average success score for 2020 was 55, the average success score for 2021 was 60, the average success score for 2022 was 60, and finally, the average success score for 2023 was 60 (12-14). Figure 1 presents a comparison between ChatGPT and real exam results.

After analyzing the numerical analysis of the answers given by ChatGPT over the years, the number of answers with correct and consistent logic has remained relatively constant. There are very few answers under the category of the correct answer, and inconsistent logic shows that ChatGPT generally gives logical answers. Although there is an increase in incorrect and consistent logic answers in 2022, the number of ChatGPT's answers with incorrect but consistent logic varies. Wrong answer, inconsistent logic: The area in which ChatGPT struggled the most was the answers with incorrect and inconsistent logic. Especially in 2020, there was a significant increase in the number of such answers (Table 2, Figure 2).

When the percentages of correct answers given by ChatGPT are analyzed by year, fluctuations are observed in its performance over time. The highest percentage of correct answers was achieved in 2019, while the lowest was recorded in 2020. Although ChatGPT's annual performance shows some variation, these fluctuations remain within a relatively limited range. This variability may stem from changes in the datasets used to train the model, updates to the model itself, or differences in the complexity of exam questions across the years.

When the correct and incorrect answer rates change over time and are visualized, differences between both rates are observed in specific years (Figure 3). In particular, the correct answer rate fluctuates over time, while the incorrect answer rate follows a similar pattern. This allows us to understand better the possible effects of yearly changes in ChatGPT's performance and how the

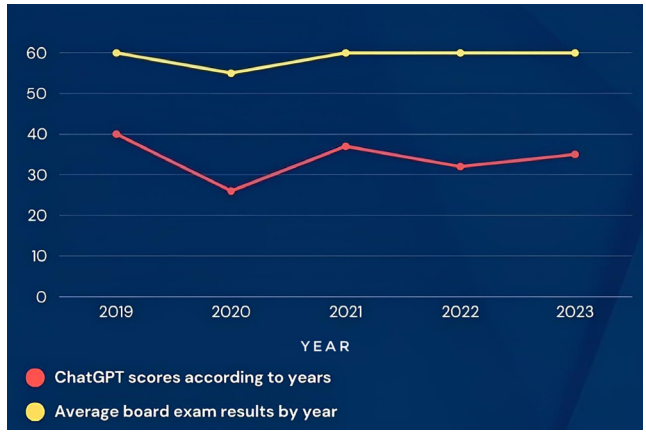


Figure 1. Comparison of ChatGPT and real exam results
ChatGPT: Chat Generative Pre-Trained Transformer

Table 2. Numerical analysis of the answers given by ChatGPT by years

Answers	2019	2020	2021	2022	2023
Correct answer, consistent logic	38	25	31	30	33
Correct answer, inconsistent logic	1	0	0	0	1
Incorrect answer, consistent logic	7	1	5	27	22
Incorrect answer, inconsistent logic	49	69	47	34	36
Question that could not be evaluated	3	5	17	9	8

ChatGPT: Chat Generative Pre-Trained Transformer

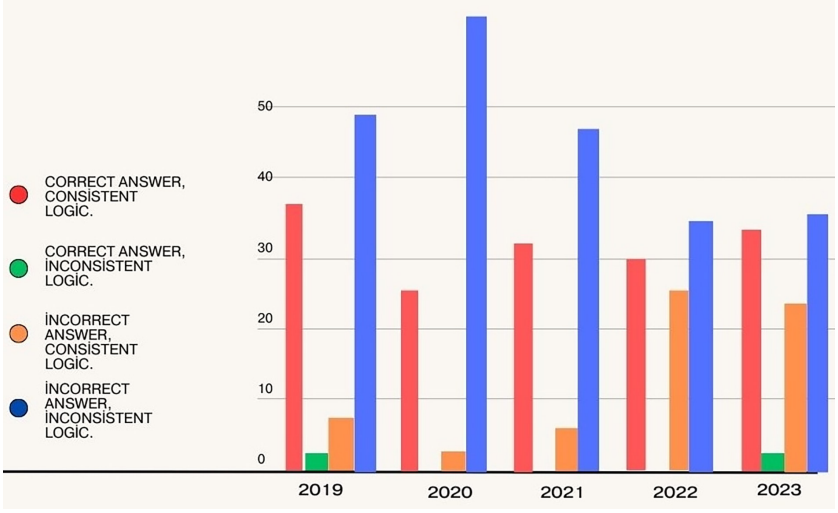


Figure 2. Detailed analysis of the answers given by ChatGPT
ChatGPT: Chat Generative Pre-Trained Transformer

model responds to certain types of questions in specific years. The Mann-Whitney U test to assess whether the differences between the "correct answer, coherent logic" and "incorrect answer, incoherent logic" categories are statistically significant shows a statistically significant difference between the distributions of the two groups (p=0.032). This result indicates that the medians of the two groups are not the same at the 5% significance level. There is a statistically significant difference between the distributions of the answers in the categories "correct answer, consistent logic" and "incorrect answer, inconsistent logic". This analysis shows that ChatGPT's tendency to give correct answers using consistent logic is statistically significantly different from its tendency to provide incorrect answers using inconsistent logic. These results can be considered when developing strategies to improve ChatGPT's performance and accuracy. For example, the focus could be increasing the correct answer rate by strengthening the model's consistent logic.

When the correct answer percentages by years and the simple linear regression model applied to these data are analyzed, a slightly decreasing trend is observed in the correct answer rates as the years progress (Figure 4). The model's slope is negative, indicating a decrease in the correct answer rates as the years progress. However, due to the low R-square (R^2) value, the model only partially explains the variability in the correct answer rates. This indicates that other factors may influence the change over the years (R^2 value 0.0366).

The t-statistic obtained from the paired sample t-test between the results of ChatGPT and the actual exam results was calculated as -15.52 and p=0.0001. This

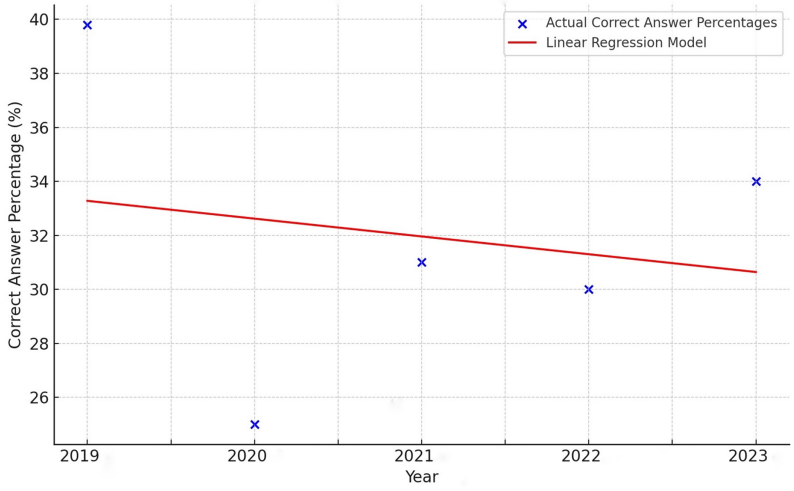


Figure 3. Percentage of correct answers by years and linear regression model

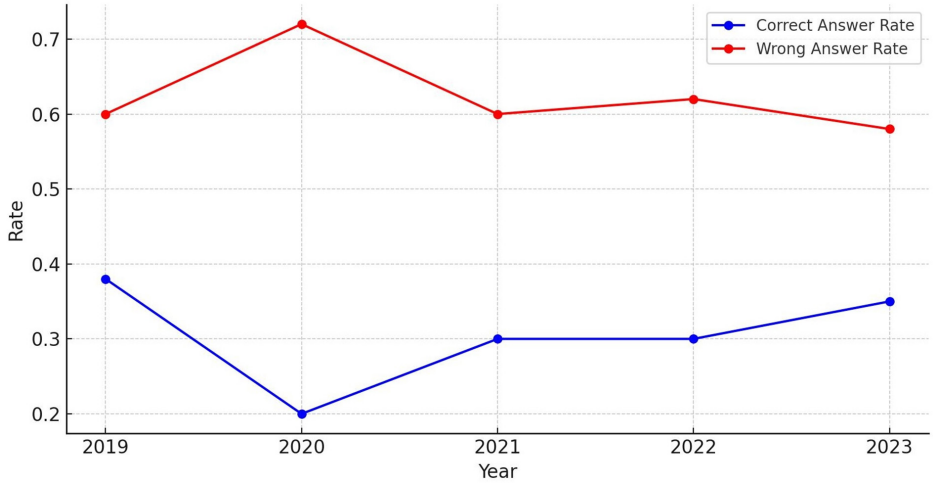


Figure 4. Simple linear regression analysis on the percentage of correct answers by year

result shows a statistically significant difference between the results of ChatGPT and the actual exam results at a 5% significance level. The low p-value indicates that this difference is not random and a significant difference exists in the general population. This analysis shows that ChatGPT's performance on the TOTEK Qualifying Written Examination significantly differs from the exam's overall pass rates. Chat Generative Pre-Trained Transformer's scores are below the average passing scores of the actual exam, indicating that the model has limitations in solving such exam questions and that ChatGPT needs to address some aspects of the exam fully. These differences point to potential areas for improvement in ChatGPT training and its ability to understand test questions.

Discussion

Chat Generative Pre-Trained Transformer in the clinical field has demonstrated its potential to revolutionize healthcare by providing accurate and understandable information on orthopedic issues. Creating interactive quizzes and educational tools for students supports learning and provides instant feedback (15). This AI-powered technology holds great promise for the future of orthopedics, as it is expected to enhance patient care, surgical planning, and medical education (16). Upon evaluating the results of this study, we first observed that ChatGPT's ability to solve exam questions offers certain advantages compared to the performance of actual physicians. Chat Generative Pre-Trained Transformer can effectively address exam questions by quickly accessing and analyzing a vast pool of medical knowledge, which is especially crucial in complex and fast-paced medical scenarios. However, there are areas where improvement is needed in terms of practical applicability and reliability. The lack of citations in the information provided by ChatGPT hinders users from verifying its accuracy, which can limit the use of AI, particularly in healthcare (17).

Nevertheless, some studies indicate that AI can help provide advice and recommendations based on medical history, symptoms, and clinical data (18,19). Chat Generative Pre-Trained Transformer's performance varies depending on specific exam formats and the characteristics of the questions. Additionally, when compared to actual physicians' clinical experience and human skills, ChatGPT's accuracy and reliability may require further improvement. Further research is needed to determine how ChatGPT can be optimally used in educational and assessment processes and identify the areas where it will be most effective. Nonetheless, AI technologies like ChatGPT are expected to play an increasingly significant role in healthcare. These technologies offer educational support to healthcare professionals and contribute to the improvement of diagnosis and treatment processes. However, careful management is necessary to ensure these technologies' effective integration and reliability (17,20-22).

When the existing literature is examined, there are very few articles comparing the board exam results of different countries with ChatGPT's performance. Jain et al. (23) evaluated ChatGPT's decision-making process to assess the performance of the ChatGPT-3.5 version on the Orthopaedic In-Service Training Examination (OITE), conducted by the American Academy of Orthopaedic Surgeons and covering 11 topics, and to determine whether it is practical to adopt it as a resource in this field. At the end of the study, they found that ChatGPT-3.5 performed at the level of a first-year postgraduate (PGY-1) based on annual OITE technical reports for residents. They reported that ChatGPT performed better in basic science and sports. However, when the whole study was evaluated, they noted that ChatGPT in its current form lacks the essential capabilities to be a comprehensive tool in orthopaedic surgery (23).

Kung et al. (4) examined ChatGPT's performance in all three sections of the USMLE directly using publicly available questions on the official website. They reported that ChatGPT performed at or near the passing threshold in all three exams without any special training or support. Furthermore, ChatGPT showed high levels of cohesion and insight in its annotations. These results suggest that large language models may assist medical education and clinical decision-making (4). Gilson et al. (24) evaluated questions from the United States Medical Licensing Examination (USMLE) using ChatGPT. They reported that ChatGPT achieved accuracy rates of 44% (44/100), 42% (42/100), 64.4% (56/87), and 57.8% (59/102) in four data sets: AMBOSS-Step1, AMBOSS-Step2, NBME-Free-Step1, and NBME-Free-Step2, respectively. They also noted that ChatGPT performed 8.15% better than InstructGPT on average across all data sets, while GPT-3 performed similarly to random chance (24). In a study conducted in Peru, the Peruvian National Licensing Medical Examination [Examen Nacional de Medicina (ENAM)] was analyzed using GPT-3.5 and GPT-4. They found that ChatGPT (GPT-3.5 and GPT-4) was able to achieve expert-level performance on ENAM and outperformed most of the examinees (25).

Study Limitations

The findings indicate that ChatGPT can play an essential role in exploring its potential in medical examinations. However, the study also highlighted several critical points and limitations of ChatGPT. First, the dataset used to evaluate ChatGPT's performance is limited in scope. It remains unclear whether the dataset on which ChatGPT is trained is comprehensive enough in the medical domain, which impacts its effectiveness in real-world scenarios (16,26).

Second, the evaluation of ChatGPT's exam performance has limitations when compared to physicians' performance. Chat Generative Pre-Trained Transformer's natural language processing capabilities differ from physicians' clinical experience and expertise. Thus, further research is needed to reach definitive conclusions regarding the real-world applicability of ChatGPT's exam performance (3,27).

Third, ethical and security issues surrounding ChatGPT should also be considered. Using AI systems like ChatGPT in medical training and assessment processes may raise patient privacy and security concerns. Therefore, it is crucial to address these concerns during the implementation of ChatGPT.

Fourth, ChatGPT's ability to process visual questions requires improvement. Compared to human perception and interpretation, ChatGPT currently has a limited capacity to understand visual information. This limitation affects its ability to accurately answer complex or specific questions requiring visual detail (26,28,29).

Conclusion

This study provided a comprehensive evaluation of the performance of AI, specifically ChatGPT, in the TOTEK Qualifying Written Examination and showed that ChatGPT-4 correctly answered less than half of the TOTEK Qualifying Written Exam questions. Our analyses revealed that ChatGPT's ability to understand exam questions and produce appropriate answers was significantly lower compared to the average exam performance of human candidates. While these findings demonstrate ChatGPT's potential as a supportive tool for learning and exam preparation, they also emphasize that it cannot replace human guidance in areas requiring expertise and in-depth knowledge.

In conclusion, we should view ChatGPT and similar AI tools in medical education as aids, not as technologies meant to replace educators. Their role should be as supportive elements in learning processes. Future developments in this technology may allow AI to take a more active role in exam preparation and training; however, this requires ongoing evaluation and a human-centered approach.

Footnote

Ethics Committee Approval: Written permission was obtained from the TOTBID board of directors for this study (document no.: 159, dated: 26.04.2024).

Informed Consent: Not required.

Financial Disclosure: This study received no financial support.

References

1. Atik OŞ. Artificial intelligence, machine learning, and deep learning in orthopedic surgery. Jt Dis Relat Surg. 2022;33:484-5.
2. Beyaz S. A brief history of artificial intelligence and robotic surgery in orthopedics & traumatology and future expectations. Jt Dis Relat Surg. 2020;31:653-5.
3. Sedaghat S. Early applications of ChatGPT in medical practice, education and research. Clin Med (Lond). 2023;23:278-9.
4. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS Digit Health. 2023;2:e0000198.
5. OpenAI. ChatGPT: optimizing language models for dialogue [Internet]. OpenAI; 2022 [cited 2024 July 24 1]. Available from: https://chatgpt.r4wand.eu.org/.
6. OpenAI Platform. Developer quickstart. Get up and running with the OpenAI API Available at: https://platform.openai.com/docs/quickstart / [Accessed: 24.07.2024].
7. Voytovich L, Greenberg C. Natural Language Processing: Practical Applications in Medicine and Investigation of Contextual Autocomplete. Acta Neurochir Suppl. 2022;134:207-14.
8. Zhou B, Yang G, Shi Z, Ma S. Natural Language Processing for Smart Healthcare. IEEE Rev Biomed Eng. 2024;17:4-18.
9. Névéol A, Zweigenbaum P. Clinical Natural Language Processing in 2014: Foundational Methods Supporting Efficient Healthcare. Yearb Med Inform. 2015;10:194-8.
10. Tözün R. The First Term Book of the Turkish Orthopaedics and Traumatology Education Council (TOTEK) (2001-2003), p:29
11. Gönen E, Berk H. Comparison of TOTEK (Turkish Orthopaedics and Traumatology Education Council) and EBOT (European Board of Orthopaedics and Traumatology Fellowship) qualification exams. TOTBİD J. 2014;13:488-99.
12. Güçlü B. The Ninth Term Book of the Turkish Orthopaedics and Traumatology Education Council (TOTEK) (2017-2019), p:43
13. Özdemir G. The Tenth Term Book of the Turkish Orthopaedics and Traumatology Education Council (TOTEK) (2019-2021), pp:26-27
14. Huri G. The Eleventh Term Book of the Turkish Orthopaedics and Traumatology Education Council (TOTEK) (2021-2023), pp:39-41
15. Huang J, Tan M. The role of ChatGPT in scientific communication: writing better scientific review articles. Am J Cancer Res. 2023;13:1148-54.
16. Uz C, Umay E. "Dr ChatGPT": Is it a reliable and useful source for common rheumatic diseases? Int J Rheum Dis. 2023;26:1343-9.

17. Morya VK, Lee HW, Shahid H, et al. Application of ChatGPT for Orthopedic Surgeries and Patient Care. Clin Orthop Surg. 2024;16:347-56.

18. van Dis EAM, Bollen J, Zuidema W, van Rooij R, Bockting CL. ChatGPT: five priorities for research. Nature. 2023;614:224-6.

19. Plana D, Shung DL, Grimshaw AA, Saraf A, Sung JJY, Kann BH. Randomized Clinical Trials of Machine Learning Interventions in Health Care: A Systematic Review. JAMA Netw Open. 2022;5:e2233946.

20. Tomar L, Govil G, Dhawan P. Closed Negative Suction Drain Entrapment in Total Knee Arthroplasty: A Report on the Implications of a Broken Drain Based on the ChatGPT Outlook. Cureus. 2023;15:e36290.

21. Seth I, Rodwell A, Tso R, Valles J, Bulloch G, Seth N. A Conversation with an open artificial intelligence platform on osteoarthritis of the hip and treatment. J Orthop Sports Med. 2023;5:112-20.

22. Bernstein J. Not the Last Word: ChatGPT Can't Perform Orthopaedic Surgery. Clin Orthop Relat Res. 2023;481:651-5.

23. Jain N, Gottlich C, Fisher J, Campano D, Winston T. Assessing ChatGPT's orthopedic in-service training exam performance and applicability in the field. J Orthop Surg Res. 2024;3;19:27.

24. Gilson A, Safranek CW, Huang T, et al. How Does ChatGPT Perform on the United States Medical Licensing Examination (USMLE)? The Implications of Large Language Models for Medical Education and Knowledge Assessment. JMIR Med Educ. 2023;9:e45312.

25. Flores-Cohaila JA, García-Vicente A, Vizcarra-Jiménez SF, et al. Performance of ChatGPT on the Peruvian National Licensing Medical Examination: Cross-Sectional Study. JMIR Med Educ. 2023;9:e48039.

26. Liu PR, Zhang JY, Xue MD, et al. Artificial intelligence to diagnose tibial plateau fractures: an intelligent assistant for orthopedic physicians. Curr Med Sci. 2021;41:1158-64.

27. Cuthbert R, Simpson AI. Artificial intelligence in orthopaedics: can Chat Generative Pre-trained Transformer (ChatGPT) pass Section 1 of the Fellowship of the Royal College of Surgeons (Trauma & Orthopaedics) examination? Postgrad Med J. 2023;99:1110-4.

28. Lum ZC. Can Artificial Intelligence Pass the American Board of Orthopaedic Surgery Examination? Orthopaedic Residents Versus ChatGPT. Clin Orthop Relat Res. 2023;481:1623-30.

29. Khan RA, Jawaid M, Khan AR, Sajjad M. ChatGPT - Reshaping medical education and clinical management. Pak J Med Sci. 2023;39:605-7.

Case Report

DOI: 10.4274/haseki.galenos.2024.9903
Med Bull Haseki 2024;62:250-254

Case Report and Current Literature Review of Adult Cerebrotendinous Xanthomatosis: Evaluation of Treatment Response Based on Gait Analysis Adult Cerebrotendinous Xanthomatosis

Eser Bulus*, Selen Ucem**, Sahin Avci***, F. Feriha Ozer*

*Koc University Hospital, Clinic of Neurology, Istanbul, Turkey
**Koc University Faculty of Medicine, Department of Neurology, Istanbul, Turkey
***Koc University Faculty of Medicine, Department of Medical Genetics, Istanbul, Turkey

Abstract

Cerebrotendinous xanthomatosis (CTX) is a rare autosomal recessive neurometabolic disease caused by a mutation in the *CYP27A1* gene and deficiency of the mitochondrial 27-sterol hydroxylase enzyme. Deficiency of this enzyme leads to the accumulation of cholestanol and cholesterol in various systems (brain, lens, tendons), resulting in chronic diarrhea, juvenile cataracts, tendon xanthoma, and progressive neurodegeneration. Neurological manifestations include ataxia, dystonia, parkinsonism, seizures, dementia, and peripheral neuropathy. We report a 55-year-old woman who presented with chronic and progressive difficulty walking with a history of juvenile cataracts and a family history of parkinsonism. She was found to have cognitive decline, pyramidal-cerebellar signs, and xanthomas at her distal extremities. The diagnosis of CTX was confirmed by a homozygous pathological variant of the *CYP27A1* gene, and treatment with chenodeoxycholic acid was initiated. Because CTX is treatable and preventable, accurate diagnosis and initiation of treatment at the earliest stages are crucial.

Keywords: Rare disease, xanthoma, treatment, chenodeoxycholic acid, gait analysis

Introduction

Cerebrotendinous xanthomatosis (CTX) is a rare autosomal recessive neurometabolic disease. Cerebrotendinous xanthomatosis develops due to a deficiency of the mitochondrial 27-sterol hydroxylase (CYP27) enzyme, which is necessary for bile acid synthesis, and mutations in the *CYP27A1* gene. First described by Van Bogaert in 1937, deficiency of the CYP27 enzyme leading to the accumulation of cholestanol and cholesterol in various systems (brain, lens, tendons), resulting in chronic diarrhea, juvenile cataracts, tendon xanthomas, and various progressive neurological symptoms (1). Neurological manifestations include ataxia, dystonia, parkinsonism, seizures, dementia, and peripheral neuropathy (2). According to our knowledge, approximately 500 cases of CTX have been reported up to date (3). The small number of reported cases suggests that the disease may not have been recognized well and may not have been adequately diagnosed. The heterogeneity of clinical findings and the absence of classical symptoms (juvenile cataracts and tendon xanthomas) in every case or their occurrence after the onset of neurological symptoms can make diagnosis challenging. However, early diagnosis and prompt initiation of appropriate treatment are crucial to prevent potential severe neurological consequences.

Herein, we aimed to describe the clinical, electrophysiological, and genetic characteristics of a 55-year-old female patient diagnosed with CTX and to report her treatment response via gait analysis.

Address for Correspondence: Eser Bulus, Koc University Hospital, Clinic of Neurology, Istanbul, Turkey
Phone: +90 553 803 22 78 **E-mail:** ebulus2004@hotmail.com **ORCID:** orcid.org/0000-0002-5534-8800
Received: 02.05.2024 **Accepted:** 02.09.2024

Case Presentation

A 55-year-old female patient presented with progressive walking difficulties, instability, and forgetfulness that persisted for over 10 years. Her walking condition worsened over the last five months, and the frequency of falls increased. She was a primary school student with a history of normal neurodevelopmental milestones. There was no history of neonatal jaundice or chronic diarrhea during childhood. It was later found out that she underwent bilateral cataract surgery for juvenile cataracts at the age of 8. In her family history, there were symptoms of Parkinsonism in her sister (Figure 1).

During the neurological examination, she presented with cerebellar-type dysarthria, bilateral dysmetria, dysdiadochokinesia, and ataxic gait. The plantar reflexes were bilaterally extensor, and the deep tendon reflexes were hyperactive. The Mini-Mental State Examination revealed moderate cognitive impairment (a score of 13). On physical examination, smooth, firm, ovoid-shaped, non-tender swellings (xanthomas) were observed on the left triceps tendon, bilateral tuberosity of the tibia, and Achilles tendons (Figure 2).

All laboratory tests were within normal limits, including complete blood counts, biochemical tests, and triglyceride and cholesterol levels. Magnetic resonance imaging (MRI) scans showed signs of cerebellar atrophy and hyperintense lesions on T2-weighted imaging in the left internal capsule's posterior limb, the mesencephalon's anterior aspect, dentate nuclei, and deep cerebellar white matter. T1-weighted imaging revealed hypointense lesions in these areas. Additionally, FLAIR imaging revealed hypointense areas in the dentate nuclei and deep cerebellar white matter (Figure 3).

Electromyography (EMG) revealed sensory-predominant axonal polyneuropathy in the lower extremities. Genetic testing was performed with a presumptive diagnosis of CTX. Complete gene sequencing of CYP27A1 (Koc University Hospital, Genetic Diseases Evaluation Center) revealed a homozygous mutation (pathogenic variant) [CYP27A1 c.646G>C (p. Ala216Pro)] compatible with CTX (OMIM 2131700). The same genetic mutation was detected in her sister by Sanger DNA sequencing analysis. The treatment was initiated with a gradual increase in the dosage of chenodeoxycholic acid (CDCA) to 750 mg/day, alongside simvastatin (20 mg/day). Walking and balance analyses were performed before treatment initiation and at the 4th and 12th weeks of treatment. Due to the patient's inability to comprehend commands and cooperate with the analysis before treatment, the assessments could not be conducted. At weeks 4 and 12 of treatment, sensory-based gait analysis and balance functions were evaluated using APDM motion sensors (Noraxon-myoMOTION Research Pro System) and Zebris gait analysis. Zebris gait analysis showed a 135.3% reduction in the center of pressure (COP) oscillation area (from 1468 to 690 mm^2) between the two assessments, with an increased weight bearing on the rear foot (14-29%), indicating that the patient exhibited a more stable and balanced posture while standing (Figure 4).

In the sensory-based gait analysis using motion sensors, comparing two assessments, spatial-temporal walking parameters (such as speed, cadence, step duration, swing, and stance phases) remained similar, but the step width was found to have increased by 19.4% (12±4 to 14±4). At the clinical follow-up in the 4th week of treatment, significant improvements in walking balance and intellectual capacity were observed. Additionally, a reduction in xanthoma size and decreased rigidity compared with the initial examination were noted. However, by the 12th week follow-up, xanthomas appeared more consolidated than

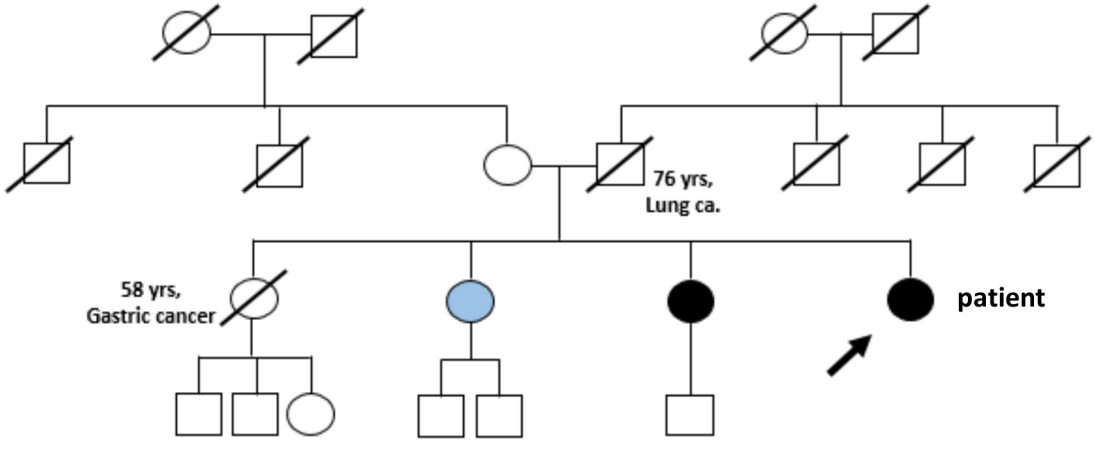


Figure 1. Pedigree, patient and her sister with homozygous pathological variant in *CYP27A1* gene

in the previous evaluation, and no changes were observed in walking and balance. Upon detailed questioning, it was found that the patient had discontinued CDCA treatment due to nausea and indigestion. Emphasizing the importance of adhering to the recommended doses and regimens, the medication dosages were readjusted, and the clinical follow-ups were continued.

Discussion

Cerebrotendinous xanthomatosis, also known as cerebral cholesterinosis, is a rare autosomal recessive disease caused by a mutation in the *CYP27A1* gene, resulting in a deficiency of the mitochondrial 27-sterol hydroxylase enzyme (1). This leads to decreased synthesis of cholic acid and chenodeoxycholic acid and accumulation of cholestanol via upregulation of 7α-hydroxy-4-cholesten-3-one (4). The abnormal accumulation of cholesterol and cholestanol compounds in various tissues leads to multisystemic involvement and heterogeneous clinical symptoms, primarily affecting the neurological, ocular, and musculoskeletal systems (2). The characteristic features of the disease are cataracts (80.3%), cognitive

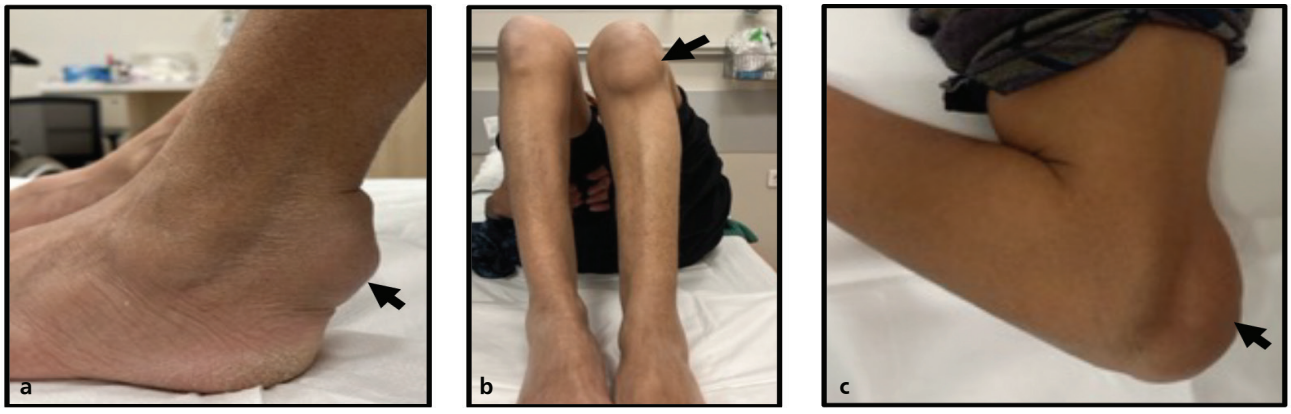


Figure 2. Xanthomas on the left achilles tendon **(a)**, bilateral tuberosity of the tibia **(b)**, and left triceps tendon **(c)**

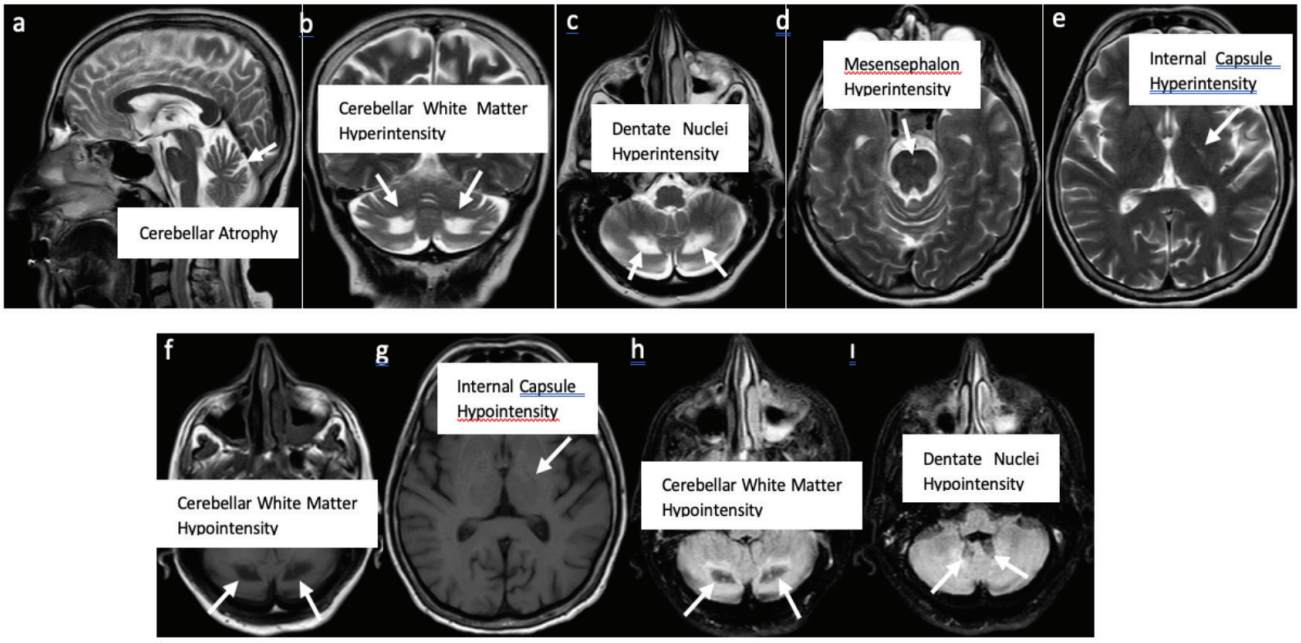


Figure 3. a) Cerebellar atrophy. **b-e)** Hyperintense lesions on T2-weighted imaging in the deep cerebellar white matter, dentate nuclei, anterior of the mesencephalon and posterior limb of the left internal capsule. **f, g)** Hypointense lesions on T1-weighted imaging in the deep cerebellar white matter and posterior limb of the left internal capsule. **h, ı)** Hypointensity areas in the deep cerebellar white matter and dentate nuclei on FLAIR imaging

FLAIR: Fluid-attenuated inversion recovery

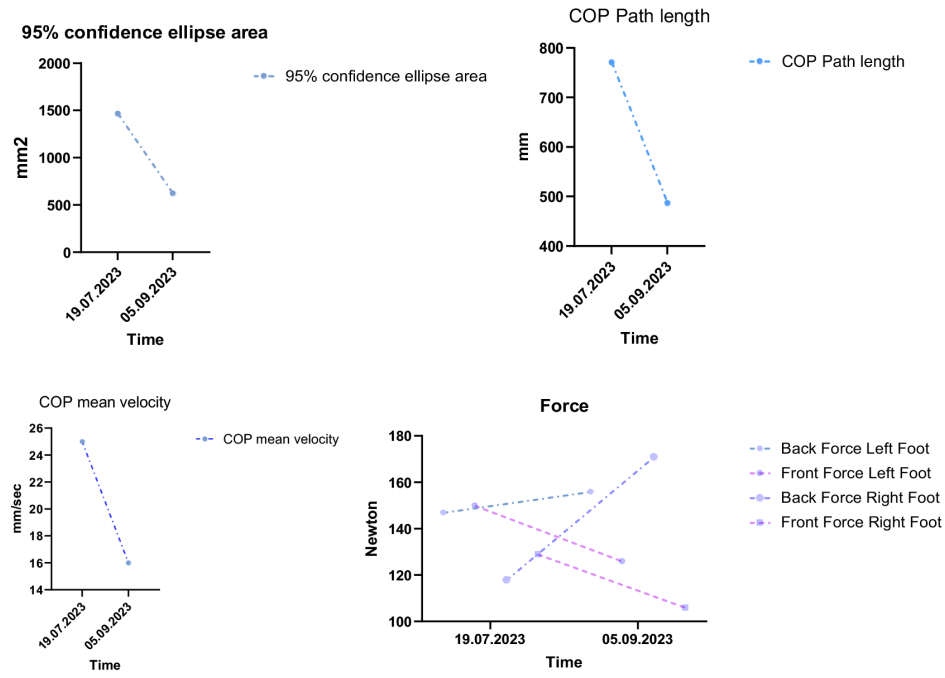


Figure 4. Zebris gait analysis data; 135.3% reduction in the COP oscillation area (135.3%, 1468 to 690 mm²) and increased weight-bearing on the rear foot (14-29%)

COP: Center of pressure

impairment (75.9%), pyramidal signs (72.9%), xanthomas (66.5%), cerebellar ataxia (63.9%), peripheral neuropathy (52.9%), chronic diarrhea (47.6%), seizures (27.4%), and parkinsonism (14.9%) (5). Additional findings include atherosclerosis, cardiovascular disease, osteoporosis, and pulmonary disease (1). In our case, there was a history of juvenile cataract, cognitive impairment, pyramidal-cerebellar signs, polyneuropathy, and xanthoma, but no history of seizures or chronic diarrhea. The patient's sister had Parkinson's disease. Typical MRI findings include cerebral and cerebellar atrophy, hyperintense lesions on T2/fluid-attenuated inversion recovery (FLAIR) sequences, and hypointense lesions on T1-weighted sequences in the periventricular white matter, posterior limb of the internal capsule, cerebral peduncles, anterior pons, cerebellar parenchyma, and dentate nuclei. These lesions appear due to demyelination and axonal damage caused by lipid accumulation. In T2/FLAIR sequences, dentate nuclei may also appear hypointense due to demyelination, hemosiderin deposition, microcalcifications, necrosis, and cystic space formation (6). Diagnosis of the disease is made through clinical, examination, biochemical, imaging, and histopathological findings and the demonstration of homozygous or compound heterozygous mutations in the *CYP27A1* gene (7). In clinical practice, the Mignarri et al. (8) predictive index, which is composed of family history, systemic features, and neurological signs of involvement, may be used to assess the likelihood of a CTX diagnosis. The diagnostic criteria for CTX were established by Stelen et al. (9) and include the measurement of plasma cholestanol levels. However, because of increased opportunities for genetic testing, a diagnosis of the disease can now be made without measuring plasma cholestanol levels. In our case, although we planned to assess plasma cholestanol levels during the diagnostic process, the inability to measure cholestanol levels at our hospital and the ability to perform genetic testing allowed us to establish a diagnosis without measuring cholestanol levels. Chenodeoxycholic acid treatment normalizes cholestanol concentrations, halting disease progression and preventing permanent neurological damage, thereby stabilizing the disease. In addition to CDCA, treatments such as HMG-CoA reductase inhibitors, cholestyramine, and ursodeoxycholic acid have been used in these patients but have not shown significant clinical differences. Early initiation of therapy can reverse or even prevent the progression of neurological symptoms in CTX. Studies have shown that patients starting treatment after the age of 25 years have worse outcomes than those starting therapy at younger ages (10). Age at diagnosis and treatment initiation, brain magnetic resonance imaging findings, and response to CDCA treatment are prognostic factors (1).

In conclusion, we described a 55-year-old adult female patient with CTX who had a CDCA treatment response verified through gait analysis. CTX should be considered among hereditary neurometabolic diseases in patients with a history of juvenile cataracts, pyramidal-cerebellar system signs, ataxia, dystonia/parkinsonism, and/or cognitive impairment, along with observed tendon xanthomas. CDCA treatment should commence as early as possible. Our case demonstrates that CDCA treatment led to an improvement in disease symptoms and prevented irreversible neurological damage.

Footnote

Informed Consent: Informed consent form the patient has been obtained.

Authorship Contributions

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

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

Financial Disclosure: This study received no financial support.

References

1. Nóbrega PR, Bernardes AM, Ribeiro RM, et al. Cerebrotendinous Xanthomatosis: A practice review of pathophysiology, diagnosis, and treatment. Front Neurol. 2022;13:1049850.
2. Carson BE, De Jesus O. Cerebrotendinous Xanthomatosis. 2023 Aug 23. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023.
3. Ejsmont-Sowała K, Książek T, Maciorowska-Rosłan K, et al. Case report: Cerebrotendinous xanthomatosis treatment follow-up. Front Neurol. 2024;15:1409138.
4. Koyama S, Sekijima Y, Ogura M, et al. Cerebrotendinous Xanthomatosis: Molecular Pathogenesis, Clinical Spectrum, Diagnosis, and Disease-Modifying Treatments. J Atheroscler Thromb. 2021;28:905-25.
5. Cao LX, Yang M, Liu Y, Long WY, Zhao GH. Chinese patient with cerebrotendinous xanthomatosis confirmed by genetic testing: A case report and literature review. World J Clin Case. 2020;8:5446-56.
6. Ma C, Ren YD, Wang JC, et al. The clinical and imaging features of cerebrotendinous xanthomatosis: A case report and review of the literature. Medicine (Baltimore). 2021;100:e24687.
7. Nie S, Chen G, Cao X, Zhang Y. Cerebrotendinous xanthomatosis: a comprehensive review of pathogenesis, clinical manifestations, diagnosis, and management. Orphanet J Rare Dis. 2014;9:179.
8. Mignarri A, Gallus GN, Dotti MT, Federico A. A suspicion index for early diagnosis and treatment of cerebrotendinous xanthomatosis. J Inherit Metab Dis. 2014;37:421-9.
9. Stelten BML, Dotti MT, Verrips A, et al. Expert opinion on diagnosing, treating and managing patients with cerebrotendinous xanthomatosis (CTX): a modified Delphi study. Orphanet J Rare Dis. 2021;16:353.
10. Stelten BML, Huidekoper HH, van de Warrenburg BPC, et al. Long-term treatment effect in cerebrotendinous xanthomatosis depends on age at treatment start. Neurology. 2019;92:83-95.