e-ISSN:2602-2079

Volume 8
Issue 1
2024

JOURNAL of Surgery and Medicine

International Medical Journal

JOSAM

Editorial Team

Prasenjit Das, Professor, MD, DNB, MNAMS, MNASc
Department of Pathology
All India Institute of Medical Sciences
New Delhi, India
Website

Seyed Vahid Hosseini, Professor
Shiraz University of Medical Sciences, Shiraz, Iran
Website

Vol. 8 No. 1 (2024)

Published: 2024-02-01

Research Article

J Surg Med. 2024;8(1):1-6.
https://doi.org/10.28982/josam.7511

JOSAM

Research article
https://jsurgmed.com/

Journal of Surgery and Medicine

e-ISSN: 2602-2079

A study of the correlation between magnesium and ferritin levels and the severity of the disease and sleep quality in restless legs syndrome

Sibel Üstün Özek

University of Health Sciences, Prof. Dr. Cemil Taşçıoğlu City Hospital, Department of Neurology, Istanbul, Turkey

ORCID of the author(s)

SÜÖ: https://orcid.org/0000-0003-1165-2648

Corresponding Author
Sibel Üstün Özek
University of Health Sciences, Prof. Dr. Cemil Taşçıoğlu City Hospital, Department of Neurology, Istanbul, Turkey
E-mail: sibelustun@hotmail.com

Ethics Committee Approval
The study was approved by the Ethics Board of the University of Health Sciences, Prof. Dr. Cemil Taşçıoğlu City Hospital (No: 1225, Date: April 2, 2019).
All procedures in this study involving human participants were performed in accordance with the 1964 Helsinki Declaration and its later amendments.

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

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

Published
2024 January 6

Abstract

Background/Aim: Restless legs syndrome (RLS) is a chronic neurological disease that impairs sleep quality, causes emotional stress and affects daily activities. While the association between disease severity and low iron and ferritin levels is known, the magnesium (Mg) results are contradictory. This study aimed to investigate the influence of low Mg and ferritin levels on the severity of the disease and sleep quality.
Methods: A case-control study included 50 RLS patients aged 18–78 years and 50 healthy control patients. Mg and ferritin levels were measured, considering values below <1.8 mg/dL and 75 ng/mL as low. Both groups completed the International Restless Legs Syndrome Study Group score (IRLSSG score) to assess the severity of RLS, as well as the Pittsburgh Sleep Quality Index (PSQI) and the Epworth Sleepiness Scale (ESS) to evaluate subjective sleep quality.
Results: The mean age of RLS patients and the control group was 47.06 (13.35) years and 43.30 (15.43) years, respectively (P=0.196). The RLS patients had an IRLSSG score of 25.16 (6.85). The PSQI total scores, subscale scores, and ESS scores of RLS patients were significantly higher than those of the control group. However, no significant difference was observed in the IRLSSG score, PSQI, and ESS scores based on Mg and ferritin levels. Sleep latency was found to be shorter in individuals with Mg deficiency.
Conclusion: Sleep disorders are prevalent among RLS patients. No correlation was found between Mg and ferritin levels and disease severity or sleep disorders. Furthermore, Mg deficiency did not appear to exacerbate the IRLSSG score or sleep disorder scores.

Keywords: restless legs syndrome, daytime somnolence, magnesium, ferritin, Pittsburgh Sleep Quality Index, Epworth Sleepiness Scale

How to cite: Özek SÜ. A study of the correlation between magnesium and ferritin levels and the severity of the disease and sleep quality in restless legs syndrome. J Surg Med. 2024;8(1):1-6.

Introduction

Restless legs syndrome (RLS), also known as Willis-Ekbom disease, is a chronic neurological condition characterized by an irresistible urge to move the legs, accompanied by sensations of pain, burning, and tingling. Although not intensely painful, these sensations can be highly disruptive [1,2]. The discomfort typically arises during periods of rest, worsens at night, and significantly impairs sleep, leading to chronic sleep disorders and emotional distress [3]. Accurate diagnosis enables effective symptom management, and in some secondary cases, a complete cure may be possible. RLS can be primary or secondary, with the latter resulting from underlying conditions that cause iron deficiency and local dopamine dysfunction in the brain. Approximately 25–30% of iron deficiency-related conditions, such as pregnancy, renal failure, and anemia, can lead to the development of RLS [4-6].

The recommended medications for treating RLS include dopamine agonists, alpha ligands, and opioid agonists [7]. Although not listed in the guidelines, treatments such as tramadol, magnesium sulfate, and baclofen have been attempted [8]. Marshall et al. conducted a comprehensive evaluation of cases involving magnesium and RLS, concluding that magnesium administration did not affect RLS symptoms [9]. However, contrasting findings have been reported in other studies [10,11].

We aimed to investigate the clinical impact of magnesium deficiency, given the lack of evidence-based research on magnesium use and the various approaches to magnesium replacement. We sought to examine and test our hypothesis that lower levels of ferritin and magnesium correspond to greater disease severity and compromised sleep quality, as well as increased daytime sleepiness. Our study aimed to determine whether magnesium and ferritin deficiencies contributed to elevated IRLSSG scores while also assessing concurrent daytime sleepiness and sleep quality using the Epworth Sleepiness Scale (ESS) and the Pittsburgh Sleep Quality Index (PSQI).

Materials and methods

Patients diagnosed with RLS according to the diagnostic criteria of the International Restless Legs Syndrome Study Group at a neurology clinic between 1 January 2017 and 1 April 2019 were included in the case-control study. A total of 100 patients with RLS (36 females and 14 males) and 50 controls (34 females and 16 males) were included. Only patients with idiopathic RLS were included, and those with secondary causes such as chronic renal failure, chronic liver failure, and pregnancy were excluded. The patients' sex, age, disease duration, medication usage, smoking habits, presence of RLS in the family, and laboratory test results (magnesium and ferritin) were recorded. For magnesium and ferritin measurements, 5 mL of blood was drawn from the patients after a 12-h fasting period using a yellow serum separation gel tube. The samples were immediately centrifuged at 4°C, 4000 rpm for 10 min. Magnesium levels were determined using a colorimetric method, and ferritin levels were determined using a chemiluminescent immunoassay with a Roche Cobas Integra 400 Plus analyzer. A magnesium value above 1.8 mg/dL was considered normal, while below 1.8 mg/dL was considered low. Recent literature suggests that the normal ferritin value should be above 75 ng/mL, so the cut-off value for ferritin was set at 75 ng/mL.

At the same time, the severity of the disease was assessed using the Restless Legs Evaluation Scale, and the concurrent sleep quality and daytime sleepiness were evaluated using the ESS and the PSQI. In our study, we assessed the IRLSSG score, as well as the magnesium and ferritin levels of patients diagnosed with RLS. The control group consisted of patients who visited our neurology outpatient clinic for a general medical examination and were assigned the code Z00.0. These patients had no psychiatric, metabolic, systemic, or sleep disorders. Their neurological examination, imaging, and neurophysiological examinations were normal. Only sleep questionnaires were administered to the control group. Ethical issues prevented the collection of magnesium and ferritin values in the control group.

The International Restless Legs Syndrome Study Group score (IRLSSG score) is a scale comprising 10 items to be completed by patients. These items assess the frequency and severity of RLS symptoms experienced during the previous week. The scale is particularly useful in evaluating the effectiveness of treatment.

The Turkish version of the PSQI was validated, and its reliability was assessed in 1996 by Ağargün et al. [12]. The PSQI is a self-report questionnaire comprising 19 items that evaluate sleep quality and disorders over the previous month. The items cover various aspects such as subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleep medication, and daytime dysfunction. The total score on the PSQI ranges from 0 to 21, with a score above 5 indicating poor sleep quality. The PSQI is a valuable, valid, and reliable tool for assessing sleep quality.

The ESS was employed to evaluate the patients' daytime sleepiness and the likelihood of dozing off in various situations. The ESS is a questionnaire that assesses behavioral sleepiness using a four-point self-rating scale. It measures the perceived likelihood of dozing off in eight different situations during recent times. The ESS was developed by Johns [13] and has been demonstrated to be a valid and reliable tool for assessing overall sleepiness levels. It has also been validated and found to be reliable for use in studies on sleep and sleep disorders in Turkey. Individuals scoring 11 and above on the ESS are considered to have excessive daytime sleepiness.

The study received approval from the Ethics Board of the University of Health Sciences, Prof. Dr. Cemil Taşçıoğlu City Hospital (No: 1225, Date: April 2, 2019). Informed consent was obtained from all participants included in the study.

Statistical analysis

Statistical analyses were performed using the Number Cruncher Statistical System software. Descriptive statistics (mean, standard deviation, median, frequency, percentage, minimum, maximum) were used to evaluate the study data. The normality of quantitative data was assessed using the Shapiro-Wilk test and graphical examinations. Student's t-test was utilized for two-group comparisons of normally distributed quantitative variables, while the Mann-Whitney U test was employed for two-group comparisons of non-normally

distributed quantitative variables. The Kruskal-Wallis and Dunn-Bonferroni tests compared non-normally distributed quantitative variables among more than two groups. For comparing qualitative data, Pearson's Chi-square, Fisher's exact, and Fisher-Freeman-Halton tests were applied. Correlations between quantitative variables were evaluated using Spearman's correlation analysis. Statistical significance was defined as $P<0.05$.

Results

The patient group consisted of 36 women (72%) and 14 men (28%), while the control group comprised 34 women (68%) and 16 men (32%). The average age of the patients with RLS and the control group was 47.06 (13.35) years and 43.30 (15.43) years, respectively ($P=0.196$). The IRLSSG scores of the RLS-positive group are shown in Table 1, indicating gender, age, family history, and treatment status (Table 1). There was no statistically significant difference between the two groups in terms of age, sex, and smoking ($P=0.363$, $P=0.663$, and $P=0.262$, respectively). The mean IRLSSG score was found to be 25.16 (6.85).

Table 1: Evaluation of the Pittsburgh Sleep Quality Index and the Epworth Sleepiness Scale Scores by IRLSSG score of RLS-positive group.

		IRLSSG score		
		Min-Max (Med)	Mean (SD)	*P*-value
Gender	Female (n=36)	12-37 (26.5)	26.28 (6.43)	0.045 [a]
	Male (n=14)	7-38 (21)	22.29 (7.29)	
Age	Female Male	18-78 (47.5) 30-77 (47.5)	46.22 (13.07) 49.21 (14.32)	0.483 [b]
PSQI Total	Good (n=4)	18-26 (23)	22.5 (3.7)	0.316 [a]
	Bad (n=46)	7-38 (25.5)	25.39 (7.03)	
Epworth Level	0-5 (n=27)	7-35 (26)	25.37 (6.98)	0.648 [c]
	6-10 (n=13)	13-37 (21)	23.38 (7.01)	
	11-12 (n=1)	30	30	
	13-15 (n=4)	21-28 (26)	25.25 (3.10)	
	>15 (n=5)	15-38 (26)	27.60 (8.91)	
Treatment	None (n=32)	7-38 (26.5)	26.03 (6.95)	0.234 [b]
	Yes (n=18)	12-37 (25)	23.61 (6.56)	
Family History	None (n=36)	12-38 (26)	26.11 (6.26)	0.116 [a]
	Yes (n=14)	7-37 (21)	22.71 (7.88)	
Magnesium	Low (n=12)	13-34 (27.5)	26.67 (6.15)	0.333 [a]
	Normal (n=38)	7-38 (25)	24.68 (7.06)	
Ferritin	Low (n=45)	12-38 (25)	25.6 (6.03)	0.418 [a]
	Normal (n=5)	7-35 (18)	21.2 (12.34)	

IRLSSG: International Restless Legs Syndrome Study Group, RLS: Restless legs syndrome, [a] Mann-Whitney U Test, [b] Student t-test, [c] Kruskal-Wallis Test

When evaluating the PSQI subscales of the patient group, statistically significant differences were observed in the scores obtained for subjective sleep quality, sleep latency, sleep duration, habitual sleep effectiveness, sleep disorder, and daytime dysfunction subscales compared to the control group ($P<0.001$). Additionally, the use of sleeping aids was significantly higher in the patient group ($P=0.002$) (Table 2).

The patient group obtained significantly higher scores for total ESS than the control group ($P=0.007$). A statistically significant difference was found in the ESS levels between the two groups ($P=0.012$). Moreover, the proportion of participants in the patient group scoring above ESS 15 was significantly higher than that of the control group.

No differences were found between the PSQI subscores and ESS severity in relation to ferritin levels (Table 3).

The scores obtained by the patients for the PSQI subscales "Subjective Sleep Quality", "Sleep Duration", "Sleep Effectiveness", "Sleep Disorder", "Sleeping Aid", and "Daytime Dysfunction", as well as for the total scale, showed no statistically significant differences in relation to magnesium levels ($P=0.289$, $P=0.924$, $P=0.435$, $P=0.233$, $P=0.350$, and $P=0.648$, respectively). However, patients with low magnesium levels had significantly lower scores in the PSQI "Sleep Latency" subscale ($P=0.019$). There was no statistically significant difference in the scores obtained by patients for total ESS in relation to magnesium levels ($P=0.615$) (Figure 1). When Epworth sleepiness levels were evaluated, there was a statistically significant difference in magnesium levels ($P=0.042$). However, the proportion of patients with low magnesium levels and ESS levels between 13 and 15 (indicating increased daytime moderate sleepiness) was higher than those with normal magnesium levels (Figure 2).

Figure 1: Distribution of sleep quality subscales by magnesium status

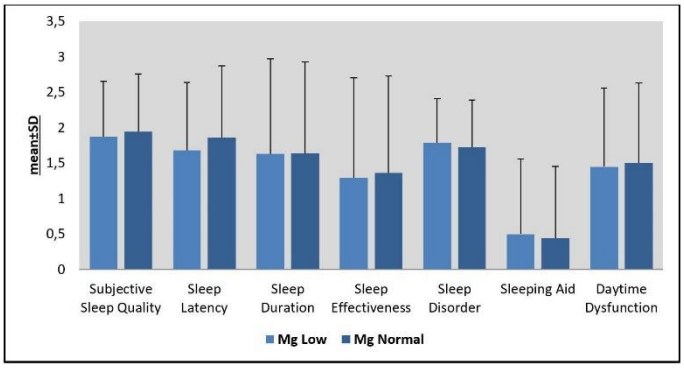
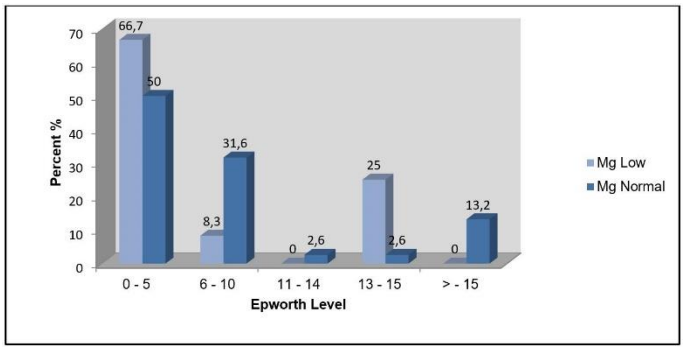


Figure 2: Distribution of Epworth level by Mg



No statistically significant differences were found between PSQI and ESS severity and the IRLSSG score ($P=0.316$ and $P=0.648$, respectively). However, the IRLSSG score of women was found to be significantly higher compared to that of men ($P=0.045$). Among the RLS cases, 64% remained untreated. Out of the RLS group, 15 out of 18 patients (36%) received pramipexole at doses ranging between 0.250 and 0.500 mg. One patient was taking gabapentin 600 mg, one patient was taking pregabalin 150 mg, and one patient was taking a combination of pramipexole and gabapentin. No significant difference in IRLSSG score was found between those who were receiving medication for RLS and those who were not ($P=0.234$). There was no difference in IRLSSG score between individuals with and without a family history of RLS ($P=0.116$). Similarly, no significant difference was found between smokers and non-smokers in terms of IRLSSG score ($P=0.312$). Additionally, there were no significant differences in RLS severities of the patients in relation to magnesium and ferritin levels ($P=0.333$, $P=0.418$, respectively). Furthermore, no significant correlation was found between age and IRLSSG score ($P=0.316$).

A statistically significant moderate-level positive correlation of 0.441 was found between the scores obtained by the patients under the PSQI "Subjective Sleep Quality" subscale and the IRLSSG score (r=0.441; $P<0.001$). Additionally, a

statistically significant weak positive correlation of 0.294 was observed between the scores obtained under the "Sleep Duration" subscale and the IRLSSG score (r=0.294; *P*=0.038). Furthermore, a statistically significant moderate positive correlation of 0.473 was identified between the scores obtained under the "Daytime Dysfunction" subscale and the IRLSSG score (r=0.473; *P*<0.001). Likewise, a statistically significant moderate positive correlation of 0.487 was found between the total scores obtained for the PSQI and the IRLSSG score (r=0.487; *P*<0.001). However, no statistically significant correlations were found between the scores obtained under the PSQI "Sleep Latency", "Sleep Effectiveness", "Sleep Disorder", and "Sleeping Aids" subscales and the IRLSSG score (*P*=0.104, *P*=0.058, *P*=0.427, and *P*=0.421) (Table 4).

No statistically significant correlation was found between the total scores obtained by the patients for the ESS and the IRLSSG score (*P*=0.809).

Table 4: Evaluation of the Pittsburgh Sleep Quality Index and the Epworth Sleepiness Scale Scores by IRLSSG score

	IRLSSG score	
	r	*P*-value
Age	0.021	0.885
Subjective Sleep Quality	0.441	<0.001
Sleep Latency	0.233	0.104
Sleep Duration	0.294	0.038
Sleep Effectiveness	0.270	0.058
Sleep Disorder	0.115	0.427
Sleeping Aid	0.116	0.421
Daytime Dysfunction	0.473	<0.001
PSQI Total	0.487	<0.001
Epworth Total	-0.035	0.809

IRLSSG: International Restless Legs Syndrome Study Group, r: Spearman's Correlation Coefficient

Table 2: Evaluation of the Pittsburgh Sleep Quality Index and the Epworth Sleepiness Scale Scores by group

			Groups		
		Total	**Control**	**Patient**	***P*-value**
Subjective Sleep Quality	Min-Max (Median)	0-3 (1)	0-2 (1)	0-3 (2)	<0.001 [a]
	Mean (SD)	1.39 (0.88)	0.84 (0.51)	1.94 (0.82)	
Sleep Latency	Min-Max (Median)	0-3 (1)	0-3 (1)	0-3 (2)	<0.001 [a]
	Mean (SD)	1.44 (1.04)	1.02 (0.89)	1.86 (1.01)	
Sleep Duration	Min-Max (Median)	0-3 (0)	0-2 (0)	0-3 (2)	<0.001 [a]
	Mean (SD)	1.00 (1.19)	0.36 (0.6)	1.64 (1.29)	
Sleep Effectiveness	Min-Max (Median)	0-3 (0)	0-2 (0)	0-3 (1)	<0.001 [a]
	Mean (SD)	0.71 (1.18)	0.06 (0.31)	1.36 (1.37)	
Sleep Disorder	Min-Max (Median)	0-3 (1)	0-2 (1)	0-3 (2)	<0.001 [a]
	Mean (SD)	1.42 (0.67)	1.12 (0.52)	1.72 (0.67)	
Sleeping Aid	Min-Max (Median)	0-3 (0)	0-0 (0)	0-3 (0)	0.002 [a]
	Mean (SD)	0.22 (0.75)	0 (0)	0.44 (1.01)	
Daytime Dysfunction	Min-Max (Median)	0-4 (1)	0-2 (0)	0-4 (1.5)	<0.001 [a]
	Mean (SD)	0.93 (1.06)	0.36 (0.56)	1.5 (1.13)	
PSQI Total	Min-Max (Median)	0-17 (5)	0-8 (4)	3-17 (11.5)	<0.001 [a]
	Mean (SD)	7.09 (4.60)	3.76 (1.68)	10.42 (4.16)	
Sleep Quality	Good	36 (36.0)	32 (64.0)	4 (8.0)	<0.001 [b]
	Bad	64 (64.0)	18 (36.0)	46 (92.0)	
Epworth Total	Min-Max (Median)	0-24 (4)	0-12 (4)	0-24 (5)	0.007 [a]
	Mean (SD)	5.66 (5.24)	4.06 (3.32)	7.26 (6.27)	
Epworth Level	0-5	63 (63.0)	36 (72.0)	27 (54.0)	0.012 [c]
	6-10	24 (24.0)	11 (22.0)	13 (26.0)	
	11-12	4 (4.0)	3 (6.0)	1 (2.0)	
	13-15	4 (4.0)	0 (0.0)	4 (8.0)	
	>15	5 (5.0)	0 (0.0)	5 (10.0)	

[a] Mann-Whitney U Test, [b] Pearson Chi-square Test, [c] Fisher-Freeman-Halton Test

Table 3: Evaluation of Pittsburgh Sleep Quality Scale and Epworth Sleepiness Scale Scores according to ferritin levels

		Ferritin		
		Low (n=45)	**Normal (n=5)**	***P*-value**
Subjective Sleep Quality	Min-Max (Median)	1-3 (3)	0-3 (2)	0.508 [a]
	Mean (SD)	2.2 (1.1)	1.94 (0.82)	
Sleep Latency	Min-Max (Median)	1-3 (2)	0-3 (2)	0.802 [a]
	Mean (SD)	2 (1)	1.86 (1.01)	
Sleep Duration	Min-Max (Median)	0-3 (2)	0-3 (2)	0.802 [a]
	Mean (SD)	1.8 (1.3)	1.64 (1.29)	
Sleep Effectiveness	Min-Max (Median)	0-3 (1)	0-3 (1)	0.900 [a]
	Mean (SD)	1.4 (1.52)	1.36 (1.37)	
Sleep Disorder	Min-Max (Median)	0-2 (2)	0-3 (2)	0.529 [a]
	Mean (SD)	1.4 (0.89)	1.72 (0.67)	
Sleeping Aid	Min-Max (Median)	0-0 (0)	0-3 (0)	0.488 [a]
	Mean (SD)	0 (0)	0.44 (1.01)	
Daytime Dysfunction	Min-Max (Median)	0-3 (2)	0-4 (1.5)	0.508 [a]
	Mean (SD)	1.8 (1.1)	1.5 (1.13)	
PSQI Total	Min-Max (Median)	5-15 (12)	3-17 (11,5)	0.925 [a]
	Mean (SD)	10.6 (4.39)	10.42 (4.16)	
Sleep Quality	Good	4 (8.9)	0 (0.0)	1.000 [b]
	Bad	41 (91.1)	5 (100.0)	
Epworth Total	Min-Max (Median)	2-9 (5)	0-24 (5)	0.571 [a]
	Mean (SD)	5 (3.08)	7.26 (6.27)	
Epworth Level	0-5	24 (53.3)	3 (60.0)	1.000 [c]
	6-10	11 (24.4)	2 (40.0)	
	11-12	1 (2.2)	0 (0.0)	
	13-15	4 (8.9)	0 (0.0)	
	>15	5 (11.1)	0 (0.0)	

[a] Mann-Whitney U Test, [b] Fisher's Exact Test, [c] Fisher-Freeman-Halton Test

Discussion

The total PSQI and all of its subscales were significantly higher in patients with RLS than the normal controls. However, no significant difference was found in the IRLSSG score, sleep quality, and daytime sleepiness based on ferritin values. Similarly, no difference was found in IRLSSG scores based on magnesium values. Apart from the sleep latency subscale, there was no difference in the total PSQI and other components. When comparing sleep latency based on magnesium values, it was observed that those with lower levels had shorter sleep latency, suggesting that magnesium deficiency does not have an aggravating effect. Although no difference was found between magnesium levels and ESS scores, there was a higher distribution of increased daytime moderate sleepiness scores among those with magnesium deficiency. However, this was unrelated to nighttime sleep quality.

The prevalence of RLS in adults in Europe and North America ranges from 2% to 3%. A prevalence study conducted by Sevim et al. [14] in Turkey, which involved 3234 individuals, reported an RLS ratio of 3.19%. RLS was found to be more common in women, smokers, and individuals living at high

altitudes. Our study also had a higher percentage of females, and the IRLSSG score was significantly higher among women.

Studies have shown that RLS can manifest as more severe in patients with familial characteristics, with higher IRLSSG scores observed in those who have a family history of the condition [15,16]. However, in our study, no significant difference was found in terms of the IRLSSG score between individuals with and without a family history of RLS.

While it has been reported in the literature that smoking and coffee can exacerbate RLS symptoms [17], there are also reports suggesting that smoking may alleviate the symptoms [18]. However, our study found no significant difference between smokers and non-smokers in relation to RLS symptoms.

The effectiveness of oral or intravenous (IV) iron treatment has been demonstrated in patients with serum ferritin levels below 50 ng/mL. Recent studies recommend maintaining ferritin levels above 75 ng/mL [19]. The treatment of RLS often involves magnesium (Mg) supplementation and ensuring adequate iron stores.

Ferritin deficiency reflects impaired iron metabolism, and studies have indicated that central nervous system iron deficiency contributes to the development of RLS in affected patients [20]. Iron deficiency is believed to potentially induce dopaminergic dysfunction, as iron serves as a cofactor in the dopamine generation process via tyrosine hydroxylase, potentially exacerbating RLS symptoms [21]. However, in our patient population, no significant correlation was found between ferritin levels and the IRLSSG score. The most commonly used medication among our patients was a dopamine agonist [7].

Magnesium is involved in more than 300 biochemical reactions in the body [22]. It acts as a natural antagonist of NMDA receptors and an agonist of GABA receptors, which have a relaxing effect on the body and contribute to improved sleep [23]. Magnesium supplementation is often recommended as a potential remedy for relieving symptoms of RLS or periodic limb movement disorder (PLMD), and it is commonly prescribed for leg cramps as well. A daily magnesium supplement of 500 mg has been associated with significant improvements in various aspects of sleep, including sleeplessness severity index, sleep duration, sleep effectiveness, sleep onset latency, as well as changes in serum cortisol concentration, renin levels, and melatonin [24].

All articles reporting the effects of magnesium supplementation on changes in RLS and/or PLMD were examined for a systematic review. No significant curative effect of magnesium was found. Following the quality assessment and synthesis of the evidence, it was reported that no conclusion was reached regarding the effectiveness of magnesium on RLS/PLMD. It is unclear whether magnesium helps to relieve RLS or PLMD or which patients benefit from it [9]. In one study, the magnesium-supplemented group showed considerable mitigation of periodic limb movements during sleep compared to the placebo group. Additionally, there was a significant increase in general sleep efficiency, rising from 75% to 85% [25].

According to a meta-analysis, magnesium was found to be ineffective in the general population when evaluated in the context of leg cramps. However, it showed a slight effectiveness in pregnant women [11].

Rondanelli et al. [26] conducted a study that found supplementation with magnesium resulted in an improvement in the total score of the PSQI. Furthermore, several studies demonstrate that magnesium deficiency affects the circadian cycle, depletes melatonin, and contributes to sleep disorders [27]. Hornyak et al. [25] demonstrated that magnesium treatment could be a beneficial alternative for patients experiencing insomnia related to RLS or periodic limb movement syndrome. Considering the existing literature on the administration and effectiveness of magnesium, our data suggest that magnesium deficiency does not exacerbate.

A study reported that magnesium deficiency could be a potential cause of increased RLS symptoms in patients undergoing dialysis [28]. Additionally, the potential therapeutic effects of magnesium and coenzyme Q replacement in patients with type 2 diabetes and RLS have been discussed [10].

While conclusive evidence is lacking, it is advisable to exercise caution regarding the extensive use of magnesium and remain vigilant for potential complications. One such complication is the increased risk of constipation, which can lead to toxic symptoms and harm prognosis, potentially even resulting in mortality. Therefore, patients should be cautioned about the risk of constipation [29].

A high percentage of patients with RLS, specifically 94%, report difficulties in both falling asleep and maintaining sleep. Additionally, 84.7% of patients experience difficulty falling asleep, while 86% face challenges exclusively in maintaining sleep [30]. Among our patients, the sleep disorder frequency encompasses various parameters, including sleep quality, difficulties in falling asleep and maintaining sleep, and a reduction in the total sleep duration.

In a study, a significant percentage of patients with mild or severe RLS reported spending over 30 min to fall asleep, as well as waking up three or more times during the night, which is a commonly mentioned symptom of sleeplessness [3]. RLS often leads to interrupted sleep and is frequently reported as a cause of insomnia. The loss and disruption of sleep adversely affect overall health and daily functioning. Among healthcare professionals, a notable proportion experiences some degree of impaired sleep quality, with RLS and depressive symptoms also being frequent in this population [31]. The primary recommendation for managing sleep disorders is practicing good sleep hygiene [32].

None of the patients we assessed were receiving sleeping medication despite experiencing sleep problems; however, this percentage was higher compared to that of the normal control group. Similar findings were observed in a study involving 133 patients with RLS, where 85% of the patients reported difficulties in falling asleep or maintaining sleep [30]. In patients with RLS, sleep disorders are associated with a lack of energy and concentration on the following day [3]. Our patients exhibited higher levels of next-day sleepiness compared to the normal controls.

Limitation

Our study had certain limitations. First, our patient cohort consisted mainly of resistant cases seeking treatment at a neurology polyclinic in a tertiary-care training and research hospital, which may limit the generalizability of our findings.

Second, being a single-center study, it may not fully represent the broader population. Additionally, our study group was relatively small in terms of sample size. Furthermore, the number of patients with low magnesium levels was also limited within this group.

Conclusion

Our study did not find any correlation between magnesium and ferritin levels, disease severity, and sleep disorders. However, there were correlations observed between disease severity and total PSQI score, as well as subjective sleep quality, sleep duration, and the severity of daytime dysfunction. Importantly, we did not observe any aggravating effects of magnesium deficiency. Therefore, it is crucial to exercise caution when prescribing supplementary treatments, such as magnesium, without conclusive evidence.

References

1. Ekbom KA. Restless legs syndrome. Neurology. 1960 Sep;10:868-73. doi: 10.1212/wnl.10.9.868. PMID: 13726241.
2. Allen RP, Picchietti DL, Garcia-Borreguero D, Ondo WG, Walters AS, Winkelman JW, et al. International Restless Legs Syndrome Study Group. Restless legs syndrome/Willis-Ekbom disease diagnostic criteria: updated International Restless Legs Syndrome Study Group (IRLSSG) consensus criteria history, rationale, description, and significance. Sleep Med. 2014 Aug;15(8):860-73. doi: 10.1016/j.sleep.2014.03.025. Epub 2014 17 May. PMID: 25023924.
3. Walters AS, LeBrocq C, Dhar A, Hening W, Rosen R, Allen RP, Trenkwalder C; International Restless Legs Syndrome Study Group. Validation of the International Restless Legs Syndrome Study Group rating scale for restless legs syndrome. Sleep Med. 2003 Mar;4(2):121-32. doi: 10.1016/s1389-9457(02)00258-7. PMID: 14592342.
4. Winkelman JW, Chertow GM, Lazarus JM. Restless legs syndrome in end-stage renal disease. Am J Kidney Dis. 1996 Sep;28(3):372-8. doi: 10.1016/s0272-6386(96)90494-1. PMID: 8804235.
5. Salih AM, Gray RE, Mills KR, Webley M. A clinical, serological and neurophysiological study of restless legs syndrome in rheumatoid arthritis. Br J Rheumatol. 1994 Jan;33(1):60-3. doi: 10.1093/rheumatology/33.1.60. PMID: 8162461.
6. O'Hare JA, Abuaisha F, Geoghegan M. Prevalence and forms of neuropathic morbidity in 800 diabetics. Ir J Med Sci. 1994 Mar;163(3):132-5. doi: 10.1007/BF02965972. PMID: 8200777.
7. Hening W, Walters AS, Allen RP, Montplaisir J, Myers A, Ferini-Strambi L. Impact, diagnosis and treatment of restless legs syndrome (RLS) in a primary care population: the REST (RLS epidemiology, symptoms, and treatment) primary care study. Sleep Med. 2004 May;5(3):237-46. doi: 10.1016/j.sleep.2004.03.006. PMID: 15165529.
8. Karadeniz Kaynak D. Sleep Disorders Behind the Complaint of Insomnia; Restless Legs Syndrome and Periodic Limb Movements Disorders of Sleep. Noro Psikiyatr Ars 2007;44:95-100.
9. Marshall NS, Serinel Y, Killick R, Child JM, Raisin I, Berry CM, et al. Magnesium supplementation for the treatment of restless legs syndrome and periodic limb movement disorder: A systematic review. Sleep Med Rev. 2019 Dec;48:101218. doi: 10.1016/j.smrv.2019.101218. Epub 2019 16 October. PMID: 31678660.
10. Metta V, Sampath N, Vm R, Iska A. Primary restless legs syndrome in patients with type 2 diabetes mellitus: Efficacy of magnesium & coenzyme q10 therapy. J Neurol Sci. 2015;357(S1):e271. doi: 10.1016/j.jns.2015.08.952.
11. Sebo P, Cerutti B, Haller DM. Effect of magnesium therapy on nocturnal leg cramps: a systematic review of randomized controlled trials with meta-analysis using simulations. Fam Pract. 2014 Feb;31(1):7-19. doi: 10.1093/fampra/cmt065. Epub 2013 26 November. PMID: 24280947.
12. Ağargün MY, Kara H, Anlar O. Pittsburgh Uyku Kalitesi İndeksi'nin Geçerliği ve Güvenirliği. Turk Psikiyatri Derg. 1996;7(2):107-11.
13. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep. 1991 Dec;14(6):540-5. doi: 10.1093/sleep/14.6.540. PMID: 1798888.
14. Sevim S, Dogu O, Camdeviren H, Bugdayci R, Sasmaz T, Kaleagasi H, et al. Unexpectedly low prevalence and unusual characteristics of RLS in Mersin, Turkey. Neurology. 2003 Dec 9;61(11):1562-9. doi: 10.1212/01.wnl.0000096173.91554.b7. PMID: 14663043.
15. Allen RP, La Buda MC, Becker P, Earley CJ. Family history study of the restless legs syndrome. Sleep Med. 2002 Nov;3 Suppl: S3-7. doi: 10.1016/s1389-9457(02)00140-5. PMID: 14592159.
16. Desai AV, Cherkas LF, Spector TD, Williams AJ. Genetic influences in self-reported symptoms of obstructive sleep apnea and restless legs: a twin study. Twin Res. 2004 Dec;7(6):589-95. doi: 10.1375/1369052042663841. PMID: 15607009.
17. Bayard M, Avonda T, Wadzinski J. Restless legs syndrome. Am Fam Physician. 2008 Jul 15;78(2):235-40. PMID: 18697508.
18. Oksenberg A. Alleviation of severe restless legs syndrome (RLS) symptoms by cigarette smoking. J Clin Sleep Med. 2010 15 October;6(5):489-90. PMID: 20957852; PMCID: PMC2952755.
19. Trenkwalder C, Winkelmann J, Oertel W, Virgin G, Roubert B, Mezzacasa A; FCM-RLS Study Investigators. Ferric carboxymaltose in patients with restless legs syndrome and nonanemic iron deficiency: A randomized trial. Mov Disord. 2017 Oct;32(10):1478-82. doi: 10.1002/mds.27040. Epub 2017 23 June. PMID: 28643901; PMCID: PMC5655783.
20. Trotti LM. Restless Legs Syndrome and Sleep-Related Movement Disorders. Continuum (Minneap Minn). 2017 Aug;23(4, Sleep Neurology):1005-16. doi: 10.1212/CON.0000000000000488. PMID: 28777173.
21. Earley CJ, Connor J, Garcia-Borreguero D, Jenner P, Winkelman J, Zee PC, et al. Altered brain iron homeostasis and dopaminergic function in Restless Legs Syndrome (Willis-Ekbom Disease). Sleep Med. 2014 Nov;15(11):1288-301. doi: 10.1016/j.sleep.2014.05.009. Epub 2014 16 June. PMID: 25201131.
22. Altura BM. Basic biochemistry and physiology of magnesium: a brief review. Magnes Trace Elem. 1991-1992;10(2-4):167-71. PMID: 1844549.
23. Kapur N, Friedman R. Oral ketamine: a promising treatment for restless legs syndrome. Anesth Analg. 2002 Jun;94(6):1558-9. doi: 10.1097/00000539-200206000-00034. PMID: 12032026.
24. Abbasi B, Kimiagar M, Sadeghniiat K, Shirazi MM, Hedayati M, Rashidkhani B. The effect of magnesium supplementation on primary insomnia in elderly: A double-blind placebo-controlled clinical trial. J Res Med Sci. 2012 Dec;17(12):1161-9. PMID: 23853635; PMCID: PMC3703169.
25. Hornyak M, Voderholzer U, Hohagen F, Berger M, Riemann D. Magnesium therapy for periodic leg movements-related insomnia and restless legs syndrome: an open pilot study. Sleep. 1998 Aug 1;21(5):501-5. doi: 10.1093/sleep/21.5.501. PMID: 9703590.
26. Rondanelli M, Opizzi A, Monteferrario F, Antoniello N, Manni R, Klersy C. The effect of melatonin, magnesium, and zinc on primary insomnia in long-term care facility residents in Italy: a double-blind, placebo-controlled clinical trial. J Am Geriatr Soc. 2011 Jan;59(1):82-90. doi: 10.1111/j.1532-5415.2010.03232.x. PMID: 21226679.
27. Durlach J, Pagès N, Bac P, Bara M, Guiet-Bara A. Biorhythms and possible central regulation of magnesium status, phototherapy, darkness therapy and chronopathological forms of magnesium depletion. Magnes Res. 2002 Mar;15(1-2):49-66. PMID: 12030424.
28. Sinniah D. Magnesium deficiency: a possible cause of restless leg syndrome in haemodialysis patients. Intern Med J. 2015 Apr;45(4):467-8. doi: 10.1111/imj.12715. PMID: 25827521.
29. Thanikonda V, Levine A. Hypermagnesemia in a Patient Taking Magnesium Oxide for Restless Leg Syndrome: Presenting as Shock. Am J Respir Crit Care Med. 2020;201:A5161 doi: 10.1164/ajrccm-conference.2020.201.1 MeetingAbstracts.A5161.
30. Montplaisir J, Boucher S, Poirier G, Lavigne G, Lapierre O, Lespérance P. Clinical, polysomnographic, and genetic characteristics of restless legs syndrome: a study of 133 patients diagnosed with new standard criteria. Mov Disord. 1997 Jan;12(1):61-5. doi: 10.1002/mds.870120111. PMID: 8990055.
31. Demir Ülkü F. Sleep quality & prevalence of restless legs syndrome among healthcare professionals. J Surg Med. 2020;4(2):144-7.
32. Burman D. Sleep Disorders: Restless Legs Syndrome. FP Essent. 2017 Sep;460:29-32. PMID: 28845959.

J Surg Med. 2024;8(1):7-11.
https://doi.org/10.28982/josam.7588

JOSAM

Research article
https://jsurgmed.com/

Journal of Surgery and Medicine

e-ISSN: 2602-2079

Relationship between HER2 and clinicopathological data in gastric adenocarcinomas

Cengiz Ceylan, Yavuz Selim Angin, Necip Tolga Baran, Hüseyin Kocaaslan, Emrah Cengiz, Omer Güngörür, Mehmet Güzel, Kutay Saglam, Cemalettin Aydin

Department of Surgery, Turgut Ozal Medical Center, Inönü University, Malatya, Turkey

ORCID of the author(s)

CC: https://orcid.org/0000-0003-3471-8726
YSA: https://orcid.org/0000-0001-5315-8360
NTB: https://orcid.org/0000-0001-6323-9885
HK: https://orcid.org/0000-0002-4590-4850
EC: https://orcid.org/0000-0003-2879-4976
OG: https://orcid.org/0000-0002-2559-1682
MG: https://orcid.org/0000-0001-6959-8539
KS: https://orcid.org/0000-0002-0919-8370
CA: https://orcid.org/0000-0002-7839-3495

Corresponding Author
Cengiz Ceylan
Department of Surgery, Turgut Ozal Medical Center, Inonu University, 44280 Battalgazi, Malatya, Turkey
E-mail: ceylancengiz@ymail.com

Ethics Committee Approval
The study was approved by the Inönü University Scientific Research and Publication Ethics Committee (No: 2022/2914, Date: February 8, 2022).
All procedures in this study involving human participants were performed in accordance with the 1964 Helsinki Declaration and its later amendments.

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

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

Published
2024 January 11

Abstract

Background/Aim: The impact of human epidermal growth factor receptor 2 (HER2) overexpression on the surveillance of gastric cancer remains uncertain. Typically, HER2 status is assessed in both locally advanced and metastatic diseases, and targeted therapies are applied to cases with HER2-positive status. Our objective was to investigate the correlation between HER2 receptor status, clinicopathological characteristics, and prognosis in gastric cancers across all stages. Based on the results from this investigation, we aim to provide clinicians with insights into the clinicopathological conditions that warrant HER2 investigation.

Methods: In this retrospective study, we conducted a comprehensive analysis of clinicopathological data from a cohort of 169 patients who underwent surgical treatment for gastric cancer between 2014 and 2022. The HER2 status was determined based on results from immunohistochemistry (IHC) and fluorescent in situ hybridization (FISH) techniques applied to gastric cancer pathology samples. Based on the HER2 positivity, the patients were classified into two distinct groups: (1) HER2-positive and (2) HER2-negative. The relationship between the clinicopathological variables, HER2 status, and overall survival (OS) was evaluated using chi-squared and Kaplan–Meier analyses. A statistical significance level of P <0.05 was applied to determine significant associations.

Results: According to the IHC analyses performed in our study population, 33 among 169 patients were HER2-positive (19.53%). Statistically significant factors related to HER2 positivity, such as male gender (P=0.009), pathological stage, N category, lymphovascular invasion status ([LVI] P=0.046), and proximal tumor location (P=0.015) were observed. In addition, OS was 40.49 (6.21) months in HER2-positive gastric cancer patients and 57.43 (3.48) months in HER2-negative gastric cancers (P=0.045).

Conclusion: Irrespective of the pathological stage, gastric cancer exhibited HER2 positivity at a ratio of 5:1. Among the clinicopathological findings, a significant correlation was observed between HER2 expression and gastric cancers characterized by aggressive features. Moreover, HER2 positivity was associated with an unfavorable prognosis in gastric cancer patients.

Keywords: clinicopathological features, gastric cancer, HER2, immunohistochemistry, prognostic factor

How to cite: Ceylan C, Angin YS, Baran NT, Kocaaslan H, Cengiz E, Güngörür O, Güzel M, Saglam K, Aydin C. Relationship between HER2 and clinicopathological data in gastric adenocarcinomas. J Surg Med. 2024;8(1):7-11.

Introduction

Gastric cancer, along with colorectal cancer, is among one of the most prevalent malignancies affecting the gastrointestinal tract. Despite a recent decrease in its incidence and associated mortality rates, GC continues to be associated with a dismal prognosis [1]. The advent of targeted therapies in clinical practice has shed light on their potential efficacy in the treatment of gastric cancers. Notably, anti-human epidermal growth factor receptor 2 (HER2) therapies, which are widely employed in breast cancer treatment, have emerged as a crucial component of targeted therapy options [2].

HER2, also known as ErbB2, belongs to the family of epidermal growth factor receptors (EGFR) located on *chromosome 17 (17q21)* [3]. HER2 mediates signal transduction that is involved with regulation of cell proliferation, differentiation, adhesion, and migration via tyrosine kinase autophosphorylation, a process that leads to the activation of downstream pathways [4].

HER2 overexpression and amplification have been extensively observed in various cancers, particularly breast cancer [5]. Its presence in gastric cancer was initially identified in 1986 [6]. The incidence of HER2 positivity in gastric cancer ranges from 7% to 34% [7,8]. Immunohistochemistry (IHC) is the primary method for assessing HER2 expression with fluorescence in situ hybridization (FISH) performed for confirmation when necessary. The association between HER2 status and prognosis in gastric cancer remains incompletely elucidated. Nonetheless, multicenter studies have consistently demonstrated that patients with HER2-positive gastric cancer (HPGC) exhibit lower overall survival (OS) rates compared to those with HER2-negative gastric cancer [9]. Furthermore, trastuzumab, one of the therapies used in the targeted approach, has shown efficacy in producing improvements in OS in advanced HER2-positive gastric and esophagogastric junction cancers [10]. In this study, we aimed to investigate HER2 positivity, which has a significant impact on survival and is a treatment target in gastric cancer patients. Also, we aimed to explore the association between HPGC and its clinicopathologic characteristics.

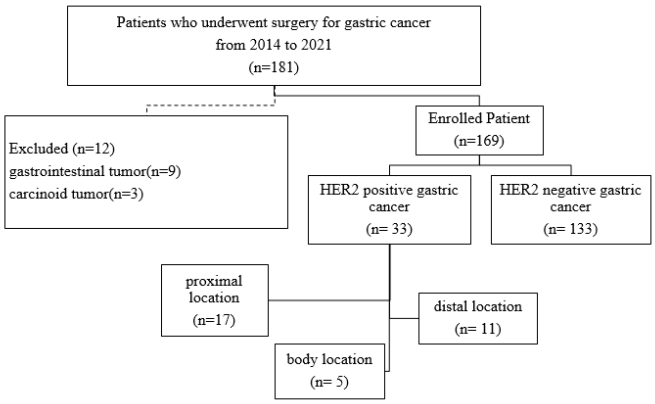
Materials and methods

Patient selection and study design

This study received approval from the Scientific Research and Publication Ethics Committee of İnönü University based on the established ethical guidelines. The approval was granted on February 8, 2022, and the assigned reference number for the study is 2022/2914. We retrospectively reviewed the data from 181 patients with gastric cancer who underwent total, completion total, or proximal gastrectomy between January 2014 and July 2021. Inclusion criteria were patients with histologically proven primary gastric adenocarcinoma regardless of pathological stage. Patients with limited electronic medical records were excluded from the study.

Among these, a total of 12 patients (nine gastrointestinal tumors, three carcinoid tumors) were excluded. Finally, 169 patients were eligible for the analysis (Figure 1).

Figure 1: Flowchart of the patient selection process.



Patient management

Following their surgical procedures, patients received either fluoropyrimidine-based or platinum-based chemotherapy with or without trastuzumab. Regular follow-up visits were conducted at 3, 6, 9, and 12 months after surgery. During these follow-up visits, patients underwent thoracoabdominal computed tomography (CT) scans every six months to screen for any signs of recurrence or metastasis. Additionally, endoscopy examinations were performed annually to monitor patients' conditions.

Another outcome measure in this study was the evaluation of the influence of HER2 status on OS. OS was defined as the period starting from the surgical intervention and ending with death from any cause. To ascertain the survival status of the patients, they were followed until July 2017.

Clinicopathological and survival data were retrieved from hospital medical records.

Evaluation of clinicopathological findings

Categorical and continuous clinicopathological data were collected and analyzed. Data on age (years), sex (male, female), Lauren's classification (intestinal type, diffuse type), tumor histology (differentiated, undifferentiated), T-stage (I–IV), N stage (0–3), pathological stage (1–4), tumor location (proximal, body, distal), lymphovascular invasion status (LVI) (absence, presence), perineural invasion status ([PNI] absence, presence), and HER2 status (positive, negative) were collected for each patient.

The World Health Organization (WHO) classification criteria and the eighth edition of the American Joint Committee on Cancer were used for pathological staging of gastric cancer [11].

Initially, the specimens were tested for HER2 expression based on IHC. IHC 3+ was defined as HER-2 positive. Those with IHC 0/1 were defined as HER-2 negative. Those with IHC 2+ were then evaluated based on FISH, and HER-2 expression was determined.

Statistical analysis

The sample size was determined using power analysis with the G-Power 3.1 software. Based on a power (1-β) of 0.80 and a confidence level of 95%, it was calculated that each group should have a sample size of 32. Therefore, the minimum total sample size for both groups was determined to be 64. Compliance of numerical data with normal distribution was checked using the Kolmogorov–Smirnov test. Continuous numerical variables were analyzed with the Mann–Whitney U test. The median, minimum, and maximum values of these variables were presented. A chi-squared analysis was performed for categorical variables. The

frequency and percentage values of these variables were presented. Univariate logistic regression analysis was performed for each variable by taking the variables with statistically significant p values in similar variables. Survival for HER2 status was calculated using the Kaplan–Meier method and log-rank test. *P*-value <0.05 was considered statistically significant. Analyses were performed using SPSS v23 (SPSS Inc., Chicago, IL, USA).

Results

Patient and sample characteristics

The median patient age was 63 (19–96) years, and the male-to-female ratio was 1.6:1. Thirty-three (19.53%) cases of HER-2 positivity gastric cancer were identified for which 17 (51.5%), 5 (15.2%), and 11 (33.3%) were found in proximal, body, and distal locations, respectively (Table 1, 2).

Based on the univariate analysis, statistically significant factors related to HER2 positive status, such as gender, pathological stage, N category, LVI, and tumor location, were found (Table 2). HER2 predominance in men (78.79%, 26/33) was detected. In addition, HER2 positivity was more commonly detected in proximal tumors (51.52%, 17/33). Most patients 84.02% (n=142) presented with advanced (T2 and above) tumors, and 21.83% (n=31) were HER2 positive. Additionally, among the 27 patients with early-stage GC, 7.41% (n=2) were HER2 positive, but this finding was not statistically significant (*P*=0.664) as shown in Table 2.

The differences between the two groups in terms of age, tumor size, PNI, Lauren's classification, tumor histology, and T category were not statistically significant (Table 2).

Survey analysis results

During follow-up, 59 of 169 (34.9%) patients died. The mean OS in all patients was 54.78 (3.18) months (95% CI: 48.55–61.01). The 3-year and 5-year survival rates were 27.2% and 12.4%, respectively. The mean OS in HER2-positive gastric cancer patients was 40.49 (6.21) months (95% CI: 28.32–52.05). However, the mean OS for HER2-negative gastric cancers was 57.43 (3.48) months (95% CI: 50.61–64.25). When survival rates were compared among groups, it could be observed that OS was better in HER2-negative gastric cancers (log-rank *P*=0.045, Hazard ratio [HR]=1.808 [1.004–3.255]) as shown in Table 3 and Figure 2.

Table 1: Patients demographics and tumor characteristics

Characteristics	Value n=169 n(%)
Gender	
Male	105(62.13)
Female	64(37.87)
Age(year)	
Median (range)	63(19–96)
Mean (SD)	62 (13)
Tumor Location	
Proximal	59(34.92)
Body	26(15.38)
Distal	84(49.70)
Histology	
Differentiated	104(61.54)
Undifferentiated	65(38.46)
Lymphovascular invasion	
Presence	135(79.88)
Absence	34(20.12)
Perineural Invasion	
Presence	108(63.9)
Absence	61(32.1)
T category	
1a	10(5.92)
1b	17(10.06)
2	9(5.32)
3	67(39.65)
4a	63(37.27)
4b	3(1.78)
N category	
0	41(24.26)
1	24(14.2)
2	27(15.98)
3a	38(22.49)
3b	39(23.07)
Pathological Stage	
1a	18(10.65)
1b	12(7.1)
2a	13(7.69)
2b	30(17.75)
3a	23(13.6)
3b	21(12.43)
3c	47(27.81)
4	5(2.97)
Tumor Size(cm)	
Median (range)	5.5(0.7-20)
Mean (SD)	5.85 (3.62)
Lauren's Classification	
Intestinal type	102(60.4)
Diffuse type	67(39.6)
HER2	
Positive	33(19.53)
Negative	136(80.47)

SD: standard deviation

Table 3: Surveillance analysis results in gastric cancers

	Mean (SD) OS (month)	HR	95%CI	*P*-value
HER2 Status		1.808	1.004-3.255	0.045
HER2-negative GC	57.43 (3.48)			
HER2-positive GC	40.49 (6.21)			
Total	54.78 (3.18)			

OS: Overall Survival, SD: standard deviation, HR: Hazard Ratio, CI: Confidence Interval, HER2: human epidermal growth factor receptor 2, GC: Gastric Cancer. *P*-value <0.05 was considered statistically significant.

Figure 2: Effect of human epidermal growth factor receptor 2 (HER2) on overall survival in gastric cancer.

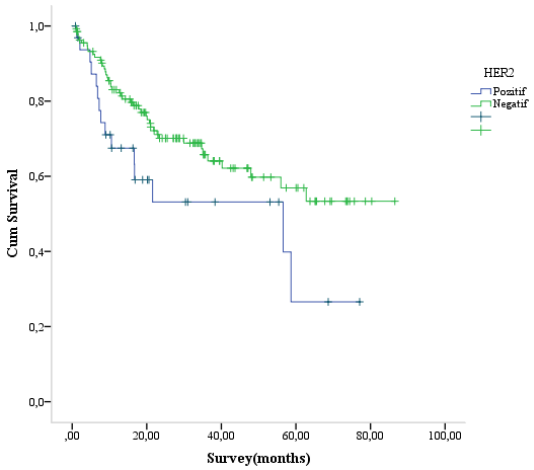


Table 2: Univariate analysis of risk factors for human epidermal growth factor receptor 2 (HER-2) positivity

	Characteristics	All patients n=169 n(%)	HER2-positive n=33 n(%)	OR	95% CI	*P*-value
Gender	Female	64(37.87)	7(21.21)	Reference		
	Male	105(62.13)	26(78.79)	2.486	1.088-6.601	0.009
Age(year)	<60	72(42.60)	12(36.36)	Reference		
	≥60	97(57.40)	21(63.64)	1.382	0.630-3.032	0.420
Tumor Location	Distal	84(49.70)	11(33.33)	Reference		
	Body	26(15.38)	5(15.15)	1.870	0.607-5.758	0.275
	Proximal	59(34.92)	17(51.52)	2.859	1.222-6.692	0.015
Histology	Differentiated	104(61.54)	7(78.79)	Reference		
	Undifferentiated	65(38.46)	26(21.21)	1.961	0.792–4.855	0.145
Lymphovascular Invasion	No Invasion	34(20.12)	7(21.21)	Reference		
	Invasion	135(79.88)	26(78.79)	4.769	1.082–21.031	0.046
Perineural Invasion	Invasion	108(63.9)	10(30.30)	Reference		
	No Invasion	61(32.1)	23(69.70)	1.380	0.608–3.132	0.441
T category	1a	10(5.92)	0(0)	Reference		
	1b	17(10.06)	2(6.06)	3.750	0.224–62.764	0.358
	2	9(5.32)	3(9.09)	1.000	0.063–15.988	1.000
	3	67(39.65)	15(45.45)	1.733	0.147–20.456	0.662
	4a	63(37.27)	12(36.36)	2.125	0.178–25.412	0.552
	4b	3(1.78)	1(3.03)	2.000	0.168–24.382	0.571
N category	0	41(24.26)	6(18.18)	Reference		
	1	24(14.2)	4(12.12)	3.646	1.238–10.735	0.019
	2	27(15.98)	2(6.06)	3.125	0.893–10.934	0.075
	3a	38(22.49)	6(18.18)	7.812	1.612–37.859	0.011
	3b	39(23.07)	15(45.45)	3.333	1.127–9.861	0.03
Pathological Stage	1a	18(10.65)	1(3.03)	Reference		
	1b	12(7.1)	4(12.12)	68.000	3.460–1336.268	0.005
	2a	13(7.69)	1(3.03)	8.000	0.658–97.311	0.103
	2b	30(17.75)	4(12.12)	48.000	2.404–958.237	0.011
	3a	23(13.6)	6(18.18)	26.000	2.287–295.637	0.009
	3b	21(12.43)	3(9.09)	11.333	1.048–122.549	0.046
	3c	47(27.81)	10(30.30)	24.000	1.952–295.061	0.013
	4	5(2.97)	4(12.12)	14.800	1.484–147.611	0.022
Tumor Size(cm)	≤5	87(51.48)	15(45.45)	Reference		
	>5	82(48.52)	18(54.55)	1.457	0.668–3.179	0.344
Lauren's Classification	Intestinal type	102(60.4)	20(60.60)	Reference		
	Diffuse type	67(39.6)	13(39.40)	0.987	0.453–2.149	0.974
Early Stage	Yes	27(15.98)	2(6.06)	Reference		
	No	142(84.02)	31(93.94)	1.261	0.443–3.589	0.664

OR: Odds Ratio, CI: Confidence Interval. *P*-value <0.05 was considered statistically significant.

Discussion

In many recent studies, several different factors have been revealed to be relevant to the relationship between HPGC and clinicopathological parameters (tumor location, LVI, hepatic metastasis, Lauren's Classification, age, gender, higher lymph node stage, and advanced staging) [9,12,13]. Although conditions such as hepatic metastasis and advanced stage can be explained by HER2 overexpression and amplification, intestinal histological type, low grade, and predominant localization of the cancer to the proximal stomach cannot be explained.

In this study consisting of a total of 169 gastric cancer patients, we identified several clinicopathological factors that were associated with HPGC based on the analysis: (1) male gender, (2) proximal tumor location, (3) higher lymph node stage, and (4) advanced staging. In addition to HER2 positive status is a poor factor for predicting survival in gastric cancer.

A meta-analysis demonstrated a significant association between HER-2 overexpression and overall survival in patients [14]. However, a study by Grabsch et al. [15] reported no relationship between HER-2 expression and prognosis. The role of HER2 expression as a prognostic factor has been confirmed in advanced gastric cancer, but it does not appear to affect disease-free survival and OS in early-stage gastric cancers [16]. Another study found no association between HER-2 status, clinical-pathological characteristics, and OS in early-stage gastric cancer [17]. In our study, we did not observe a correlation between HER2 and early-stage gastric cancers. However, regardless of the pathological stage, HPGC patients exhibited poorer survival. In conclusion, the impact of HER2 on overall survival remains controversial.

The Lauren classification categorizes gastric cancer into intestinal, diffuse, and mixed subtypes, and this classification has been recognized as an important prognostic factor in previous studies [18]. Specifically, patients with the intestinal subtype were found to have a higher likelihood of HPGC.

Several studies have reported a predominance of the intestinal subtype among HER2-positive patients [19,20]. However, in our study, we did not observe a significant relationship between HPGC and the intestinal subtype.

In a study conducted in South Korea, the rate of HER2 was 7.3%. As in our study, male gender, proximal tumor location, higher lymph node stage, and advanced pathological stages were found to be correlated with HPGC in the South Korean study [21]; however, no correlation was found between older age and the intestinal subtype in our study. In addition, the HPGC in our study was found to be 19.53%.

In addition to being a poor prognostic factor in many malignancies, LVI is also associated with metastatic disease, recurrence, and poor prognosis in gastric cancers [22]. Laboissiere et al. [23] found that LVI correlated with HER2 overexpression in their gastric cancer study. In our study, it was shown that LVI and HER2 were correlated.

Limitations

The primary limitation of our study is the lack of consensus regarding the determination of HER2 receptor status. As a result, HER2 positivity rates vary significantly across the existing literature. Another limitation is the retrospective nature of our study, which inherently carries potential biases and limitations. However, to overcome these limitations, it is crucial

to establish a pathology-related consensus regarding HER2 determination and gather data from well-designed randomized prospective studies for further validation and confirmation.

Conclusion

Similar to results from other studies in the literature, our study revealed an HER2 incidence rate of 19.53% in gastric cancers. Furthermore, HER2 positivity was found to be associated with unfavorable prognostic indicators, including high lymph node ratios, advanced stage, and LVI. Additionally, HER2 positivity was more commonly observed in males and proximal tumors. Surveillance analysis demonstrated that HER2-positive gastric cancer patients had a shorter OS, indicating that HER2 serves as a negative prognostic marker for gastric cancers. Given the availability of targeted therapy, assessing HER2 receptor status is recommended to guide clinicians in line with the identified factors.

References

1. Van Cutsem E, Sagaert X, Topal B, Haustermans K, Prenen H. Gastric cancer. Lancet. 2016 Nov 26;388(10060):2654-64. doi: 10.1016/S0140-6736(16)30354-3. Epub 2016 May 5. PMID: 27156933.
2. Meric-Bernstam F, Johnson AM, Dumbrava EEI, Raghav K, Balaji K, Bhatt M, et al. Advances in HER2-Targeted Therapy: Novel Agents and Opportunities Beyond Breast and Gastric Cancer. Clin Cancer Res. 2019 Apr 1;25(7):2033-41. doi: 10.1158/1078-0432.CCR-18-2275. Epub 2018 Nov 15. PMID: 30442682; PMCID: PMC6445731.
3. Tebbutt N, Pedersen MW, Johns TG. Targeting the ERBB family in cancer: couples therapy. Nat Rev Cancer. 2013 Sep;13(9):663-73. doi: 10.1038/nrc3559. Epub 2013 Aug 16. PMID: 23949426.
4. Choi B, Cha M, Eun GS, Lee DH, Lee S, Ehsan M, et al. Single-molecule functional anatomy of endogenous HER2-HER3 heterodimers. Elife. 2020 Apr 8;9:e53934. doi: 10.7554/eLife.53934. PMID: 32267234; PMCID: PMC7176432.
5. Vranić S, Bešlija S, Gatalica Z. Targeting HER2 expression in cancer: New drugs and new indications. Bosn J Basic Med Sci. 2021 Feb 1;21(1):1-4. doi: 10.17305/bjbms.2020.4908. PMID: 32530388; PMCID: PMC7861626.
6. Sakai K, Mori S, Kawamoto T, Taniguchi S, Kobori O, Morioka Y, et al. Expression of epidermal growth factor receptors on normal human gastric epithelia and gastric carcinomas. J Natl Cancer Inst. 1986 Nov;77(5):1047-52. PMID: 3464796.
7. Gravalos C, Jimeno A. HER2 in gastric cancer: a new prognostic factor and a novel therapeutic target. Ann Oncol. 2008 Sep;19(9):1523-9. doi: 10.1093/annonc/mdn169. Epub 2008 Apr 25. PMID: 18441328.
8. Hofmann M, Stoss O, Shi D, Büttner R, van de Vijver M, Kim W, et al. Assessment of a HER2 scoring system for gastric cancer: results from a validation study. Histopathology. 2008 Jun;52(7):797-805. doi: 10.1111/j.1365-2559.2008.03028.x. Epub 2008 Apr 18. PMID: 18422971.
9. Kurokawa Y, Matsuura N, Kimura Y, Adachi S, Fujita J, Imamura H, et al. Multicenter large-scale study of prognostic impact of HER2 expression in patients with resectable gastric cancer. Gastric Cancer. 2015 Oct;18(4):691-7. doi: 10.1007/s10120-014-0430-7. Epub 2014 Sep 16. PMID: 25224659.
10. Bang YJ, Van Cutsem E, Feyereislova A, Chung HC, Shen L, Sawaki A, et al. ToGA Trial Investigators. Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial. Lancet. 2010 Aug 28;376(9742):687-97. doi: 10.1016/S0140-6736(10)61121-X. Epub 2010 Aug 19. Erratum in: Lancet. 2010 Oct 16;376(9749):1302. PMID: 20728210.
11. Brierley JD, Gospodarowicz MK, Wittekind C. The TNM classification of malignant tumours. 8. Oxford: Wiley Blackwell; 2017.
12. Matsusaka S, Nashimoto A, Nishikawa K, Miki A, Miwa H, Yamaguchi K, et al. Clinicopathological factors associated with HER2 status in gastric cancer: results from a prospective multicenter observational cohort study in a Japanese population (JFMC44-1101). Gastric Cancer. 2016 Jul;19(3):839-51. doi: 10.1007/s10120-015-0518-8. Epub 2015 Aug 12. Erratum in: Gastric Cancer. 2016 Jul;19(3):1026. PMID: 26265390; PMCID: PMC4906061.
13. Tang D, Liu CY, Shen D, Fan S, Su X, Ye P, et al. Assessment and prognostic analysis of EGFR, HER2, and HER3 protein expression in surgically resected gastric adenocarcinomas. Onco Targets Ther. 2014 Dec 16;8:7-14. doi: 10.2147/OTT.S70922. PMID: 25565860; PMCID: PMC4274138.
14. Liang JW, Zhang JJ, Zhang T, Zheng ZC. Clinicopathological and prognostic significance of HER2 overexpression in gastric cancer: a meta-analysis of the literature. Tumour Biol. 2014;35(5):4849–58.
15. Grabsch H, Sivakumar S, Gray S, Gabbert HE, Müller W. HER2 expression in gastric cancer: Rare, heterogeneous and of no prognostic value - conclusions from 924 cases of two independent series. Cell Oncol. 2010;32(1-2):57-65. doi: 10.3233/CLO-2009-0497. PMID: 20208134; PMCID: PMC4619246.
16. Li H, Li L, Zhang N, Wang Z, Xu N, Linghu E, et al. Relationship between HER2 overexpression and long-term outcomes of early gastric cancer: a prospective observational study with a 6-year follow-up. BMC Gastroenterol. 2022 May 13;22(1):238. doi: 10.1186/s12876-022-02309-7. PMID: 35562663; PMCID: PMC9102633.
17. Fisher SB, Fisher KE, Squires MH 3rd, Patel SH, Kooby DA, El-Rayes BF, et al. HER2 in resected gastric cancer: Is there prognostic value? J Surg Oncol. 2014 Feb;109(2):61-6. doi: 10.1002/jso.23456. Epub 2013 Oct 10. PMID: 24122802.
18. Wang H, Xing XM, Ma LN, Liu L, Hao J, Feng LX, et al. Metastatic lymph node ratio and Lauren classification are independent prognostic markers for survival rates of patients with gastric cancer. Oncol Lett. 2018 Jun;15(6):8853-62. doi: 10.3892/ol.2018.8497. Epub 2018 Apr 13. PMID: 29844813; PMCID: PMC5958805.
19. Liu X, Xu P, Qiu H, Liu J, Chen S, Xu D, et al. Clinical utility of HER2 assessed by immunohistochemistry in patients undergoing curative resection for gastric cancer. Onco Targets Ther. 2016 Feb 26;9:949-58. doi: 10.2147/OTT.S100979. PMID: 27013889; PMCID: PMC4777257.
20. Ngaiza A, Vuhahula E, Yahaya J, Ndayisaba MC, Kawishe GJ, Grenert JP, et al. Evaluation of Human Epidermal Growth Factor Receptor 2 Expression in Gastric and Gastroesophageal Cancers in Tanzania. Arch Pathol Lab Med. 2022 Dec 1;146(12):1523-9. doi: 10.5858/arpa.2021-0394-OA. PMID: 35344993; PMCID: PMC9515243.
21. Choi S, Song JH, Lee S, Cho M, Kim YM, Kim HI, et al. Lymphovascular Invasion: Traditional but Vital and Sensible Prognostic Factor in Early Gastric Cancer. Ann Surg Oncol. 2021 Dec;28(13):8928-35. doi: 10.1245/s10434-021-10224-6. Epub 2021 Jun 1. PMID: 34075484.
22. Fujikawa H, Koumori K, Watanabe H, Kano K, Shimoda Y, Aoyama T, et al. The Clinical Significance of Lymphovascular Invasion in Gastric Cancer. In Vivo. 2020 May-Jun;34(3):1533-9. doi: 10.21873/invivo.11942. PMID: 32354959; PMCID: PMC7279808.
23. Laboissiere RS, Buzelin MA, Balabram D, De Brot M, Nunes CB, Rocha RM, et al. Association between HER2 status in gastric cancer and clinicopathological features: a retrospective study using whole-tissue sections. BMC Gastroenterol. 2015 Nov 4;15:157. doi: 10.1186/s12876-015-0384-1. PMID: 26530403; PMCID: PMC4632681.

J Surg Med. 2024;8(1):12-16.
https://doi.org/10.28982/josam.8020

JOSAM

Research article
https://jsurgmed.com/

Journal of Surgery and Medicine

e-ISSN: 2602-2079

Deep-learning-based diagnosis and grading of vesicoureteral reflux: A novel approach for improved clinical decision-making

Osman Ergün [1], Tekin Ahmet Serel [1], Sefa Alperen Öztürk [1], Hüseyin Bulut Serel [2], Sedat Soyupek [1], Burak Hoşcan [3]

[1] Süleyman Demirel University, Medical School, Department of Urology, Isparta, Turkey
[2] Getir Company, Software Engineering, Ankara, Turkey
[3] Bilim University, Vocational School of Health Services, Antalya, Turkey

ORCID of the author(s)

OE: https://orcid.org/0000-0001-7611-0933
TAS: https://orcid.org/0000-0002-2135-353X
SAÖ: https://orcid.org/0000-0003-4586-9298
HBS: https://orcid.org/0000-0002-8955-7942
SS: https://orcid.org/0000-0002-7345-3452
BH: https://orcid.org/0000-0002-7887-9575

Corresponding Author
Osman Ergün
Süleyman Demirel University Faculty of Medicine Hospital, Urology Academic Floor Secretariat, Eastern Campus, Çünür, Isparta, Turkey
E-mail: osmanergun77@gmail.com

Ethics Committee Approval
Ethical approval was not required, as the data was sourced from an open-access archive. Informed consent was not required due to design of the study. This study does not contain identifying information of the patients. However, it is crucial to emphasize that all data used in this study has been anonymized and treated with the utmost confidentiality.

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

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

Published
2024 January 17

Abstract

Background/Aim: Vesicoureteral reflux (VUR) is a condition that causes urine to flow in reverse, from the bladder back into the ureters and occasionally into the kidneys. It becomes a vital cause of urinary tract infections. Conventionally, VUR's severity is evaluated through imaging via voiding cystourethrography (VCUG). However, there is an unresolved debate regarding the precise timing and type of surgery required, making it crucial to classify VUR grades uniformly and accurately. This study's primary purpose is to leverage machine learning, particularly convolutional neural network (CNN), to effectively identify and classify VUR in VCUG images. The aspiration is to diminish classification discrepancies between different observers and to create an accessible tool for healthcare practitioners.

Methods: We utilized a dataset of 59 VCUG images with diagnosed VUR sourced from OpenI. These images were independently classified by two seasoned urologists according to the International Reflux Classification System. We utilized TensorFlow, Keras, and Jupyter Notebook for data preparation, segmentation, and model building. The CNN Inception V3 was employed for transfer learning, while data augmentation was used to improve the model's resilience.

Results: The deep-learning model attained exceptional accuracy rates of 95% and 100% in validation and training, respectively, after six cycles. It effectively categorized VUR grades corresponding to the global classification system. Matplotlib tracked loss and accuracy values, while Python-based statistical analysis assessed the model's performance using the F1-score.

Conclusion: The study's model effectively categorized images, including those of vesicoureteral reflux, which has significant implications for treatment decisions. The application of this artificial intelligence model may help reduce interobserver bias. Additionally, it could offer an objective method for surgical planning and treatment outcomes.

Keywords: vesicoureteral reflux, voiding cystourethrography, artificial intelligence, convolutional neural network, deep learning

How to cite: Ergün O, Serel TA, Öztürk SA, Serel HB, Soyupek S, Hoşcan B. Deep-learning-based diagnosis and grading of vesicoureteral reflux: A novel approach for improved clinical decision-making. J Surg Med. 2024;8(1):12-16.

Introduction

Vesicoureteral reflux (VUR) is a condition where urine flows backward from the bladder into the ureters and, occasionally, the kidneys. It is a significant contributory factor to urinary tract infections and primarily arises from anatomical or functional abnormalities [1]. The ureter does not close during voiding or under conditions of high intravesical pressure, such as neurogenic bladder, urethral stenosis, and posterior urethral valve [2]. Untreated, VUR may result in complications like reflux nephropathy, kidney damage leading to renal failure, end-stage renal disease, or growth retardation in children [3].

Voiding cystourethrography (VCUG) is a common radiological test employed to detect vesicoureteral reflux (VUR). VCUG enables comprehensive insights into the presence, absence, and severity of VUR, as defined by international standards [4-7]. It is often ordered by specialists such as urologists, pediatric urologists, pediatricians, pediatric nephrologists, and pediatric surgeons. For patients with persistent high-grade reflux (grades 4/5), surgical correction should be considered. However, consensus is lacking on the timing and selection of surgical methods. Reimplantation is generally preferable for higher reflux grades, whereas endoscopic injections can produce satisfactory results for lower grades. Thus, standardizing and accurately grading VUR is crucial. It strongly impacts treatment choices and promotes clear communication among health professionals.

Machine learning, a facet of artificial intelligence, involves teaching a computer to develop programs based on given data and anticipated results. This approach is increasingly prevalent in the biomedical field. While humans can easily understand the context and importance of an image, translating this ability to machine comprehension is a complex task. The convolutional neural network (CNN), a prominent machine learning technique, has made significant strides in the field of medical imaging. CNN architectures are widely employed in image detection and classification [8]. In simple terms, an application, often shortened to 'app', is a type of software or collection of programs designed to assist users in completing specific tasks. The use of CNNs holds notable implications for healthcare professionals in clinical practice.

This research focused on using deep-learning methods to identify and categorize vesicoureteral reflux (VUR) in images from Voiding Cystourethrogram (VCUG) studies. The intention was to decrease discrepancies in classification between observers and to create a corresponding application.

Materials and methods

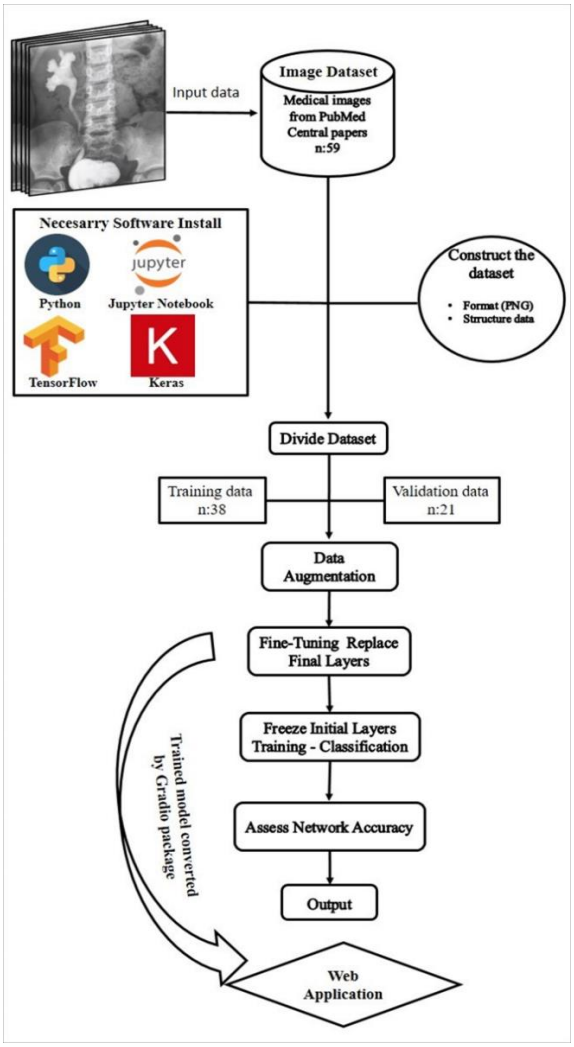
Data collection

This study used a dataset of 59 images, all previously diagnosed with VUR using VCUG. The images were sourced from OpenI, a PubMed Central online medical image archive. Because the data was obtained from an open-access archive, there was no need for ethical approval. All analyses were carried out on this readily available data.

Image classification

Two seasoned urologists (OE, SAÖ), with a decade of experience each, classified all images based on the International Reflux Classification System, independently and unaware of the other's determinations. In instances of conflict, a senior urologist (SS) with 20 years of experience was called in. Images causing unresolved disagreements were omitted from the study. The study's flow is illustrated in Figure 1.

Figure 1: The flowchart of the study.



Data preparation

For data analysis, we used TensorFlow, Keras, and Jupyter Notebook. Jupyter Notebook, an open-source web application, enabled us to create and share documents containing text and live code [9,10].

Data splitting

The images were divided at random into two groups: a training group of 38 cases and a validation group of 21 cases. This method aimed to represent the dataset accurately and evenly [11,12].

Model construction

A Jupyter Notebook environment was set up using prerequisites from the Keras library. We used a convolutional neural network model called Inception V3 for this task [13]. The Cross-Entropy loss function was used for classification, and we applied the Adam optimizer with standard settings [14].

Transfer learning

We designed a two-step transfer learning strategy, which included freezing and training the final layers. We then added extra layers with random initialization to the pre-trained Inception V3 model, originally trained on ImageNet. Finally, we fine-tuned the model with a weight of 0.0001.

Data augmentation

The Keras Image Data Generator was used for image enhancement to improve the model's strength. This generator was set up with the training directory, file requirements, picture dimensions, and batch size.

Model deployment

To develop an easy-to-use application, we used the Gradio package in Python. This facilitated the transformation of the trained model into a web application. You can access the web application at https://huggingface.co/spaces/Ragio/VUR_grade_prediction.

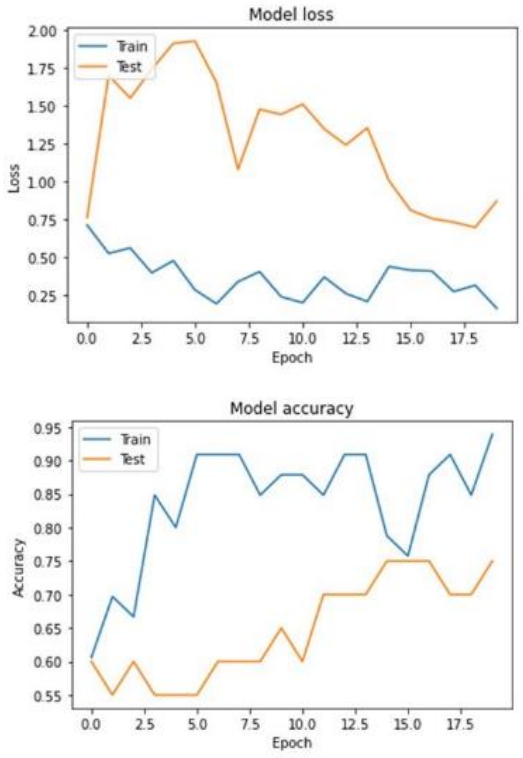
Statistical analysis

All statistical analyses were conducted using Python 3.6. The model's performance was evaluated using the F1-score, which ranges from 0 to 1, with 1 demonstrating flawless positive predictive value and sensitivity.

Results

Once the code is run, the model begins its training process. After six epochs, the model yields a training accuracy of 100% and a validation accuracy of 95%. The validation accuracy is lower due to the smaller number of images (n=21) in this group. If both validation and training values decrease, it demonstrates effective learning by the model. Matplotlib is a Python package utilized for plotting and generating figures in multiple formats. It captures the loss and accuracy values in arrays (Figure 2).

Figure 2: Loss and accuracy arrays of the created model.

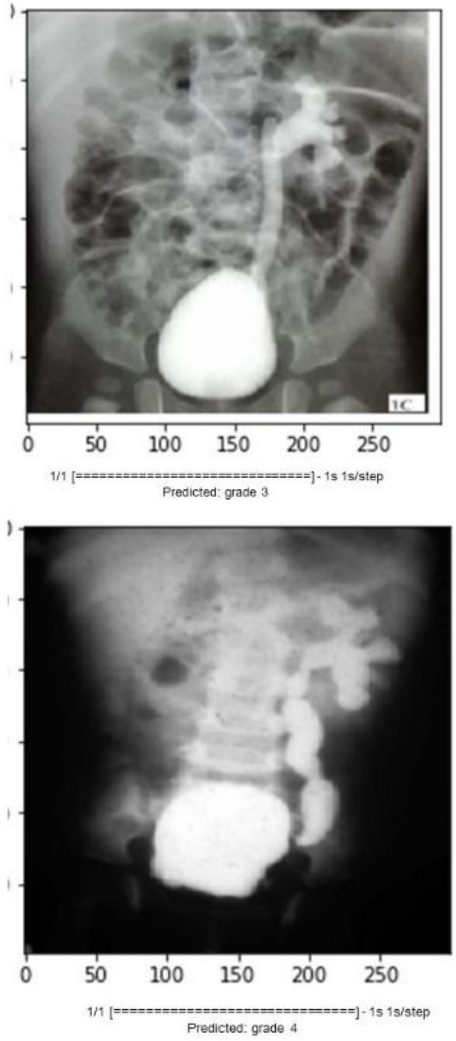


The hold-out test is employed to gauge the efficiency of our trained model as well as to evaluate the training and validation of the images. Keras, a Python-based deep-learning library, is utilized in training models within neural networks and executing the data generator on a batch of test images. It operates on a directory of issues using a for-loop or evaluates one point at a time.

Our model successfully estimated all vesicoureteral reflux (VUR) grades according to the International Reflux Classification System. Specifically, it categorized: Grade I as the presence of contrast medium solely in the non-dilated ureter; Grade 2 as the presence of contrast in both the ureter and the renal pelvis, with no significant dilatation; Grade 3 as mild dilation in the ureter and the pyelocalyceal system; Grade 4 as tortuosity and moderate dilation in the ureter, along with blunted renal fornices; and Grade 5 as tortuosity and severe dilation in the ureter, dilated pyelocalyces, loss of fornices, and papillary impression.

Figure 3 visually represents the estimated grade 4 and grade 5 VUR. The potential to distinguish between these two grades is indicated with numbers in parentheses, which range from 0 to 1.

Figure 3: Estimated values of grade III and IV vesicoureteral reflux.



Discussion

Vesicoureteral reflux (VUR) is a perilous disease that poses significant risks to a child's kidney health, alongside various other health complications [15]. Early diagnosis of this disease is critical to prevent recurring urinary tract infections and further kidney damage in children [16]. Urologists universally advise that all children with recurrent febrile urinary tract infections undertake Voiding Cystourethrograms (VCUGs). The VCUG's findings allow them to gauge the severity of VUR, enabling them to prescribe accurate treatments. This can range from frequent follow-ups, prophylactic antibiotics, or even necessitate endoscopic or surgical procedures.

Studies have shown that the combination of human analysis and deep-learning systems is more effective than using either independently [17]. The employment of CNNs in reviewing

radiological images can decrease the workload of healthcare professionals and lower clinical practice costs. Such technology can also reduce variability in observation and promote uniform interpretation. As a result, integrating deep learning into decision-making processes can significantly benefit specialists. Developers are creating deep-learning algorithms to analyze medical images, such as ultrasound and MRI scans, to help identify and categorize urological diseases [18-24]. Additionally, tools powered by deep learning are under examination for their potential to assist in surgical planning and decision-making in the field of urology.

Technologies play a crucial role in diagnosing and managing urological diseases. However, their accuracy and therapeutic effectiveness are often reliant on the operator's skills and the urologist's expertise. Pediatric urologists demonstrate an increased interest in standardizing classification methods and reporting in the pediatric urology field. Prenatal hydronephrosis, seen in about 5% of pregnancies, commonly leads to urology clinic referrals. Recognizing this, Lorenzo et al. [25] developed a deep-learning model to predict the need for surgical intervention in children diagnosed with hydronephrosis during prenatal screening. This early study offers substantial potential for improving clinical decision-making by identifying which patients are more likely to require surgery. Another study reported that deep learning could accurately detect VUR and hydronephrosis [26]. This underscores the strength of deep learning in quickly developing an accurate differentiation algorithm for recognizing hydronephrosis and VUR using minimal code and training cases.

Serrano-Durba et al. [27] developed a deep-learning model to predict the results of endoscopic treatment for VUR. Compared to traditional statistical methods, the deep-learning model proved superior in all evaluated variables, including sensitivity, specificity, and positive and negative predictive values. A related study involved 96 children with VUR and aimed to generate a deep-learning model for predicting the outcomes of various VUR treatments. The authors observed that deep learning outperformed traditional statistical methods [28]. The model accurately predicted VUR resolution and suggested that deep learning could potentially enhance traditional methods for more precise clinical outcome predictions.

There is no substantial evidence showing significant benefits of correcting persistent, low-grade reflux (grades I-III) when there are no symptoms and kidney function remains normal. Only those enduring persistent high-grade reflux (grades IV/V) should contemplate surgical correction. Reimplantation tends to yield better results than endoscopic corrections for higher reflux grades. Hence, patients with persistent high-grade reflux should be offered reimplantation, while endoscopic correction might be more fitting for lower reflux grades.

Current data insinuates that about one-third of Voiding Cystourethrogram (VCUG), results display inconsistent grading among clinicians, especially with moderate (grade 3-4) Vesicoureteral Reflux (VUR) [29]. Khondker et al. [30] developed a deep-learning model to gauge the reliability of reflux grading by assessing VCUGs for four features: ureteral tortuosity, proximal, distal, and maximum ureteral dilatation. This feature set was used to train the model to predict VUR grades.

The team reported that the developed model determined VUR grades with human-like accuracy, and there was a strong correlation between VUR grade and the four features mentioned above. Unlike the International Reflux Classification System, their model ignored the appearances of renal calyces in grade classification, posing a notable limitation. Additionally, the use of data obtained by measurements defined by established mathematical relationships might pose a disadvantage over the International Reflux Classification System.

Our study underscores the accurate recognition of VUR grades through the application of a deep-learning model, indicating the model's ability to correctly classify and categorize images, particularly those with grade 3 and 4 VUR. This accuracy is consequential as it can influence treatment decisions. The use of deep learning also curbs interobserver bias, helps cut costs, and reduces patients' radiation exposure. In addition, it facilitates objective surgical planning and the attainment of treatment goals.

We have developed a deep-learning model and an accompanying web application. This app enables healthcare professionals to expedite image interpretation, thereby speeding up diagnosis times. Moreover, it lessens healthcare workers' workload and enhances patient care quality. This application also proves useful for preliminary diagnosis in non-specialist environments.

We have converted our deep-learning model into an interactive web application using Gradio that is accessible to users worldwide. This user-friendly interface lets us integrate our model smoothly into a web application, providing real-time predictions based on data users input. Gradio's customization options and user-friendly design can create a smooth user experience, enabling easy interaction with our model. The web application is unrestricted, providing open access to all users.

Limitations

This study has several limitations. First, the number of images available for analysis was limited. However, we used a novel method that includes stock images from OpenI to construct the training set for the artificial intelligence. This innovative approach greatly improves the data's generalizability. Second, it is important to note that reflux can occur during both bladder filling and voiding. Reflux during voiding has a higher chance of resolution than during bladder filling. Regrettably, due to the nature of our data source, we are unable to determine whether the observed reflux in the images occurred during the bladder filling or voiding phase.

Conclusions

Vesicoureteral Reflux (VUR) is a serious disease that can cause significant mortality and morbidity, making early and accurate diagnosis crucial, particularly in pediatric patients. In our study, we developed a deep-learning model and application aimed at diagnosing and grading VUR in voiding cystourethrography (VCUG) images. Our model proves highly accurate in both diagnosis and staging of this disease, supported by a user-friendly web application that we also developed.

References

1. Cetin N, Gencler A, Kavaz Tufan A. Risk factors for development of urinary tract infection in children with nephrolithiasis. J Paediatr Child Health. 2020;56(1):76-80.
2. Weitz M, Schmidt M. To screen or not to screen for vesicoureteral reflux in children with ureteropelvic junction obstruction: a systematic review. Eur J Pediatr. 2017;176(1):1-9.
3. Wadie GM, Moriarty KP. The impact of vesicoureteral reflux treatment on the incidence of urinary tract infection. Pediatr Nephrol. 2012;27(4):529-38.
4. Siomou E, Giapros V, Serbis A, Makrydimas G, Papadopoulou F. Voiding urosonography and voiding cystourethrography in primary vesicoureteral reflux

associated with mild prenatal hydronephrosis: a comparative study. Pediatr Radiol. 2020;50(9):1271-6.
5. Baydilli N, Selvi I, Pinarbasi AS, Akinsal EC, Demirturk HC, Tosun H, et al. Additional VCUG-related parameters for predicting the success of endoscopic injection in children with primary vesicoureteral reflux. J Pediatr Urol. 2021;17(1):68.e1-68.e8.
6. Schaeffer AJ, Greenfield SP, Ivanova A, Cui G, Zerin JM, Chow JS, et al. Reliability of grading of vesicoureteral reflux and other findings on voiding cystourethrography. J Pediatr Urol. 2017;13(2):192-8.
7. Çelebi S, Özaydın S, Baştaş CB, Kuzdan Ö, Erdoğan C, Yazıcı M, et al. Reliability of the Grading System for Voiding Cystourethrograms in the Management of Vesicoureteral Reflux: An Interrater Comparison. Adv Urol. 2016;2016:1684190.
8. Ilhan HO, Sigirci IO, Serbes G, Aydin N. A fully automated hybrid human sperm detection and classification system based on mobile-net and the performance comparison with conventional methods. Med Biol Eng Comput. 2020;58(5):1047-68.
9. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. TensorFlow: large-scale machine learning on heterogeneous distributed systems V2 (Version 2). 2016, March 16. http://arxiv.org/abs/1603.04467.
10. Kluyver T, Ragan-Kelley B, Pérezc F, Granger B, Bussonnier M, Frederic J, et al. Jupyter Notebooks – a publishing format for reproducible computational workflows. In: Loizides F, Schmidt B, eds. Positioning and Power in Academic Publishing: Players, Agents and Agendas. Amsterdam, IOS Press; 2016:87-90. doi: 10.3233/978-1-61499-649-1-87.
11. Johnson J. Cnn-benchmarks. Accessed 13 March 2023. https://github.com/jcjohnson/cnn-benchmarks/;2023.
12. Wang S, Summers RM. Machine learning and radiology. Med Image Anal. 2012;16(5):933-51.
13. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE; 2016:2818-26.
14. Kingma DP, Ba J. Adam: a method for stochastic optimization. In: Bengio Y, LeCun Y, eds. Proceedings of the 3rd International Conference on Learning Representations. San Diego, CA: USA; 2015:1-15. https://arxiv.org/abs/1412.6980
15. Meena J, Hari P. Vesicoureteral reflux and recurrent urinary tract infections. Asian J. Pediatric Nephrol. 2019;2:61-70.
16. De Palma D. Radionuclide Tools in Clinical Management of Febrile UTI in Children. Semin Nucl Med. 2020;50(1):50-5.
17. Fazal MI, Patel ME, Tye J, Gupta Y. The past, present and future role of artificial intelligence in imaging. Eur J Radiol. 2018;105:246-50.
18. Batuello JT, Gamito EJ, Crawford ED, Han M, Partin AW, McLeod DG, et al. Artificial neural network model for the assessment of lymph node spread in patients with clinically localized prostate cancer. Urology. 2001;57(3):481-5.
19. Poulakis V, Witzsch U, de Vries R, Emmerlich V, Meves M, Altmannsberger HM, et al. Preoperative neural network using combined magnetic resonance imaging variables, prostate-specific antigen, and Gleason score for predicting prostate cancer biochemical recurrence after radical prostatectomy. Urology. 2004;64(6):1165-70.
20. Wu S, Chen X, Pan J, Dong W, Diao X, Zhang R, et al. An artificial intelligence system for the detection of bladder cancer via cystoscopy: a multicenter diagnostic study. J Natl Cancer Inst. 2022;114(2):220–7.
21. Fenstermaker M, Tomlins SA, Singh K, Wiens J, Morgan TM. Development and Validation of a Deep-learning Model to Assist with Renal Cell Carcinoma Histopathologic Interpretation. Urology. 2020;144:152-7.
22. Hobbs KT, Choe N, Aksenov LI, Reyes L, Aquino W, Routh JC, et al. Machine Learning for Urodynamic Detection of Detrusor Overactivity. Urology. 2022;159:247-54.
23. Katz JE, Abdelrahman L, Nackeeran S, Ezeh U, Visser U, Deane LA. The Development of an Artificial Intelligence Model Based Solely on Computer Tomography Successfully Predicts Which Patients Will Pass Obstructing Ureteral Calculi. Urology. 2023;174:58-63.
24. Çelik Ö, Aslan AF, Osmanoğlu UÖ, Çetin N, Tokar B. Estimation of renal scarring in children with lower urinary tract dysfunction by utilizing resampling technique and machine learning algorithms. J Surg Med. 2020;4(7):573-7.
25. Lorenzo AJ, Rickard M, Braga LH, Guo Y, Oliveria JP. Predictive Analytics and Modeling Employing Machine Learning Technology: The Next Step in Data Sharing, Analysis, and Individualized Counseling Explored with a Large, Prospective Prenatal Hydronephrosis Database. Urology. 2019;123:204-9.
26. Serel A, Ozturk SA, Soyupek S, Serel HB. Deep Learning in Urological Images Using Convolutional Neural Networks: An Artificial Intelligence Study. Turk J Urol. 2022 Jul;48(4):299-302.
27. Serrano-Durba A, Serrano AJ, Magdalena JR, Martín JD, Soria E, Domínguez C, et al. The use of neural networks for predicting the result of endoscopic treatment for vesico-ureteric reflux. BJU Int. 2004;94:120–2.
28. Seckiner I, Seckiner SU, Erturhan S, Erbagci A, Solakhan M, Yagci F. The use of artificial neural networks in decision support in vesicoureteral reflux treatment. Urol Int. 2008;80(3):283-6. doi: 10.1159/000127342.
29. Khondker A, Kwong JCC, Yadav P, Chan JYH, Singh A, Skreta M, et al. Multi-institutional Validation of Improved Vesicoureteral Reflux Assessment With Simple and Machine Learning Approaches. J Urol. 2022 Dec;208(6):1314-22.
30. Khondker A, Kwong JCC, Rickard M, Skreta M, Keefe DT, Lorenzo AJ, et al. A machine learning-based approach for quantitative grading of vesicoureteral reflux from voiding cystourethrograms: Methods and proof of concept. J Pediatr Urol. 2022 Feb;18(1):78.e1-78.e7.

J Surg Med. 2024;8(1):17-21.
https://doi.org/10.28982/josam.7799

JOSAM

Journal of Surgery and Medicine

e-ISSN: 2602-2079

Research article
https://jsurgmed.com/

Assessment of maternal and fetal outcomes according to induction methods following negative oxytocin challenge test

Okan Aytekin [1], Sezin Ertürk Aksakal [2], Mehmet Metin Altay [3]

[1] Department of Gynecologic Oncology, Ankara City Hospital, University of Health Sciences, Ankara, Turkey
[2] Department of Obstetrics and Gynecology, Etlik Zubeyde Hanim Women's Health Training and Research Hospital, University of Health Sciences, Ankara, Turkey
[3] Department of Perinatology, Etlik Zubeyde Hanim Women's Health Training and Research Hospital, University of Health Sciences, Ankara, Turkey

ORCID of the author(s)

OA: https://orcid.org/0000-0002-6430-4607
SEA: https://orcid.org/0000-0002-4418-7319
MMA: https://orcid.org/0000-0001-6772-858X

Corresponding Author
Okan Aytekin
Department of Gynecologic Oncology, Ankara City Hospital, University of Health Sciences, Ankara, Turkey
E-mail: okanaytekin13@hotmail.com

Ethics Committee Approval
The study was approved by the Etlik Zubeyde Hanim Women's Health Training and Research Hospital Local Ethics Committee (Approval Date/No: February 22, 2017/02).
All procedures in this study involving human participants were performed in accordance with the 1964 Helsinki Declaration and its later amendments.

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

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

Previous Presentations
The study is published as a preprint in Research Square (https://doi.org/10.21203/rs.3.rs-1232960/v1).

Published
2024 January 22

Abstract

Background/Aim: There is insufficient information about how long fetal well-being will last after a negative oxytocin challenge test (OCT) and the factors affecting this process. We aim to evaluate maternal and perinatal outcomes in high-risk patients who had negative OCTs and to investigate the effects of methods of induction on the development of fetal distress.

Methods: The study was designed as a retrospective cohort study. Data of patients who were hospitalized in the perinatal intensive care unit due to high-risk pregnancies between January 2016 and December 2016 were reviewed retrospectively. The patient's gestational age, gravidity, parity, and body mass index (BMI), risk factors leading to the OCT, labor induction methods used following a negative OCT, time from negative OCT to delivery, mode of delivery, and indications for cesarean section were recorded. In addition, data regarding fetal sex, birth weight, birth height, labor complications, Apgar scores at minutes 1 and 5, admission to the neonatal intensive care unit (NICU), indications for NICU admission, length of NICU stay, and stillbirth were also recorded.

Results: OCT was performed on 551 patients and was negative in 447 patients. Among patients with a negative OCT, labor induction was preferred in 427 (95.5%) patients. When fetal distress development was assessed according to the induction method used following a negative OCT, fetal distress developed in 9.1% of 427 patients who underwent labor induction.

Conclusion: When outcomes were considered in pregnant women with a negative OCT, it was observed that there were no fetal deaths and a limited number of newborns with low Apgar scores. Further randomized studies are needed to draw definitive conclusions.

Keywords: fetal outcomes, labor induction, oxytocin challenge test, maternal outcomes

How to cite: Aytekin O, Aksakal SE, Altay MM. Assessment of maternal and fetal outcomes according to induction methods following negative oxytocin challenge test. J Surg Med. 2024;8(1):17-21.

Introduction

Detection of fetal hypoxia during labor to minimize fetal death and neurologic sequelae related to fetal asphyxia is highly important. Electronic fetal monitoring (non-stress test [NST]), contraction stress test (CST), fetal biophysics profile, amniotic fluid index, Doppler sonography, fetal scalp blood testing, umbilical cord blood gas analysis, and neonatal Apgar scores are commonly used parameters in the assessment of intrauterine fetal state and distress [1–7]. The oxytocin challenge test (OCT) is one of the CST methods and is commonly used to monitor the fetus during the antepartum period. OCT is a test used to evaluate fetal well-being based on uterine contractions in suspected placental insufficiency. The negative OCT test was defined as the presence of accelerations in fetal heart rate, normal variability, and absence of slowdown in uterine contractions [8]. Freeman et al. [9] used the CST for follow-up in 679 post-term cases and observed no perinatal mortality. In other studies, it was shown that the perinatal mortality rate was lower than 0.1% during follow-up within the first week following a negative CST [10–13].

Delivery is induced due to maternal (preeclampsia, cardiac or renal disease), fetal (intrauterine growth retardation), or combined causes (uncontrolled diabetes mellitus, premature rupture of membrane, or post-term pregnancy) [14]. There are several methods of induction. However, no method has been shown to be superior to others [15]. Our study aimed to evaluate maternal and perinatal outcomes in high-risk patients (preeclampsia, intrauterine growth retardation, gestational diabetes mellitus, post-term pregnancy, and oligohydramnios, among others.) who had negative OCTs. The secondary aim was to investigate the effects of methods of induction on the development of fetal distress.

Materials and methods

The hospital records of all patients hospitalized in the perinatal intensive care unit due to high-risk pregnancies between January 2016 and December 2016 were reviewed retrospectively. There were 950 high-risk pregnancies. Of these, 551 patients underwent OCT and met the patient selection criteria. A total of 447 individuals with negative test results were included in the study. This study was approved by the Ethical Committee of the University of Health Sciences Turkey, Etlik Zübeyde Hanim Women's Health Training and Research Hospital (Approval Date/No: February 22, 2017/02).

Patient selection

Patients who had ≥34 weeks of gestation, singleton pregnancy with vertex presentation, and a negative OCT result were included in the study. Patients who had a history of cesarean delivery, uterine surgery, suspected cephalopelvic disproportion, inconclusive and hyper-stimulated or suspected OCT results, and a fetus with major congenital anomalies were excluded. The research included patients who had at least one of the risk factors listed below.

The following definitions were used when assessing risk factors:

- Oligohydramnios was defined as amnion fluid index <50 mm,
- Polyhydramnios was defined as the deepest vertical pocket >80 mm,
- The decreased fetal movement was defined as experiencing fewer or no fetal movements by the mother,
- Intrauterine growth retardation was defined as estimated fetal weight <3rd percentile according to the gestational week or confirmation by birth weight,
- Non-reactive NST was defined as lacking fetal heart rate acceleration at a 40-min period in NST monitoring,
- Preterm pregnancy was defined as gestational age between 34 – 36 6/7 weeks,
- Term pregnancy was defined as gestational age between 37 – 40 6/7 weeks,
- Post-term pregnancy was defined as gestational age ≥41 weeks,
- Gestational diabetes mellitus (GDM) was defined as the presence of elevation in at least two values by 100 g glucose and in at least one value by 75 g glucose in an oral glucose tolerance test,
- Preeclampsia was defined as arterial blood pressure >140/90 mm Hg and the presence of at least one of the following: proteinuria, thrombocytopenia, hepatic dysfunction, pulmonary edema, or cerebral and visual symptoms [16].

In the OCT, ten units of oxytocin in 500 milliliters (mL) of normal saline were given at a rate of 15 mL/hour using an infusion pump. The dose was doubled at 20-minute (min) intervals until achieving three uterine contractions with sufficient intensity in a 10-min period. The OCT was considered negative in patients who showed three uterine contractions with moderate intensity lasting 40–60 seconds but not late deceleration [17].

If an induction of labor is decided following an OCT, the appropriate method was selected according to the patients' Bishop scores. In our clinic, oxytocin (SYNPITAN®, DEVA, Turkey), Dinoproston (PROPESS Ovul®, FERRING, Switzerland), transcervical balloon catheter (COOK® Cervical Ripening Balloon, Cook Medical, Ireland), and Foley balloon catheter (Latex Foley Catheter®, Weel Lead Medical, China) were used for inducing labor. Ten units of oxytocin (SYNPITAN®) in 500 mL normal saline were given at a rate of 15 mL/h using an infusion pump; the dose was doubled at 20-min intervals until achieving three uterine contractions with sufficient intensity in a 10-min period. Dinoprostone was given via a slow-release vaginal delivery system (PROPESS Ovule®). It was removed in patients who achieved active labor, those with membrane rupture, and those who failed to induce labor within 12 hours. A transcervical balloon catheter and a Foley balloon catheter were administered through the cervix, as shown in studies in the extra-amniotic region [18,19].

The patient's gestational age, gravidity, parity, and body mass index (BMI), risk factors leading to the OCT, labor induction methods used following a negative OCT, time from negative OCT to delivery, mode of delivery, and indications for cesarean section were recorded. In addition, data regarding fetal sex, birth weight, birth height, labor complications, Apgar scores at minutes 1 and 5, admission to the neonatal intensive care unit (NICU), indications for NICU admission, length of NICU stay, and stillbirth were also recorded.

Statistical analysis

All statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS) for Windows version 22.0. Descriptive values are expressed as arithmetic mean (standard deviation), median, and percent. The Chi-square test

was used to analyze independent quantitative data, and Fisher's exact test was used when terms for the Chi-square test were inappropriate. A *P*-value <0.05 was considered statistically significant.

Results

It was found that the OCT was performed on 551 patients between January 2016 and December 2016. It was seen that the OCT was negative in 447 (81.1%) patients, positive in 50 patients (9.1%), and inconclusive in 54 (9.8%) patients. It was found that the mean maternal age was 25.7 (5.1) years, and the mean gestational age was 39.2 (1.5) weeks. The mean gravidity was 1.8 (1.1), and the mean BMI was 29.1 (4.2) kg/m^2 in 447 patients with a negative OCT.

When the induction methods used were assessed, among patients with a negative OCT, it was seen that labor induction was preferred in 427 (95.5%) patients, and spontaneous delivery occurred in 20 patients (4.5%). In addition, it was found that labor induction was performed using dinoprostone in 237 (53%) patients, oxytocin alone in 117 (26.2%) patients, Cook balloon catheter with oxytocin inductions in 47 (10.5%) patients, and Foley balloon catheter with oxytocin inductions in 26 (5.8%) patients (Table 1).

Table 1: Labor induction methods used in patients with a negative OCT

Labor induction method used	n	%
No induction	20	4.5%
Labor induction	427	95.5%
Dinoprostone	237	53.0%
Oxytocin	117	26.2%
Cook Balloon + Oxytocin	47	10.5%
Foley Balloon + Oxytocin	26	5.8%

Of the patients, 310 (72.6%) gave birth via vaginal delivery (VD) and 117 (27.4%) via cesarean section (C/S). When indications for C/S were assessed, it was found that C/S was performed due to fetal distress in 39 (33.3%) patients, non-progressive labor in 74 (63.2%) patients, placental abruption in two (1.7%) patients, umbilical cord prolapse in one (0.9%) patient, and risk for chorioamnionitis in one (0.9%) patient.

When the causes of fetal distress following OCT were assessed according to the risk factors prompting OCT, it was seen that C/S was performed in 10 (5%) patients who underwent OCT due to oligohydramnios, comprising 25.6% of all cases with fetal distress, followed by six (15.4%) patients with post-term pregnancy plus oligohydramnios, and six (15.4%) patients with post-term pregnancy alone. A significant difference was detected in oligohydramnios, preeclampsia, polyhydramnios, and gestational cholestasis between groups with or without fetal distress when they were assessed according to the risk factors prompting OCT (Table 2).

When fetal distress development was assessed according to the induction method used following a negative OCT, it was seen that fetal distress developed in 25.0% of 20 patients (n=5) who did not undergo labor induction and 9.1% of 427 patients (n=39) who underwent labor induction (Table 2). No significant difference was detected in fetal distress development according to the use of labor induction. When the induction methods used were assessed, it was seen that the dinoprostone group (n=237) had the highest rate of fetal distress development (10.5%). No significant difference was detected in fetal distress development according to the induction method used (*P*=0.257, *P*=0.906, *P*=0.077, *P*=0.792; Table 2).

Table 2: Fetal distress development following a negative OCT according to risk factors and induction methods in the induction group

Risk factor prompting OCT	Fetal distress (+) n=39 (9.1%)	Fetal distress (-) n=388 (90.9%)	Total n=427	*P*-value
Oligohydramnios	10(5%)	166 (95%)	176	0.038 $^{X^2}$
Post-term and oligohydramnios	6 (11%)	48 (89%)	54	0.589 $^{X^2}$
Post-term	6 (11%)	48 (89%)	54	0.589 $^{X^2}$
IUGR and Oligohydramnios	2 (5%)	36 (95%)	38	0.386 $^{X^2}$
Decreased fetal movements	3 (8%)	33 (92%)	36	0.862 $^{X^2}$
IUGR	4 (12%)	27 (88%)	31	0.449 $^{X^2}$
Non-reactive NST	3 (13%)	19 (87%)	22	0.452 $^{X^2}$
Preeclampsia	3 (33%)	6 (67%)	9	0.011 $^{X^2}$
Gestational Diabetes Mellitus	0 (0)	3 (100%)	3	0.582 $^{X^2}$
Polyhydramnios	1 (50%)	1 (50%)	2	0.044 $^{X^2}$
Gestational cholestasis	1 (50%)	1 (50%)	2	0.044 $^{X^2}$
Labor induction method				
Dinoprostone	25 (10.5)	212 (89.5)	237	0.257 $^{X^2}$
Oxytocin	11 (9.4)	106 (90.6)	117	0.906 $^{X^2}$
Cook Balloon + Oxytocin	1 (2.1)	46 (97.9)	47	0.077 $^{X^2}$
Foley Balloon + Oxytocin	2 (7.7)	24 (92.3)	26	0.792 $^{X^2}$

OCT: oxytocin challenge test, $^{X^2}$ Chi-square test

Perinatal outcomes

When NICU admission and its indications were assessed, it was seen that 397 neonates (93%) did not require NICU admission. Of the neonates who required NICU admission, it was found that 53.3% (n=16) were admitted due to neonatal respiratory distress, 33.3% (n=10) due to neonatal jaundice, 10% due to prematurity, and one (3.3%) due to brachial plexus injury.

Perinatal outcomes were assessed by stratifying according to gestational age: ≥41 weeks and between ≥37 and <41 weeks. It was found that of the 311 newborns with gestational age between ≥37 and <41 weeks, 46.9% were boys, and 53.1% were girls. It was found that the mean birth weight was 3011 (383) g, and the mean birth length was 49.2 (2.4) cm. The median Apgar scores were 9 and 10 at minutes 1 and 5, respectively. Again, it was found that of the 108 newborns with gestational age ≥41 weeks, 45.4% were boys, and 54.6% were girls. The mean birth weight was 3394 (349) g, and the mean birth length was 51.2 (1.5) cm. The median Apgar scores were 9 and 10 at minutes 1 and 5, respectively. It was seen that 31 newborns required NICU admission, including six newborns with gestational age <37 weeks. Prematurity was the underlying cause of NICU admission in 50% of six newborns with gestational age <37 weeks. In the remaining 25 newborns with gestational age ≥37 weeks, respiratory distress was the underlying cause of NICU admission in 45.1% (Table 3).

When the relationship between the time from OCT to delivery and fetal distress was assessed, fetal distress development was detected after OCT in 39 patients. Of these 39 patients, fetal distress developed within the first 24 hours in 24 (10.4%) and between hours 24 and 72 in 15 (8.2%). No fetal distress development was detected beyond 72 hours after the OCT (*P*=0.388). When Apgar scores were assessed in infants born following a negative OCT, it was found that Apgar scores were <7 at minute 1 in 10 newborns and four newborns at minute 5. When the relationship between the time from the OCT to delivery and Apgar scores was assessed, it was seen that there was no significant difference between Apgar scores <7 and >7 (*P*=0.416) (Table 4).

Table 3: Perinatal outcomes according to gestational age in the induction group

		Gestational age ≥37 – <41 weeks n=311	Gestational age ≥41 weeks n=108	*P*-value
		Mean (SD)/n (%)	Mean (SD)/n (%)	
Birth weight (g)		3011 (383)	3394 (349)	<0.001 [t]
Birth length (cm)		49.2 (2.4)	51.2 (1.5)	<0.001 [t]
Sex	Boy	146 (46.9)	49 (45.4)	0.777 [X²]
	Girl	165 (53.1)	59 (54.6)	
Apgar score on minute 1		8.9 (0.6)	8.8 (0.6)	0.594 [t]
Apgar score on minute 5		9.9 (0.5)	9.8 (0.5)	0.480 [t]
Perinatal complications				
None		310 (99.7)	107 (99.1)	0.450 [X²]
Shoulder dystocia		1 (0.3)	1 (0.9)	
Length of hospital stay (days)		3.0 (1.7)	3.8 (2.5)	0.449 [t]
NICU admission				
None		292(93.9)	102 (94.4)	0.391 [X²]
Yes		19 (6.1)	6 (5.6)	
Cause of NICU admission				
Neonatal jaundice		7 (36.8)	3 (50.0)	
Neonatal respiratory distress		11 (57.9)	3 (50.0)	
Brachial plexus injury		1 (5.3)	0 (0.0)	
Prematurity		0 (0.0)	0 (0.0)	

[X²] Chi-square test, [t] Student T test

Table 4: Relationship between time from OCT to delivery, fetal distress and Apgar score at minute 5 in the induction group

Time from OCT to delivery	Fetal distress (-)	Fetal distress (+)	*P*-value	APGAR ≥7	APGAR <7	*P*-value
	n (%)	n (%)		n (%)	n (%)	
0–24 hours	208 (89.6)	24 (10.4)	0.388 [X²]	231 (99.6)	1 (0.4)	0.416 [X²]
24–72 hours	167 (91.8)	15 (8.2)		179 (98.4)	3 (1.6)	
>72 hours	13 (100)	0 (0.0)		13 (100)	0 (0.0)	

OCT: oxytocin challenge test, [X²] Chi-square test

Discussion

This descriptive study investigated perinatal outcomes in pregnant women with a negative OCT. It was found that labor was induced in 95% of pregnant women with a negative OCT, dinoprostone and oxytocin were the most commonly used methods of labor induction, and delivery was via vaginal delivery in 72.6% and via C/S in 27.4%. No significant difference was detected in fetal distress rates between patients who did and did not undergo labor induction. In addition, there was no significant difference in fetal distress rates across the induction methods used. Of the patients, 7% were admitted to the NICU, most commonly due to respiratory distress and jaundice. No significant relationship was detected between fetal distress development, time from the OCT to delivery, and Apgar scores at minute 5.

In our study, delivery was via C/S in 117 (27.4%) patients. When C/S indications were assessed, fetal distress was the most common indication (33.3%). In the study of Różańska-Waledziak et al. [8], 69 (33.8%) of 204 patients with negative OCT underwent C/S. The authors found that abnormal heart traces comprised 37.6% of C/S indications. The C/S rate in our study was in agreement with the study mentioned above.

Our study observed fetal distress findings during follow-up and labor in 39 (9.1%) of 427 pregnant women who were followed after a negative OCT. This rate indicated that OCT was not effective in predicting fetal distress during labor. In a study by Schifrin et al. [20], fetal distress findings during labor were observed in three (3%) of 101 patients with a negative OCT. The difference may be attributed to the difference in sample sizes. Again, Apgar scores at minute 1 were <7 in 10 of 100 newborns, three of which had Apgar scores <7 (3%) at minute 5. In our study, Apgar scores at minute 1 were <7 in 10 of 427 patients, four of whom had Apgar scores at min 5 of <7 (0.9%). In a study by Hayden et al. [21], fetal distress findings during labor were observed in five of 97 patients with a negative OCT. The authors provided no data regarding Apgar scores, diagnoses, and labor induction methods.

Our study used labor induction methods in 427 of 447 patients with a negative OCT, and fetal distress was observed in 9.1%. No significant difference was found in fetal distress development across the induction methods. In a study by Cammu et al. [22], labor was induced in 7,683 nulliparous pregnant women, and fetal distress findings were observed in 2.6% of patients. However, the study had some limitations, including the selection of low-risk pregnancies, excluding post-term and low-birth-weight pregnancies, and a lack of data regarding fetal well-being before induction. In a study by Vahratian et al. [23], labor was induced in 429 low-risk, nulliparous pregnant women, and fetal distress findings were observed in 5.1% of patients. The authors assigned patients into two groups: (1) labor was directly induced using oxytocin in 286 patients and (2) using oxytocin following cervical preparation with a Foley balloon catheter in 143 patients. The fetal distress rate was 7% among patients who underwent cervical preparation with a Foley balloon catheter and 4.2% in the oxytocin group [23]. Again, the study was conducted on low-risk term pregnant women, and no test was performed to assess fetal well-being before induction. In our study, the fetal distress rate was 7.7% in the Foley balloon catheter and oxytocin group and 9.4% in the oxytocin group. Again, this study, like other studies, was applied to term pregnant women with low risk, and no test was applied for fetal well-being before induction. However, our study was limited by the small number of patients in the Cook balloon catheter and Foley balloon catheter groups and patients who did not undergo labor induction.

During follow-up from OCT to delivery, no stillbirths were observed in 427 pregnant women with a negative OCT. In a study by Evertson et al. [24], seven cases of stillbirth (1%) were observed during a one-week follow-up among 680 pregnant women with a negative OCT. The authors reported that causes of stillbirth included cord injury in three cases, placental detachment in one case, multiple anomalies in one case, and undefined in two cases, suggesting that fetal deaths were not due to disruption of fetal well-being. In a study on 1,337 high-risk pregnant women, Nageotte et al. [25] reported only one fetal death within seven days following a negative CST. Again, in a study on 679 women who underwent CST due to post-term pregnancy, no fetal death was observed by Freeman et al. [9]. In another study, CST was used to assess fetal well-being in 337 pregnant women with a previous history of stillbirth, and no fetal deaths were observed [10].

Our study also has limitations. The fact that all groups underwent induction, including high-risk patients, limited the ability to determine whether disruption of fetal well-being was caused by the risk status of patients or the induction method used.

Our study has some strengths. In studies about outcomes after negative OCT, fetal mortality rates were investigated during one-week follow-up in general [9,10,24,25]. However, our study evaluated fetal distress and Apgar scores in addition to fetal mortality. Low-risk pregnancies were preferred in studies that investigated outcomes according to induction methods, and no data were provided regarding fetal well-being before induction [22,23].

Conclusions

When outcomes are considered in pregnant women with a negative OCT, there are no fetal deaths and a limited number of newborns with low Apgar scores. It could be suggested that OCT is an effective method to predict fetal well-being but may not provide data regarding fetal distress development during labor. The induction methods used after negative OCT and the time until delivery have no effect on the development of fetal distress. Further randomized studies are needed to draw definitive conclusions.

References

1. Boylan P. Intrapartum fetal monitoring. Bailliere's Clinical Obstetrics and Gynaecology. 1987;1(1):73-95.
2. Apgar V. A proposal for a new method of evaluation of the newborn. Classic Papers in Critical Care. 1952;32(449):97.
3. Aarnoudse JG, Huisjes HJ, Gordon H, Oeseburg B, Zijlstra WG. Fetal subcutaneous scalp PO2 and abnormal heart rate during labor. Am J Obstet Gynecol. 1985 Nov 1;153(5):565-6. doi: 10.1016/0002-9378(85)90476-4. PMID: 4061520.
4. Holtzman RB, Banzhaf WC, Silver RK, Hageman JR. Perinatal management of meconium staining of the amniotic fluid. Clinics in Perinatology. 1989;16(4):825-38.
5. Riley RJ, Johnson J. Collecting and analyzing cord blood gases. Clinical Obstetrics and Gynecology. 1993;36(1):13-23.
6. Sarnat HB, Sarnat MS. Neonatal encephalopathy following fetal distress: a clinical and electroencephalographic study. Archives of Neurology. 1976;33(10):696-705.
7. Vintzileos AM, Gaffney SE, Salinger LM, Kontopoulos VG, Campbell WA, Nochimson DJ. The relationships among the fetal biophysical profile, umbilical cord pH, and Apgar scores. American Journal of Obstetrics and Gynecology. 1987;157(3):627-31.
8. Różańska-Walędziak A, Czajkowski K, Walędziak M, Teliga-Czajkowska J. The Present Utility of the Oxytocin Challenge Test—A Single-Center Study. Journal of Clinical Medicine. 2020;9(1):131.
9. Freeman RK, Garite TJ, Modanlou H, Dorchester W, Rommal C, Devaney M. Postdate pregnancy: utilization of contraction stress testing for primary fetal surveillance. American Journal of Obstetrics and Gynecology. 1981;140(2):128-35.
10. Freeman RK, Dorchester W, Anderson G, Garite TJ. The significance of a previous stillbirth. American Journal of Obstetrics and Gynecology. 1985;151(1):7-13.
11. Druzin M, Karver M, Wagner W, Hutson J, Waltner A, Kogut E. Prospective evaluation of the contraction stress and nonstress tests in the management of post-term pregnancy. Surgery Gynecology & Obstetrics. 1992;174(6):507-12.
12. Gabbe SG, Mestman JH, Freeman RK, Anderson GV, Lowensohn RI. Management and outcome of class A diabetes mellitus. American Journal of Obstetrics and Gynecology. 1977;127(5):465-9.
13. Lagrew DC, Pircon RA, Towers CV, Dorchester W, Freeman RK. Antepartum fetal surveillance in patients with diabetes: when to start? American Journal of Obstetrics and Gynecology. 1993;168(6):1820-6.
14. Obstetricians ACo, Gynecologists. ACOG practice bulletin no. 107: induction of labor. Obstet Gynecol. 2009;114:386-97.
15. Alfirevic Z, Keeney E, Dowswell T, Welton NJ, Medley N, Dias S, et al. Which method is best for the induction of labor? A systematic review, network meta-analysis and cost-effectiveness analysis. Health Technology Assessment. 2016;20(65):1-583.
16. Gabbe S, Jennifer R, Simpson J. Obstetrics: normal and problem pregnancies. Philadelphia. Elsevier/Saunders; 2012.
17. Freeman RK. The use of the oxytocin challenge test for antepartum clinical evaluation of uteroplacental respiratory function. American Journal of Obstetrics and Gynecology. 1975;121(4):481-9.
18. Hadi H. Cervical ripening and labor induction: clinical guidelines. Clinical Obstetrics and Gynecology. 2000;43(3):524-36.
19. Adair CD. Nonpharmacologic approaches to cervical priming and labor induction. Clinical Obstetrics and Gynecology. 2000;43(3):447-54.
20. Schifrin BS, Lapidus M, Doctor G, Leviton A. Contraction stress test for antepartum fetal evaluation. Obstetrics and Gynecology. 1975;45(4):433-8.
21. Hayden BL, Simpson JL, Ewing DE, Otterson WN. Can the oxytocin challenge test serve as the primary method for managing high-risk pregnancies? Obstet Gynecol. 1975 Sep;46(3):251-4. PMID: 1161226.
22. Cammu H, Martens G, Ruyssinck G, Amy J-J. Outcome after elective labor induction in nulliparous women: a matched cohort study. American Journal of Obstetrics and Gynecology. 2002;186(2):240-4.
23. Vahratian A, Zhang J, Troendle JF, Sciscione AC, Hoffman MK. Labor progression and risk of cesarean delivery in electively induced nulliparas. Obstet Gynecol. 2005 Apr;105(4):698-704. doi: 10.1097/01.AOG.0000157436.68847.3b. PMID: 15802393.
24. Evertson LR, Gauthier RJ, Collea JV. Fetal demise following negative contraction stress tests. Obstet Gynecol. 1978 Jun;51(6):671-3. PMID: 662243.
25. Nageotte MP, Towers CV, Asrat T, Freeman RK, Dorchester W. The value of a negative antepartum test: contraction stress test and modified biophysical profile. Obstet Gynecol. 1994 Aug;84(2):231-4. PMID: 8041536.