J Surg Med. 2022;6(11):893-898.
DOI: 10.28982/josam.1120336 Research article

Journal of Surgery and Medicine
e-ISSN: 2602-2079

The evaluation of ultrasonographic hip measurement differences among physicians according to the Graf method in newborns

Tarık Altunkılıç [1], Bünyamin Arı [1], Mehmet Yetiş [2], Nihat Kılıçaslan [3], Feyza İnceoğlu [4]

[1] Malatya Turgut Ozal University Faculty of Medicine, Department of Orthopedics and Traumatology, Turkey
[2] Ahi Evran Unıversity Faculty of Medicine, Department of Orthopedics and Traumatology, Turkey
[3] Mehmet Akif İnan Training and Education Hospital University, Department of Radiology, Turkey
[4] Malatya Turgut Ozal University Faculty of Medicine, Department of Biostatistics, Turkey

ORCID ID of the author(s)
TA: 0000-0002-1640-4275
BA: 0000-0001-9720-1869
MY: 0000-0002-8193-4344
NK: 0000-0003-4031-5798
Fİ: 0000-0003-1453-0937

Corresponding Author
Tarık Altunkılıç
Malatya Turgut Ozal University Faculty of Medicine, Department of Orthopedics and Traumatology, Turkey
E-mail: tarik.altunkilic@ozal.edu.tr

Ethics Committee Approval
This study was carried out after obtaining the approval of the Clinical Research Ethics Committee of Malatya Turgut Özal University, with the protocol number 2021/56 dated August 20, 2021.
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
2022 November 7

Abstract

Background/Aim: Hip ultrasonography (USG) is the most important diagnostic method in developmental hip dysplasia in newborns. However, a disadvantage of the ultrasonography method is that there can be measurement differences among doctors measuring the same hip. We aimed to investigate the causes and solutions of this situation. We further strived to measure the hip ultrasonography performed by different physicians using the Graf method and comparing the obtained values.

Methods: Hip USGs of newborns admitted to Malatya Turgut Ozal University Faculty of Medicine Hospital between Jan. 8, 2020 and Jan. 5,.2021 were measured and classified using the Graf method. The study type is consistent with retrospective cohort studies. Newborns aged 0-22 weeks without any additional pathology were included in the study. A radiologist and two orthopedists measured and interpreted the images separately in accordance with the Graf method. The first hip measurements (R1) were made by the radiologist (R) with the USG device, and they were classified according to alpha and beta angles; two printouts were made. The first orthopedic specialist (OS1) and the second orthopedic specialist (OS2) made their measurements with printouts. Subsequently, the results from the physicians were compared.

Results: A statistically significant difference was found between R1-OS2 ($P < 0.001$) and OS1-OS2 ($P < 0.001$) in terms of the Graf classifications. No statistically significant difference was found between R1 and OS1 in terms of the Graf classification ($P = 0.562$). A statistically significant difference was found between R1-OS2 ($P < 0.001$) and OS1-OS2 ($P = 0.048$) angles (alpha and beta) measurements. While R1 and OS1 measurements were compatible with each other, OS2 measurements were found to be inconsistent.

Conclusion: We think that there may be differences in angle measurements and the Graf classification among physicians who perform hip ultrasonography in newborns, and the most important way to correct this is through regular participation of physicians in subject-specific trainings.

Keywords: Hip, Ultrasonography, Newborn

How to cite: Altunkılıç T, Arı B, Yetiş M, Kılıçaslan N, İnceoğlu F. The evaluation of ultrasonographic hip measurement differences among physicians according to the Graf method in newborns. J Surg Med. 2022;6(11):893-898.

Introduction

Developmental hip dysplasia (DHD) is a progressive disease that occurs as a result of disruption of the integrity of the acetabulum and femoral head due to many risk factors during pregnancy. In the past, the incidence of DHD was thought to be 1 in 1000 live births [1]. In the study of Kutlu et al. [2] the incidence of DHD was found to be 1.34%. The incidence of DDH may vary according to race and geographical regions. DHD can be seen at a rate of 1.5-2% in Europe and at a rate of 1-1.5% in Turkey [3, 4].

If DHD is not diagnosed and treated early, it can cause serious arthrosis and deformities in later life [5]. Early diagnosis is the basis of treatment and also increases the success of treatment. Although Barlow and Ortolani tests performed by physical examination have an important place in the diagnosis, hip ultrasonography (USG) performed in the first six months after birth is the most valuable method [6]. In children older than six months, pelvic roentgenogram has an important place in the diagnosis [7]. Combined clinical examination and USG are the most important methods in diagnosis and follow-up. USG is routinely used and accepted, because it is an internationally standardized diagnostic method, can diagnose up to the sixth month, and helps in the planning of follow-up and treatment [8, 9].

Although USG is used in diagnosis and treatment, we think that different results may be obtained according to the measurement of the radiologist or orthopedist performing the USG. It was stated in the studies of Bar-on and Zieger et al. [10, 11] that it may vary according to the person making the measurement.

Hip ultrasonography is the most important diagnostic method in developmental hip dysplasia in newborns. However, a disadvantage of the ultrasonography method is that there can be measurement differences among doctors measuring from the same hip. We aimed to investigate the causes and solutions of this situation. We further strived to measure the hip ultrasonography performed by different physicians using the Graf method and comparing the obtained values.

Materials and methods

Hip USGs of newborns admitted to our hospital between Jan. 8, 2020 and Jan. 5, 2021 were measured and classified using the Graf method. This study was carried out after obtaining the Clinical Research Ethics Committee of Malatya Turgut Özal University's approval, with the protocol number 2021/56 dated Aug. 20, 2021. Fifty five (49.5%) of the newborns were girls and 56 (50.5%) were boys. The study type is consistent with retrospective cohort studies. Newborns aged 0-22 weeks without any additional pathology were included in the study. Two printouts were obtained for each hip from each newborn. A radiologist and two orthopedists measured and interpreted the images separately in accordance with the Graf method. The first hip measurements (R1) were made by the radiologist (R) with the USG device, and they were classified according to alpha and beta angles, and two printouts were made. Five days later, participants were called, and R1 measured a second time (R2). The printouts were given to the first orthopedic specialist (OS1) and the second orthopedic specialist (OS2). The radiologist had previously had hip USG training. OS1 had a Graf method pediatric hip ultrasonography course certificate. OS2 had training only during the residency period.

They measured the hips manually with a protractor ruler in accordance with the Graf method and classified them according to alpha and beta angles. The alpha and beta angle measurements and classification according to the Graf method performed by the radiologist were compared with the alpha and beta angle measurements made by the two orthopedic specialists manually from the printout and classification according to the Graf method. In the study, the radiologist's first measurement (R1) was compared with the measurements of OS1 and OS2.

Type 3 and type 4 hips were not found in the study. In addition, hips were divided into two groups; Graf type 1 hips (Group 1), which did not need follow-up, and Graf type 2a - type 4 hips (Group 2), which required follow-up and treatment in order to calculate the agreement between the physicians.

Ultrasound scanning was performed with the Esaote Mylab Seven model 2019 linear probe manufactured in Genoa, Italy. The radiologist examined all hip joints with and without stress prior to USG (using Barlow and Ortolani). Then, USG was performed by the same radiologist to evaluate the stability of the hip structure and femoral head at rest and under stress (Figure 1). When the baby was in the lateral decubitus position, the probe was inserted using the lateral approach. Scanning was evaluated by fixing the hips in 20-30 degrees of flexion and 5-10 degrees of internal rotation in extension. The bone and cartilage components of the hip joint and the structure of the femoral head within the acetabulum were evaluated. Graf's classification was used for measurements. Alpha and beta angles were calculated on coronal images. In coronal images, an imaginary line is drawn from the baseline of the iliac bone and extended along the femoral head. The angle between the tangent drawn from the distal end of the deepest point of the acetabulum to the bony corner was calculated as alpha. Beta angle was calculated with the line drawn from the midpoint of the labrum to the acetabular corner (Figure 2). Two printouts were made for each hip.

Figure 1: USG measurement of radiologist

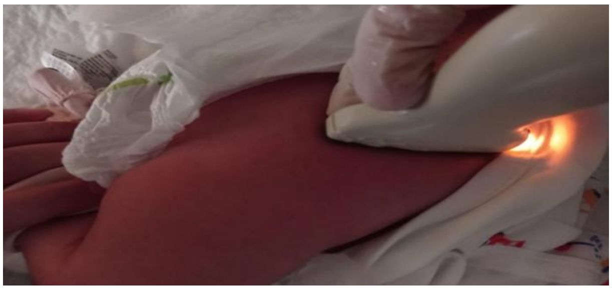
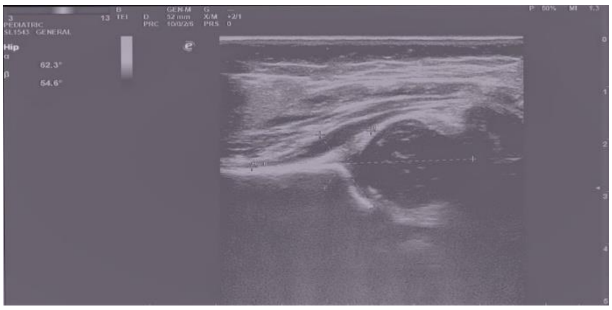


Figure 2: USG alpha and beta angle measurement



Statistical analysis

The analysis of the data included in the research was carried out using the Statistical Package for the Social Sciences (SPSS) 25 program and Kolmogorov-Smirnov test [12]. Two paired sample t-tests were used to evaluate the observers. The Pearson correlation coefficient was also calculated. Values that were frequently used in the evaluation of the findings were as follows: 0.00 – 0.19 no relationship (negligible low relationship), 0.20 – 0.39 weak relationship, 0.40 – 0.69 moderate relationship, 0.70 – 0.89 strong relationship, 0.90 – 1.00 a very strong relationship [12].

The sample of this study was determined by power analysis, using the G*Power 3.1 program. According to the power analysis, the sample was determined as 198 with an effect size of 0.25, a margin of error of 0.05, and a confidence level of 0.95.

Interobserver reliability is used to show variability between two or more raters measuring the same group of participants [13]. Observers were evaluated using the intraclass correlation coefficient (ICC). Since the participants were evaluated by the same observer, the ICC (3,2) model and two-way mixed effects model were used. ICC (3,2), standard error of measurement (SEM) and inter-measurement reliability, which is the smallest detectable difference (SDD), were evaluated [14-16]. SDD was calculated to generate random error scores. The low value of the calculated SEM indicates that the participant was consistent throughout the test trials [15, 17]. In order to compare the measurements among observers, comparisons between groups were made using the ANOVA test. The measurement values of the tests were recorded by providing similar test setups and conditions for all participants included in the study.

Results

One hundred eleven newborns were included in the study. A total of 55 (49.5%) were female and 56 (50.5%) were male. The mean age of the participants was 69.72 (30.78) days, with the youngest age being 22 days, and the highest, 155 days. A total of 222 hips of the participants, both right and left hips, were evaluated.

Mean, standard deviation, minimum and maximum values of both R1 and R2 alpha (α) and beta (β) angles taken by R, alpha (α), beta (β) angles measured by OS1, alpha (α), beta (β) angles measured by OS2 were calculated and results are given in Table 1.

The number and percentage values of R1 and R2 Graf classification (Type 1, 2A, 2B, 2C), OS1 Graf classification (Type 1, 2A, 2B, 2C) and OS2 Graf classification (Type 1, 2A, 2B, 2C) distributions of participants in the study were calculated and the results are given in Table 2. No Type 3 or Type 4 patients were found in the study.

To calculate the agreement among the physicians, measurements were divided into two groups, Group 1 (type 1) and Group 2 (type 2a to type 4). Group 1 and Group 2 classification numbers and percentage values calculated by R1, OS1 and OS2 are given in Table 3.

There was no statistically significant difference between R1 and R2 measurements made by R for alpha (α) angle variable (P^1 = 0.708, Table 4). It was observed that the values obtained with the second measurement were higher than the values obtained with the first measurement (difference = -0.05). A very high level (r = 0.872) of statistically significant correlation was found between R1 and R2 (P^1 = 0.708, Table 4).

Table 1: Alpha and beta angle measurement averages of physicians

Angle	Mean (SD)	Min	Max
R1 α angle	61.92 (3.84)	48	73
R1 β angle	55.4 (7.66)	35	72
R2 α angle	61.97 (3.95)	46	70
R2 β angle	56.36 (6.78)	40	71
OS1 α angle	60.02 (12.1)	62	71
OS1 β angle	57.9 (12.32)	55	72
OS2 α angle	59.22 (5.31)	43	75
OS2 β angle	60.64 (5.82)	42	80

SD: standard deviation, Min: lowest score received, Max: highest score

Table 2: Number of hips by Graf classification

	R1 n(%)	R2 n(%)	OS1 n(%)	OS2 n(%)
Type 1	171 (77%)	173 (77.9%)	146 (73%)	120 (54.1%)
2A	41 (18.5%)	39 (17.6%)	48 (21.6%)	76 (34.2%)
2B	9 (4.1%)	9 (4.1%)	11 (5%)	21 (9.5%)
2C	1 (0.5%)	1 (0.5%)	1 (0.5%)	5 (2.3%)

n: frequency, %: percent

Table 3: Group 1 and Group 2 distribution of physicians' measurements

	R1	OS1	OS2
Group 1	171 (77%)	162 (73%)	120 (54,1)
Group 2	51 (23%)	60 (%27%)	102 (45.9%)

%: percent, Group 1: Graf type 1 hips that do not need follow-up, Group 2: Hips between Graf type 2a + type 4 requiring follow-up and treatment

There was no statistically significant difference between R1 and R2 performed by the radiologist for the beta (β) angle variable (P^1 = 0.220, Table 4). It was observed that the values obtained with the second measurement were higher than the values obtained with the first measurement (difference = -0.95). A very high level (r = 0.952) statistically significant correlation was found between the first and second measurements (P^1 = 0.220, Table 4).

No statistically significant difference was found between R1 and R2 measurements in the Graf classification using R (*P* = 0.058, Table 5).

Table 4: R1 and R2 angle measurements of the radiologist

Angle	Measurement	Mean (SD)	P^1-value	r	P^2-value
Alfa (α)	R1	61.92 (3.84)	0.708	0.872	0.001*
	R2	61.97 (3.95)			
Beta (β)	R1	55.40 (7.66)	0.220	0.952	0.001*
	R2	56.36 (6.78)			

SD: standard deviation, P^1: statistical significance of two paired samples t test, r: between two observation correlation coefficients, P^2: **P* <0.05. There is a statistically significant relationship between the two tests.

Table 5: Comparison of two measurements of the radiologist according to Graf classification

		R2				Total	*P*-value
		Type 1	2A	2B	2C		
R1	Type 1	168 (97.1%)	3	0	0	171 (77%)	0.058
	2A	5 (2.9%)	36 (92.3%)	0	0	41 (18.5%)	
	2B	0	0	9 (%100)	0	9 (4.1%)	
	2C	0	0	0	1 (100%)	1 (0.5%)	
Total		173	39	9	1	222	

%: percent, *P*: test statistical significance's value of kappa statistics

It was found that there was a statistically significant difference among the physicians in the alpha and beta measurements of the participants included in the study. To measure from which group the difference originates, pairwise comparisons were made with the Tukey test. According to the paired comparisons:

- No statistically significant difference was found between R1 and OS1 angle measurements (P^2 = 0.051, Table 6).
- A statistically significant difference was found between R1 and OS2 angle measurements (alpha angle: P^2 < 0.001, Table 6).
- A statistically significant difference was found between OS1 and OS2 angle measurements (P^2 = 0.048, Table 6).

Table 6: Comparison of alpha and beta angle measurements among physicians

Angle	Measurement	Mean (SD)	P^1-value	Multiple comparisons		
				First	Second	P^2-value
Alfa	R1	61.92 (3.84)	<0.003*	R1	OS1	0.051
	OS1	60.02 (4.27)			OS2	0.001**
	OS2	59.22 (5.31)		OS1	OS2	0.048**
Beta	R1	55.4 (7.66)	<0.001*	R1	OS1	0.053
	OS1	57.9 (5.96)			OS2	0.001**
	OS2	60.64 (5.85)		OS1	OS2	0.001**

SD: standard deviation, *P*: statistical significance of repeated measures of ANOVA, * $P^1 < 0.05$: There is a statistically significant difference between observers ** $P^2 < 0.05$: There is a statistically significant difference between observers.

In the measurements made for the alpha angle, the R1 and OS1 ICC value was calculated as 0.892 (95% CI; 0.859-0.917; Table 7). The reliability between measurements was evaluated with the Standard Error of Measurement (SEM) and the smallest detectable difference (SDD). High ICC value, low SEM value indicates good inter-participant agreement [17, 18].

The R1 and OS2 ICC value was calculated as 0.327 (95% CI; 0.205-0.440). The reliability between measurements was evaluated with the Standard Error of Measurement (SEM) and the smallest detectable difference (SDD). A low ICC value and a high SEM value indicate that the agreement among participants is not good [17, 18]. OS1 and OS2 ICC values were calculated as 0.347 (95% CI; -0.226-0.458; Table 7). In the measurements made for the beta angle, the R1 and OS1 ICC value was calculated as 0.734 (95% CI; 0.654-0.796). R1 and OS2 ICC value was calculated as 0.284 (95% CI; 0.159-0.401). OS1 and OS2 ICC values were calculated as 0.389 (95% CI; 0.271-0.495; Table 7).

Table 7: ICC, SEM, and SD values of physicians' angle measurements

Angle	Measurements	Measurements	ICC (3.2)	% 95 CI		SEM	SD
				Lower Bound	Upper Bound		
Alfa	R1	OS1	0.892	0.859	0.917	0.101	0.281
		OS2	0.327	0.205	0.440	1.361	3.772
	OS1	OS2	0.347	0.226	0.458	1.257	3.484
Beta	R1	OS1	0.734	0.654	0.796	0.521	1.443
		OS2	0.284	0.159	0.401	2.767	7.670
	OS1	OS2	0.389	0.271	0.495	2.514	6.969

ICC: Intraclass Correlation Coefficient, CI: Confidence Interval, SEM: Standard Error of Measurement, SD: Smallest detectable difference.

No statistically significant difference was found between R1 and OS1 in terms of the Graf classification ($P = 0.562$, Table 8). A statistically significant difference was found between R1 and OS2 in terms of the Graf classification ($P = 0.001$, Table 8). A statistically significant difference was found between OS1 and OS2 in terms of the Graf classification ($P < 0.001$, Table 9). No statistically significant difference was found between Group 1 and Group 2 measurements of R1 and OS1 ($P = 0.058$, Table 10). A statistically significant difference was found between R1 and OS2 measurements ($P = 0.001$, Table 10). A statistically significant difference was found between Group 1 and Group 2 measurements of OS1 and OS2 ($P < 0.001$, Table 11).

Table 8: Inter-physician comparison of Graf type classification

		R1				Total	*P*-value
		Type 1	2A	2B	2C		
OS1	Type 1	160	2	0	0	162	0.562
	2A	9	39	0	0	48	
	2B	2	0	9	0	11	
	2C	0	0	0	1	1	
OS2	Type 1	108	10	1	1	120	0.001*
	2A	48	28	0	0	76	
	2B	13	0	8	0	21	
	2C	2	3	0	0	5	

* $P < 0.05$: There is a difference between groups. *P*: test statistical significance's value of kappa statistics.

Table 9: Inter-orthopedist comparison of Graf type classification

		OS1				Total	*P*-value
		Type 1	2A	2B	2C		
OS2	Type 1	105	11	3	1	120	<0.001*
	2A	42	34	0	0	76	
	2B	13	0	8	0	21	
	2C	2	3	0	0	5	

* $P < 0.05$: There is a difference between groups. *P*: test statistical significance's value of kappa statistics.

Table 10: Comparison of R1 between OS1 and OS2 in Group 1 - Group 2 classifications

		R1		Total	*P*-value
		Group 1	Group 2		
OS1	Group 1	159	3	162	0.058
	Group 2	12	48	60	
OS2	Group 1	108	12	120	0.001*
	Group 2	63	39	102	
Total		171	51	222	

* $P < 0.05$: There is a difference between groups. *P*: test statistical significance's value of kappa statistics, Group 1: Graf type 1 hips that do not need follow-up, Group 2: Hips between Graf type 2a + type 4 requiring follow-up and treatment

Table 11: Comparison between OS1 and OS2 in Group 1 - Group 2 classifications

			OS1		Total	*P*-value
			Group 1	Group 2		
OS2	Group 1	n	105	15	120	<0.001*
	Group 2	n	57	45	102	

n: frequency, %: percent, * $P < 0.05$: There is a difference between groups. *P*: test statistical significance's value of kappa statistics, Group 1: Graf type 1 hips that do not need follow-up, Group 2: Hips between Graf type 2a + type 4 requiring follow-up and treatment

Discussion

DHD is a progressive disease that occurs in newborns due to many prenatal and postnatal causes. The main deterrent of this disease is early diagnosis and treatment [19]. If the correct diagnosis and treatment is not made in the newborn, it causes permanent deformities, loss of ability to work, and a decrease in the quality of life of the patient [20]. It should be noted that 9.1% of cementless total hip prostheses and 5.2% of cemented total hip prostheses are performed due to hip dysplasia or dislocation [21]. Barlow and Ortolani examinations are also frequently used in the diagnosis of newborns. However, the specificity of these methods was found to be high >99% and the sensitivity to be 60% [22, 23]. Pelvic roentgenogram is not used for early diagnosis in children under six months, since the femoral head ossifies after four to six months [24]. For all these reasons, the importance and value of USG has been increasing.

Ultrasonography will not only prevent the occurrence of such problems in adults, but will also contribute to the reduction in surgical treatment costs of DHD [25, 26]. Although USG reduces the problems and treatment costs of the family and the child in the future, it may cause false positives and unnecessary treatments [27]. In the study of Dias et al. [28], two pediatric radiologists and three orthopedic physicians reproduced the hip ultrasound images of 209 newborns and evaluated them with the Graf method. It was stated that USG is an important diagnostic method in DHD, although the reliability among physicians is weak.

The Graf ultrasonographic hip measurement method has an important place in determining hip dislocation, as it allows us to make USG and angle measurements. In order to reduce the errors of physicians who perform USG measurements and increase the validity, reference points, drawing lines, and accurate measurement of angles were determined [29]. It has been proven in publications that errors can be minimized as a result of following these guidelines [30].

In the study of Bar-On et al. [31], hip ultrasonography was performed on 75 infants by two different physicians consecutively. Ultrasound images were printed and analyzed twice by three pediatric orthopedic surgeons and classified using the Graf method. Intra- and inter-observer agreement between interpretations was classified as normal (types 1 to 2A) and abnormal (types 2B to 4) using the Graf classification and analyzed using kappa coefficients. When examining the same printout, the mean of intra-observer agreement kappa value for Graf classification was 0.6, but the inter-physician agreement kappa value was found to be moderate, 0.50. For normal and abnormal grouping, the mean of intra-observer agreement kappa value was 0.67 and inter-physician agreement kappa value was 0.57. According to the findings of this study, Bar-On et al. indicated that both the ultrasonography technique and the physician interpreting the image may affect the result, and that there was a weak consensus among orthopedists [31].

In the study of Roover et al. [32], 200 ultrasonography images were classified according to Graf's method by four radiographers and a radiologist. They stated that the interobserver agreement regarding the ultrasound evaluation of the hip was good enough and that observer variability did not cause any serious cases to be overlooked. In the study of Ömeroğlu et al. [33] the rates of intra-observer and inter-physician agreement in the Graf classification were 65% and 51%, respectively. Intra-observer and inter-physician agreement rates in treatment method by hip type were 76% and 64%, respectively. In the study, it was stated that having basic knowledge about the Graf method was the most important point and the number of previous examinations by the observer had no effect on the results. In a study by Roposch et al. [34] intra- and inter-observer variability in the interpretation of ultrasound was examined. Training materials developed by Graf were given to four orthopedic specialists in their last year of training. However, two physicians (Group A) attended the training courses on the technique, while the other two (Group B) did not attend the training courses. A misclassification affecting treatment was rare; one patient received unnecessary treatment and three patients did not receive the necessary treatment. In the study, it was concluded that how to do and interpret ultrasound should be done correctly through training courses and self-study is insufficient [34]. In our study, no statistically significant difference was found between R1 and OS1 in terms of Graf type and normal-abnormal classification, but a significant difference was found when these physicians were compared with OS2.

Ozgun et al. [35] treated 210 babies. Ultrasound images were evaluated by two pediatric orthopedic professors, two orthopedic specialists, and two orthopedic assistants. In their study, they stated that more importance should be given to the beta angle and cartilage labrum in the resident training program and this is directly proportional to clinical experience. In our study, a statistically significant difference was found between physicians in alpha and beta measurements. It was seen that this difference was caused by the measurements of OS2 rather than the measurements of R1 and OS1.

Simon et al. [27] evaluated 158 ultrasonographic images. There was better agreement among physicians in the assessment of immature hips compared to mature hips. The least agreement was between the least experienced and most experienced physicians. It was observed that the agreement between physicians was higher in regard to immature hips than in mature hips. Therefore, they stated that USG is important in the evaluation of immature hips. The study indicated that although there were differences between the measurements in alpha and beta angles measured among physicians, the agreement between physicians was good in the classification, and the experience and education of the researchers played an important role in the compliance [27]. Melzer et al. [36] in their study, determined a mean error of 3.2° for the alpha angle and 11.9° for the beta angle. According to Neither et al [37], it is assumed that there may be a 10° margin of error in the measurements of alpha and beta angles. Roovers et al. [32] mean standard deviations of 3.2° and 6° were reported for alpha and beta angles, respectively. According to the Graf method, the mean standard deviation for alpha and beta angles is 4° [38]. In the study of Pedrotti et al. [39] ultrasonography of 798 patients was examined by different physicians. In the study, alpha and beta angle changes were not found to be statistically significant. In our study, it was seen that the ICC values obtained as a result of alpha and beta measurement showed good agreement between R1 and OS1, but the agreement between the two physicians with OS2 was not good.

The study of Rosendhal et al. [40] aimed to determine the agreement between physicians in the evaluation of hip morphology and stability with ultrasound. Three groups of infants were examined. Intra-observer and inter-physician agreements were determined for reading and reviewing the recorded ultrasound scans. A high degree of agreement was found for the morphological classification based on repeated examinations by two physicians of the scans of 206 infants enrolled by the same physician (doctors' kappa values of 0.7 and 0.8, respectively). The degree of agreement between the physicians was found to be moderate (kappa value = 0.5). There was moderate agreement among physicians in determining hip stability in 70 infants (kappa value = 0.4). In their study, they stated that in order to achieve a high level of agreement among physicians in the evaluation of hip morphology and stability in newborns, appropriate training and attention to detail in the measurement technique should be taken [40].

We examined the variation and variables between physicians' own measurements and among physicians in the interpretation of infant hip ultrasound. There was no pediatric radiologist in our study, but we had a radiologist and two orthopedic specialists. OS1 had the Graf method pediatric hip ultrasonography course certificate. OS2 only trained for developmental dysplasia during his residency. Among the three physicians, R and OS1 were the most compatible with each other, and OS1 and OS2 were the most incompatible. We attribute this situation to the fact that OS2 did not receive hip USG training.

Limitations

The low number of patients in this study can be considered a limitation. In addition, the inclusion of more doctors in this study would have been beneficial.

Conclusion

While there was agreement between trained doctors in the Graf measurements made by more than one doctor, it was seen that there was no agreement between the doctor who was not trained and the other doctors. We think that the most important way to increase this compliance is that doctors should regularly attend subject-specific training courses.

References

1. Ellen de OG, Miguel A, Juliana Pp, Claudio S. The epidemıology of developmental dysplasıa of the hip in males. Acta Ortop Bras. 2020;28:26–30.
2. Kutlu A, Memik R, Mutlu M. Congenital dislocation of the hip and its relation to swadding in Turkey. J Pediatr Orthop. 1992;12:80-2.
3. Falliner A, Schwinzer D, Hahne H-J, Hedderich J, Hassenpflug J. Comparing Ultrasound Measurements of Neonatal Hips Using the Methods of Graf and Terjesen. JBJS. 2006;88:104–6.
4. Sadık B, Bartu S. Gelişimsel Kalça Displazisi Gelişimsel Kalça Displazisi. Güncel Pediatri. 2005;3:18-21.
5. Keller MS, Nijs EL. The role of radiographs and US in developmental dysplasia of the hip: how good are they? Pediatr Radiol. 2009;39:211–5.
6. Graf R. The diagnosis of congenital hip-joint dislocation by the ultrasonic Combound treatment. Arch Orthop Trauma Surg. 1980;97:117–33.
7. Grissom L, Harcke HT, Thacker M. Imaging in the surgical management of developmental dislocation of the hip. Clin Orthop Relat Res. 2008;466:791–01.
8. Graf R. Fundamentals of sonographic diagnosis of infant hip dysplasia. J Pediatr Orthop 1984;4:735–40.
9. Graf R. New possibilities for the diagnosis of congenital hip joint dislocation by ultrasonography. J Pediatr Orthop. 1983;3:354–9.
10. Bar-On E, Meyer S, Harari G, Porat S. Ultrasonography of the hip in developmental hip dysplasia. J Bone Joint Surg Br. 1998;80:319–22.
11. Zieger M. Ultrasound of the infant hip. Part 2. Validity of the method. Pediatr Radiol. 1986; 16:488–92.
12. Alpar, R, Spor, Sağlık ve Eğitim Bilimlerinde Örneklerle Uygulamalı İstatistik ve Geçerlik-Güvenirlik. Ankara: Detay; 2020. Pp. 147.
13. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 1979;86:420–8.
14. Rankin G, Stokes M . Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. Clinical Rehabilitation. 1988;12:187-99.
15. Thomas JR, Nelson JK, Silverman J. Research methods in physical activity. Champaign: Human Kinetics; 2005. pp. 66-7.
16. Kropmans, TJB, Dijkstra PU, Stegenga B, Stewart R, De Bont, LGM. Smallest detectable difference in outcome variables related to painful restriction of the temporomandibular joint. J Dent Res.1999;78:784-9.
17. Coppieters M, Stappaerts K, Janssens K, Jull G. Reliability of detecting 'onset of pain' and 'submaximal pain' during neural provocation testing of the upper quadrant. Physiother Res Int. 2002;7:146-56.
18. Joseph P Weir. Quantıfyıng Test-Retest Relıabılıty Usıng The Intraclass Correlatıon Coeffıcıent And The Sem. J Strength Cond Res. 2005;19:231–40.
19. Köse N, Ömeroğlu H, Dağlar B. Gelişimsel Kalça Displazisi Ulusal Erken Tanı ve Tedavi Programı. 2010;2-19.
20. Hoaglund FT, Steinbach LS. Primary osteoarthritis of the hip: etiology and epidemiology. J Am Acad Orthop Surg. 2001;9:320–7.
21. Furnes O, Lie SA, Espehaug B, Vollset SE, Engesaeter LB, Havelin LI. Hip disease and the prognosis of total hip replacements. A review of 53,698 primary total hip replacements reported to the Norwegian Arthroplasty Register 1987–99. J Bone Joint Surg Br. 2001;83:579–86.
22. Macnicol MF. Results of a 25-year screening programme for neonatal hip instability. J Bone Joint Surg Br. 1990;72:1057–60.
23. Jones D. An assessment of the value of examination of the hip in the newborn. J Bone Joint Surg Br. 1977;59:318–22.
24. Grissom L, Harcke HT, Thacker M. Imaging in the surgical management of developmental dislocation of the hip. Clin Orthop Relat Res. 2008;466:791–801.
25. Thallinger C, Pospischill R, Ganger R, Radler C, Krall C, Grill F. Long-term results of a nationwide general ultrasound screening system for developmental disorders of the hip: the austrian hip screening program. J Child Orthop. 2014;8:3–10.
26. Riccabona M, Schweintzger G, Grill F, Graf R, Graf R. Screening for developmental hip dysplasia (DDH)-clinically or sonographically? Comments to the current discussion and proposals. Pediatr Radiol. 2013;43:637–40.
27. Simon EA, Saur F, Buerge M, Glaab R, Roos M, Kohler G. Inter-observer agreement of ultrasonographic measurement of alpha and beta angles and the final type classification based on the graf method. Swiss Med Wkly. 2004;134:671–7.
28. Dias JJ, Thomas IH, Lamont AC, Mody BS, Thompson JR. The reliability of ultrasonographic assessment of neonatal hips. J Bone Joint Surg Br. 1993;75:479–82.
29. Graf R. Hip sonography: background; technique and common mistakes; results; debate and politics; challenges. Hip Int. 2017;27:215–9.
30. Graf R, Mohajer M, Plattner F. Hip sonography update. Quality-management, catastrophes - tips and tricks. Med Ultrason. 2013;15:299–3.
31. Bar-On E, Meyer S, Harari G, Porat S. Ultrasonography of the hip in developmental hip dysplasia. J Bone Joint Surg Br. 1998;80:323-5.
32. Roovers EA, Boere-Boonekamp M, Geertsma T, Zielhuis G, Kerkhof A. Ultrasonic screening for developmental dysplasia of the hip in infants. Reproducibility of assessments made by radiographers. J Bone Joint Surg. 2003;85:726–30.
33. Omeroğlu H, Biçimoğlu A, Koparal S, Seber S. Assessment of variations in the measurement of hip ultrasonography by the graf method in developmental dysplasia of the hip. J Pediatr Orthop B. 2001;10:89–95.
34. Roposch A, Graf R, Wright JG. Determining the reliability of the graf classification for hip dysplasia. Clin Orthop Relat Res. 2006;447:119–24.
35. Ozgun K, Ozgur K, Ahmet SS, Mehmet MO, Hasan HM. Is it difficult to obtain inter-observer agreement in the measurement of the beta angle in ultrasound evaluation of the paediatric hip. J Orthop Surg Res. 2019; 17;14:221.
36. Melzer Ch. Röntgenbild-Sonographie-Anatomie, Orthopädische Klinik Giessen, Angeborene Hüftdysplasie und -luxation vom Neugeborenen bis zum Erwachsenen. 1993 Nov; Symposium Zürich Universität-Irchel. 1993.p. 69-77.
37. Niethard FU, Roesler H. Die Genauigkeit von Längen- und Winkelmessungen im Röntgenbild und Sonogramm des kindlichen Hüftgelenkes. Z Orthop. 1987;125:170–6.
38. Graf R. Die anatomischen Strukturen der Sauglingshüfte und ihre sonographische Darstellung. Morphol Med.1982;2:29–38.
39. Luisella P, Ilaria C, Alessandro D, Federica D, Francesca R, Mario M. Interpreting Neonatal Hip Sonography: Intraobserver and Interobserver Variability. J Pediatr Orthop B. 2020;29:214–8.
40. Rosendahl K, Aslaksen A, Lie RT, Markestad T. Reliability of ultrasound in the early diagnosis of developmental dysplasia of the hip. Pediatr Radiol. 1995;25:219-24.

The National Library of Medicine (NLM) citation style guide has been used in this paper.