JOURNAL OF HEALTH SCIENCES AND MEDICINE

Sağlık Bilimleri ve Tıp Dergisi

J Health Sci Med 2020; 3(2): 121-124

Research Article / Araştırma Makalesi

Interobserver reliability in the ultrasonography evaluation with Graf method of developmental dysplasia of the hip: the importance of education for ultrasonography classification

Gelişimsel kalça displazisinin Graf yöntemiyle ultrasonografi değerlendirmesinde gözlemciler arası güvenirlilik: ultrasonografi sınıflamasında eğitimin önemi

Kadri Yıldız[1], Hayrunnisa Bekis Bozkurt[2], Türkhun Çetin[3], Vahit Yıldız[4]

[1]Kafkas University, School of Medicine, Department of Orthopedics and Traumatology, Kars, Turkey
[2] Kafkas University, School of Medicine, Department of Pediatric Health and Diseases, Kars, Turkey
[3]Kafkas University, School of Medicine, Department of Radiology, Kars, Turkey
[4]Adnan Menderes University, School of Medicine, Department of Orthopedics and Traumatology, Aydın, Turkey

ABSTRACT

Aim: Developmental dysplasia of the hip is a hip deformity that can be diagnosed early by numerous ultrasonography measurements. The purpose of this study was to evaluate the interobserver reliability of ultrasonography measurements using the Graf ultrasonography (USG) method.

Material and Method: Ultrasonography measurements that were used by Graf ultrasonography (USG) method were obtained at presentation for 62 randomized and consecutive patients. Bilateral hip USG measurements were made for each patient. Each USG outcome was evaluated on multiple occasions by an orthopedic surgery specialist, a pediatric specialist, and a 25-year experienced radiologist.

The statistical measurements were made by SPSS 20.0 (Windows, IL, USA). Intraclass Correlation Coefficient was determined. 95% Confidence Interval and F Test with True Value 0 was detected. P values were accepted as statistically significantly less than 0,05.

Results: At the evaluation before education for Graf USG classification, the results were established. In the interobserver reliabilities, Intraclass classification correlations (ICC) values were 0,939 for USG classification evaluation with Graf method for right hips. ICC values were 0,907 for USG classification evaluation with Graf method for left hips. At the evaluation after education for Graf USG classification, the results were established. ICC values were 0,975 for USG classification evaluation with Graf Method for right hips. ICC values were 0,970 for USG classification evaluation with Graf Method for left hips. All p values were significant. While before education for Graf USG classification ICC was 0,838 in right hips, 0,765 in left hips; after education, it was 0,928 in right hips, 0,915 in left hips.

Conclusion: Diagnosis of DDH deformity with USG is a complex and difficult condition that needs a serious education period. Graf USG method was found to have high interobserver reliability. And also, we detected that education for Graf USG classification increased intra-observer reliability, especially among pediatrists.

Keywords: Developmental dysplasia of the hip, ultrasonography, Graf, inter-observer reliability, intra-observer reliability

ÖZ

Amaç: Birkaç ultrasonografi (USG) ölçümüyle gelişimsel kalça displazisi (GKD) deformitesi erken dönemde teşhis edilebilmektedir. Bu çalışmanın amacı, Graf USG yöntemini kullanarak gözlemciler arası güvenirliliği değerlendirmekti.

Gereç ve Yöntem: Altmış iki rastgele ve ardışık olarak sağlanan hasta için Graf USG yöntemi ölçümlerde kullanıldı. Her hasta için her iki kalça US ölçümü yapıldı. Her US sonucu, multidisipliner yaklaşımla, ortopedik cerrahi uzmanı, çocuk hastalıkları uzmanı ve 25 yıl tecrübeli radyoloji uzmanıyla değerlendirildi. İstatistiksel ölçümler SPSS 20.0 (Windows, IL, ABD) ile yapıldı. Sınıf İçi Korelasyon Katsayısı belirlendi. % 95 Güven Aralığı ve Gerçek Değer 0 ile F Testi tespit edildi. 0,05'ten daha düşük p değerleri istatistiksel olarak anlamlı kabul edildi.

Bulgular: Graf USG eğitiminden önceki değerlendirmede bulgular belirlendi. Gözlemciler arası güvenirlilikte, Sınıf içi sınıflama korelasyonları (ICC) değerleri, sağ kalçalar için Graf yöntemiyle değerlendirmede USG sınıflamasında 0,939 idi. ICC değerleri, sol kalçalar için Graf yöntemiyle değerlendirmede USG sınıflamasında 0,907 idi.

Graf USG sınıflaması eğitiminden sonraki değerlendirmede, bulgular belirlendi. Graf yöntemiyle USG sınıflama değerlendirmesinde sağ kalçalar için ICC değerleri 0,975 idi. Graf yöntemiyle USG sınıflama değerlendirmesinde sol kalçalar için ICC değerleri 0,970 idi. Bütün p değerleri anlamlıydı.

Graf USG yöntemi için eğitimden önce ICC değerleri sağ kalçalarda 0.838, sol kalçalarda 0.765 iken, eğitim sonrası sağ kalçalarda 0.928, sol kalçalarda 0.915 idi.

Sonuç: USG ile GKD deformitesinin tanısı ciddi bir eğitim dönemi gerektiren karmaşık ve zor bir durumdur. Graf USG yöntemi yüksek gözlemciler arası güvenirliliğe sahip bulundu. Ayrıca Graf USG sınıflamasına yönelik eğitimin, özellikle çocuk doktorları arasında gözlemci içi güvenilirliği arttırdığını tespit ettik.

Anahtar Kelimeler: Gelişimsel kalça displazisi, ultrasonografi, Graf, gözlemciler arası güvenirlilik, gözlemci içi güvenirlilik

Corresponding Author: Kadri Yıldız, Kafkas Üniversitesi Tıp Fakültesi, Ortopedi ve Travmatoloji Anabilim Dalı, Ana Kampüs, 36000, Kars, Türkiye
E-mail: drkadri1980@hotmail.com
Received: 18/01/2020 **Accepted:** 17/02/2020 **Doi:** 10.32322/jhsm.676820

Cite this article as: Yıldız K, Bekis Bozkurt H, Çetin T, Yıldız V. Interobserver reliability in the ultrasonography evaluation with Graf method of developmental dysplasia of the hip: the importance of education for ultrasonography classification. J Health Sci Med 2010; 3(2): 121-124.

INTRODUCTION

Developmental dysplasia of the hip (DDH) deformity is a possible disability condition of the hip (1). And it can result in gait dysfunction and severe functional limitations. Although surgical treatment for developmental hip dysplasia tends to decrease by the Graf ultrasonography (USG) methods, still more surgical treatment is an option for DDH management (2). Although early radiographs cannot be applied because of the radiation risks, hip ultrasound is a qualified diagnosis method. Developmental dysplasia of the hip measurements utilizing USG evaluation with Graf method and Graf DDH classification views allow one to characterize the hip deformity (3).

The purpose of this study was to determine interobserver reliability with varying levels of observer experience on Graf USG classification. And it aims to detect the importance of education on Graf USG classification personally.

METHODS

USG evaluation according to Graf measurements on 62 patients who presented to the clinic with complaints of DDH were examined. The group was chosen randomly. USG views were evaluated at the routine clinical trials. USG evaluation and classification according to Graf method were made by a 15-year experienced orthopedic surgery specialist, a 25-year experienced radiologist and an 11-year experienced pediatric specialist. Each observer obtained Graf USG education firstly by individual reading. This reading session was based on Graf's Hip USG evaluation of literature. And these articles were discussed with the senior author prior to making measurements. Measurements were made by each observer using the computer tools of our hospital. Each observer made measurements and classification independently and blindly. All measurements were made on three separate occasions with the order of images randomized.

These databases were collected from measurements made by an orthopedic surgery specialist who had been orientated on pediatric orthopedia, a pediatric specialist and a pediatric radiologist. These measurements were made in separate settings and were analyzed to determine interobserver reliability. A statistician at our institution's statistical consulting center determined interobserver reliability between three measurements. And then the results were compared to the same measurements made on USG.

Statistics

The statistical measurements were made by SPSS 20.0 (Windows, IL, USA). Intraclass Correlation Coefficient was determined. 95% Confidence Interval and F Test with True Value 0 were detected. P values were accepted as significantly less than 0,05.

Ethical Declaration

In this study, national and international ethical rules are observed. Ethical approval was obtained for this study from the Committee of Ethics Committee of Kafkas University, School of Medicine (Date 26/03/2019, 05/ 80576354-050-99/174 number and decision no).

RESULTS

At the evaluation before USG education, the results were established. In the interobserver reliabilities, Intraclass Classification Correlations coefficient (ICC) values were 0,939 for Graf USG classification evaluation for right hips. ICC values were 0,907 for USG classification evaluation with Graf Method for left hips. **Table 1** and **2** show these values along with the difference in interobserver reliability between three specialists before USG education.

At the evaluation after USG education, the results were established. ICC values were 0,975 for USG classification evaluation with Graf Method for right hips. ICC values were 0,970 for USG classification evaluation with Graf Method for left hips. **Table 3** and **4** show these values along with the difference in interobserver reliability between three specialists after USG education.

The interobserver reliabilities for all measurements on USG were statistically equal to zero according to p-values (**Table 1, 2, 3** and **4**). The difference between sessions as before and after education are statistically significant. The measurements had consistent high interobserver reliability on USG with Graf method. These values are listed in all tables, and all measures were significant as statistically (p=0.000). Intra Class Correlations were 0,838 in right hips, 0,765 in left hips before education; and 0,928 in right hips, 0,915 in left hips after education on Graf USG classification. All p values were equal to zero (p=0.000).

DISCUSSION

Hip ultrasonography for DDH diagnosis was first introduced in the 1980s by Graf. The Graf USG classification of hip development has improved early detection and accuracy for DDH (4). In 1992, Matos et al (5) stated that ultrasound of the infantile hip using the Graf USG method was an important tool for the assessment of developmental dysplasia of the hip and it could be used interpretation training and frequent use. Spaans et al (6) reported a correlated relationship between USG and radiographic imaging outcomes, in patients for treatment and follow-up of DDH.

In the criteria of Graf USG classification, there are four important points: i) angles (alpha and beta), ii) labrum position, iii) position of femoral head, iv) the ossification of the femoral head. The most important point is to make a differential diagnosis among type 2b and type 2c (7). Because patients with Type 2c needs treatment. In a study, 4222 hips were evaluated for DDH by Fan et al. (7). They emphasized that femoral head coverage can be used as a reference indicator for DDH classification. And they established femoral

Table 1. Right hip Graf USG evaluations before education

	Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	Lower Bound	Upper Bound	
Single Measures	,838[a]	,335	,941	57,905	61	122	,000
Average Measures	,939[c]	,602	,979	57,905	61	122	,000

Two-way mixed effects model where people's effects are random and measure effects are fixed.
a. The estimator is the same, whether the interaction effect is present or not.
b. Type A intraclass correlation coefficients using an absolute agreement definition.
c. This estimate is computed assuming the interaction effect is absent because it is not estimable otherwise.

Table 2. Left hip Graf USG evaluations before education

	Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound		Lower Bound	Upper Bound	
Single Measures	,765[a]	,349	,898	,765[a]	,349	,898	,765[a]
Average Measures	,907[c]	,617	,964	,907[c]	,617	,964	,907[c]

Two-way mixed effects model where people's effects are random and measure effects are fixed.
a. The estimator is the same, whether the interaction effect is present or not.
b. Type A intraclass correlation coefficients using an absolute agreement definition.
c. This estimate is computed assuming the interaction effect is absent because it is not estimable otherwise.

Table 3. Right hip Graf USG evaluations after education

	Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound		Lower Bound	Upper Bound	
Single Measures	,928[a]	,885	,956	,928[a]	,885	,956	,928[a]
Average Measures	,975[c]	,958	,985	,975[c]	,958	,985	,975[c]

Two-way mixed effects model where people's effects are random and measure effects are fixed.
a. The estimator is the same, whether the interaction effect is present or not.
b. Type A intraclass correlation coefficients using an absolute agreement definition.

Table 4. Left hip Graf USG evaluations after education

	Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound		Lower Bound	Upper Bound	
Single Measures	,915[a]	,859	,949	,915[a]	,859	,949	,915[a]
Average Measures	,970[c]	,948	,982	,970[c]	,948	,982	,970[c]

Two-way mixed effects model where people's effects are random and measure effects are fixed.
a. The estimator is the same, whether the interaction effect is present or not.
b. Type A intraclass correlation coefficients using an absolute agreement definition.
c. This estimate is computed assuming the interaction effect is absent because it is not estimable otherwise.

head coverage (FHC) at different positions corresponds to different reference values, and they stated Neutral and Flexion Positions (FHC-D) can be used as a quantitative indicator for assessment of hip stability (8). Choudry et al. (9) founded the positive predictive value (PPV) of clinical screening was found as 4.0% and the PPV of sonography was found as 16.1%. They evaluated previously published 10 or 15-year studies, and they found a deterioration in the PPV in those with potential instability of the hip. These findings were evaluated as a paradox Roposch et al. (10) have found that Graf classification showed moderate reliability. And they emphasized using self-study was not quite as effective as by a structured program. In another study about the USG evaluation for DDH, Kolb pointed to the examiner should pay attention to avoid transducer inclinations in the frontal plane and a combination of posterior and cranial inclination (11). Bilgili et al. (12) studied the reliability of computer-assisted and manual measurement methods for assessment of Graf type 1 and type 2 hip sonograms. They have established the alpha angle measurements had a high concordance, but low concordance for beta angle measurements. The two measurement methods were reliable and consistent with each other.

In this study, the interobserver reliability of the Graf USG classification among three-branch specialists before and after education was evaluated. The evaluation consisted of two steps. The first step was self-study and the second step two was learning the method by using standard teaching material developed by Graf USG professional personal trainer. Training studies were assumed at the same by three observers as performed self-study (before USG education) and teaching sessions (after USG education). Interobserver reliability was 0.602 (95% CI=0.335-0.979) for right hips, and 0.617 (95% CI=0.349-0.964) for left hips. After education, interobserver reliability was 0.958 (95% CI=0.885-0.985) for right hips, and 0.948 (95% CI=0.859-0.982) for left hips. All p values were equal to zero (p=0,000). Intra class correlation was 0,838 in right hips, 0,765 in left hips before education for Graf USG classification, and 0,928 in right hips, 0,915 in left hips after education for Graf USG classification.

In this study, our intention was to assess the interobserver reliabilities on the differing capability of three observers on Graf USG classification. Our hypothesis was that the USG education for the Graf classification method was better in improving interobserver reliability as statistically significant. Education increased the ability to make a differential diagnosis between Type 2b and 2c. And this effect reflected in conservative treatment decisions. Also, the variability in the three evaluations reflects each individual's experience and training level.

The most important point that must be emphasized is the misclassification may effect treatment incorrectly. An education for Graf classification can increase interobserver reliability, and it can provide a safer differential diagnosis for Type 2b and 2c for early diagnosis and treatment.

Limitation of this study is that the USG measurements were made in separate settings, which could cause bias among observers. Also, the study group could be larger.

CONCLUSION

We believe that the USG evaluation chosen to characterize developmental dysplasia of the hip deformity should be easy to teach to observers and have high reliability. This is the only way that results from different specialists can be meaningfully compared. These measurements must also be simple to perform uniform and adequate treatment.

We found that the Graf USG method is the most reliable measurement for early diagnosis and treatment on DDH management. Better education for Graf USG classification provides more accurate USG evaluation for early treatment.

DECLARATION OF INTEREST STATEMENT

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

REFERENCES

1. Kotlarsky P, Haber R, Bialik V, Eidelman M. Developmental dysplasia of the hip: What has changed in the last 20 years? World J Orthop 2015 Dec 18; 6: 886–901.
2. Alexander NB, Goldberg A. Gait disorders: Search for multiple causes. Cleveland Clin J Med 2005 July; 72: 586-600.
3. Bracken J, Ditchfield M. Ultrasonography in developmental dysplasia of the hip: What have we learned? Pediatric Radiol 2012 Sep; 42: 12.
4. Lussier EC, Sun YT, Chen HW, Chang TY, Chang CH. Ultrasound screening for developmental dysplasia of the hip after 4 weeks increases exam accuracy and decreases follow-up visits. Pediatr Neonatol 2018 Jul; 18: 26; 1-8.
5. Matos MA. Ultrasonographic assessment of infantile hip by the Graf method. Rev Assoc Med Bras 2006 Jan-Feb; 52: 53-5.
6. Spaans AJ, Beek FJ, Uiterwaal CSPM, Pruijs JEH, Sakkers RJ. Correlation between ultrasonography and radiographic imaging of developmental dysplasia of the hip. J Child Orthop 2019 Apr 1; 13: 155-60.
7. Feldman D, Nasiadko CN, Riga LV. Ultrasonographic evaluation of developmental dysplasia of hip-ddh. Why? How? What does it mean? Poster No: C-0155 Congress: ECR 2016 pages: 1-18.
8. Fan W, Li XJ, Gao H, Yi X, Liu QJ. Exploration of femoral head coverage in screening developmental dysplasia of the hip in infants. J Med Ultrason 2019 Jan; 46: 129-35.
9. Choudry QA, Paton RW. Neonatal screening and selective sonographic imaging in the diagnosis of developmental dysplasia of the hip. Bone Joint J 2018 Jun 1; 100: 806-10.
10. Roposch A, Graf R, Wright JG. Determining the reliability of the Graf classification for hip dysplasia. Clin Orthop Relat Res 2006 Jun; 447: 119-24.
11. Kolb A, Benca E, Willegger M, Puchner SE, Windhager R, Chiari C. Measurement considerations on examiner-dependent factors in the ultrasound assessment of developmental dysplasia of the hip. Int Orthop 2017 Jun; 41: 1245-50.
12. Bilgili F, Bilgili CO, Cetinkaya E, et al. Reliability of computer-assisted and manual measurement methods for assessment of Graf type 1 and type 2 hip sonograms. J Ultrasound Med 2016 Jun; 35: 6: 1269-75.