SYSTEMATIC REVIEW

Diagnostic Accuracy of Ultrasonography Method of Graf in the Detection of Developmental Dysplasia of the Hip: A Meta-Analysis and Systematic Review

Mohammadreza Chavoshi, MD[1]; Seyed Ali Mirshahvalad, MD, MPH[2]; Mehrzad Mahdizadeh, MD[3]; Fatemeh Zamani, MD[3]

Research performed at Children's Medical Center, Tehran University of Medical Sciences, Tehran, Iran

Abstract

Background: The present study aimed to perform a meta-analysis on the performance of Graf's ultrasonography method in the detection of developmental dysplasia of the hip (DDH).

Methods: A query was conducted on electronic bibliographic databases until the end of October 2020. The inclusion criteria entailed: 1. the use of Graf method in less than 12 weeks of age, 2. the use of follow-up as reference test, and 3. provision of crude data. Pooled diagnostic performance measures were calculated. Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) checklist was utilized to assess the quality of the included studies. The hierarchical summary receiver-operating characteristic (HROC) curves were also drawn.

Results: Six articles (including seven populations, 11,012 patients) were considered eligible. The pooled sensitivity and specificity were obtained at 93% (95% CI: 0.57-0.99) and 97% (95% CI:0.86-0.99), respectively (area under curve= 0.99). The pooled positive and negative likelihood ratio, as well as diagnostic odds ratio, was reported as 28.4, 0.07, and 396, respectively.

Conclusion: As evidenced by the obtained results, Graf's method is a useful ultrasonography technique with acceptable accuracy for screening DDH in neonates. However, there are uncertainties about the best population and age for screening. Furthermore, more attention should be paid to the proper training of this method to reduce the number of operator errors.

Level of evidence: I

Keywords: DDH, Diagnostic performance, Graf, Meta-analysis, Pediatrics, Ultrasonography

Introduction

Developmental dysplasia of the hip (DDH) is a hip pathology in which the proximal femur and acetabulum cannot make normal maturity progress, resulting in malformation and instability of the hip (1, 2). Although there are known risk factors for DDH (e.g., female gender, positive family history, breech position, congenital deformities), the majority of infants may not have any known risk factors.

Both missing and over-diagnosis of DDH can impose a substantial burden on health care systems; moreover, late diagnosis results in huge costs and treatment complexities. For instance, patients can usually be treated non-surgically in the first three months (3, 4). On the other hand, complications of unnecessary treatments and costs of ultrasonography are of utmost importance for screening. Therefore, an accurate and cost-effective modality should be selected for the detection of DDH.

Corresponding Author: Fatemeh Zamani, Department of Radiology, Children's Medical Center, Tehran University of Medical Sciences, Tehran, Iran
Email: doctorzamani@gmail.com

Physical examinations (e.g., Ortolani and Barlow tests) are recommended as the initial evaluation for each infant. It has been observed that approximately half of the infants would be missed in the initial examination. Moreover, X-ray-based modalities, regardless of their radiation harm, cannot be helpful in early diagnosis since the needed bone parameters appear after four months of age (5). Therefore, ultrasonography is a valuable tool for the detection of DDH in the first months, evaluating bone and cartilage components (6, 7). Ultrasonography is more commonly used in infants with risk factors or any degree of instability upon examination, called selective screening. In another program, universal screening is the primary modality for the evaluation of newborns, regardless of their history (8).

Ultrasonography is an operator-dependent modality requiring a skilled operator to reach an acceptable performance in abnormality detection. Therefore, a set of criteria should be developed to reach a diagnosis. In the 1980s, professor Graf published a technique to be used as a diagnostic tool based on different introduced angels (9). Although some modifications or new methods (e.g., modified Graf, dynamic, Harcke, Terjesen) have been introduced, Graf's technique is the most commonly utilized method.

In 2005, Woolacott et al. published a systematic review consistent with the aim of the present study. Nonetheless, only one study was eligible based on their inclusion criteria, and it was not possible to perform a meta-analysis (10). To the best of our knowledge, the current study is the first meta-analysis to calculate the performance of Graf's ultrasonography method in the detection of DDH. It is believed that the added value can be utilized in screening guidelines to adopt an efficient approach.

Materials and Methods

The present study was designed based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols (PRISMA-P) checklist (11).

Search strategy

A systematic query was conducted on electronic bibliographic databases, including the Scopus, Web of Science (Science and Social Science Citation Index), and PubMed, for the English-language original peer-reviewed full-length articles until the end of October 2020. According to each database-specific Boolean search strategies, following terms were used: ("hip dysplasia" OR DDH OR "dysplasia of the hip") AND (pediatric OR children OR neonate OR newborn OR infant) AND (ultrasound OR ultrasonography). An extra search was performed through the added articles' references to augment the search. There was no time limitation for the articles.

Study selection

The inclusion criteria were as follows:
1. The use of ultrasonography Graf method to diagnose the DDH in neonates less than 12 weeks
2. Diagnosis documentation by clinical or radiologic (US or conventional pelvic X-ray) follow-up examinations

3. Provision of adequate data to extract the true positive, true negative, false positive, and false negative

On the other hand, the exclusion criteria entailed: Abstracts without full articles, unpublished studies, notes, letters, comments, conference articles, the studies which used methods other than follow-up exam as the reference test, studies using other ultrasonography methods to diagnose DDH, studies without extractable crude data, and studies with unclear follow up strategies. The titles and abstracts were blindly reviewed by two authors (M. C. & S. M.) after the removal of the duplicated articles to check the eligibility of studies. Any study that received two positive reviews entered the full-text review phase. All articles in this phase were reviewed by both reviewers, and the article was considered eligible if both come to a consensus. Discrepancies were solved by consensus or referral to a third reviewer (M. M.).

Data extraction

The following data were extracted from each article: title, abstract, authors' name, journal title, year of publication, details of study design, study population, patients sample size, demographic features (such as numbers of patients, mean age, percentage of each gender), ultrasound technique, week of ultrasound, type of follow up (clinical or radiologic or both), inclusion and exclusion criteria, values of true positive, false positive, true negative, and false negative, as well as sensitivity, specificity, positive predictive value, negative predictive value, and accuracy. Detailed data of each subgroup of patients (gender, different ages, etc.) were also extracted if provided.

Assessment of methodological quality

Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) checklist was used to assess the quality of studies, utilizing Review Manager 5.2 software. Each question was specified based on the study and was assigned a response of yes, no, or unclear.

Statistical analysis

Mean and percentage were reported for continuous or categorical variables, respectively. Diagnostic performance measures, including sensitivity, specificity, and diagnostic odds ratio (DOR), were calculated by a hierarchical method according to the two-by-two contingency tables. A bivariate model was used to find the summary points for sensitivity and specificity and their 95% confidence intervals (CI), considering the inter-and intra-study heterogeneity. The hierarchical summary receiver-operating characteristic (HROC) model was performed to construct the HSROC curve and area under the curve (AUC). The 95% confidence region and prediction region were calculated to demonstrate the uncertainty degree of the summary sensitivity and specificity (12, 13). Higgins' I^2 statistics and Cochran's Q test were employed to evaluate the heterogeneity of studies for DOR, pooled sensitivity, and specificity. Furthermore, a linear correlation between sensitivity and false-positive was calculated

to assess the threshold bias. The result of r ≥ 0.6 was considered significant.

Deeks' funnel plot was used to assess the publication bias. A *P-value* of < 0.10 for the slope coefficient was regarded as significant(14). All analyses were conducted in R-package mada software (version 4.0.2, R Foundation for Statistical Computing, Vienna, Austria) and the "Midas" module of STATA 16 (15, 16).

Results

Study characteristics

A total of 1,504 studies were retrieved after searching the databases. Thereafter, 627 duplicated articles were removed. Title and abstract review were performed on 858 studies, and 156 studies were selected for full-text review. Finally, six articles were considered eligible to be included. One of the studies had provided separate data for two specific populations (patients who had ultrasonography test at four weeks of age and who had the test on eight weeks). Both populations fulfilled our inclusion criteria, and the data were provided independently for each group; therefore, we included them separately in the study (17).

Three different age groups were evaluated in one study. There was an overlap of the populations among these three groups; therefore, we only included the group with the most population (18). Finally, seven populations (11,012 patients, from 6 studies) were entered into the final analysis (17-22). The flow chart of the study selection process is illustrated in Figure 1. Among these six studies, the ultrasonography was performed in the second month of life in three studies, while in one study, it was conducted on the first three days of life, and two articles had not heightened the exact age. The characteristics of included studies are presented in Table 1. Although not quite satisfactory about the number of studies, a subgroup analysis was performed on the three studies with second-month data, and pooled sensitivity and specificity were calculated.

The QUADAS-2 checklist was assessed for all studies [Figure 2]. A primary concern in diagnostic test studies is referral bias, especially in retrospective studies. The reference standard in the current study was clinical
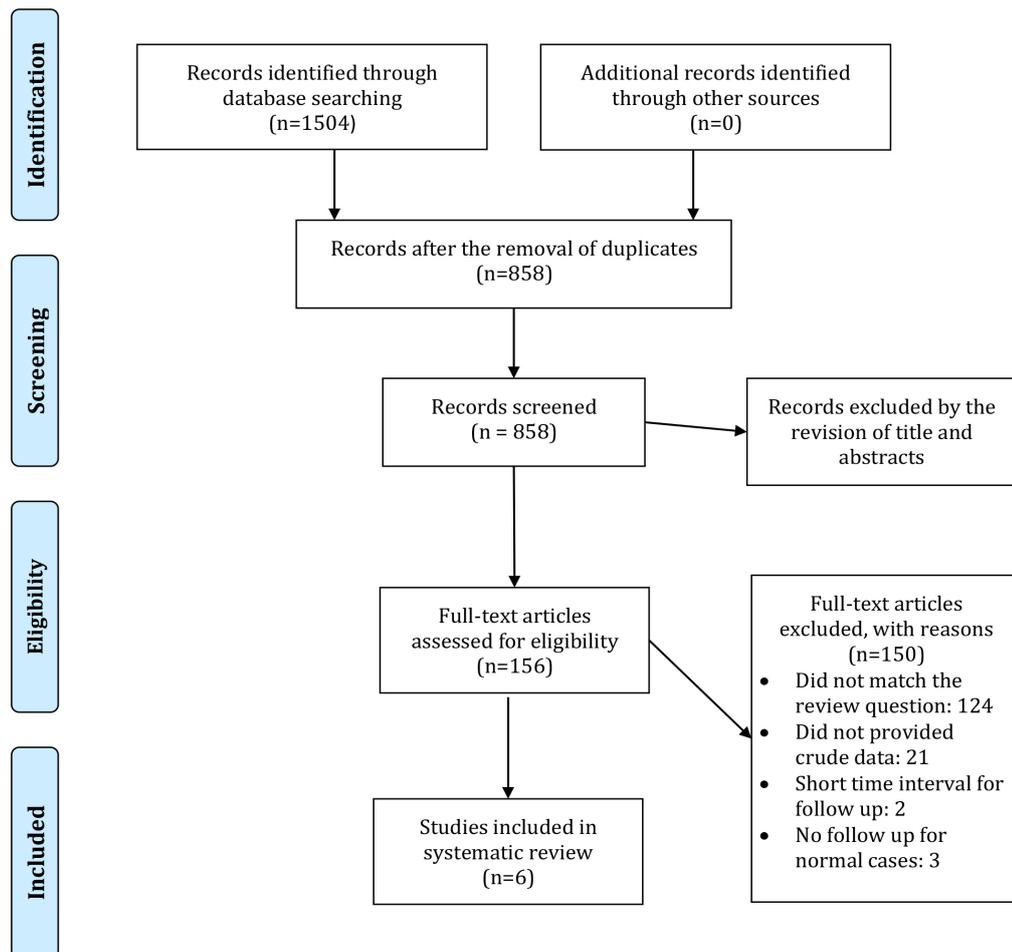

Figure 1. Flow chart of the study selection process.

PERFORMANCE OF GRAF'S ULTRASONOGRAPHY IN DEVELOPMENTAL DYSPLASIA OF THE HIP

Table 1. characteristics of the included studies

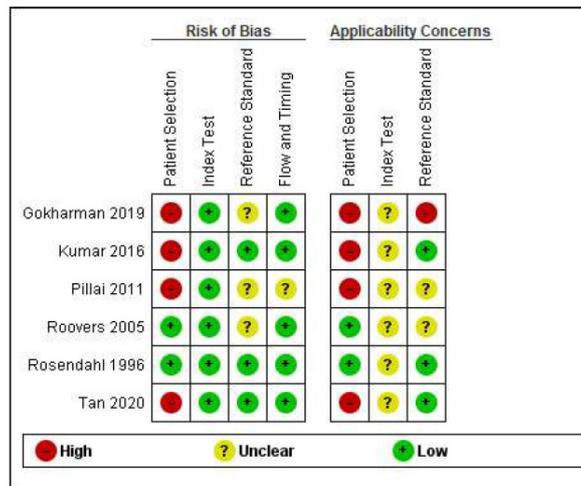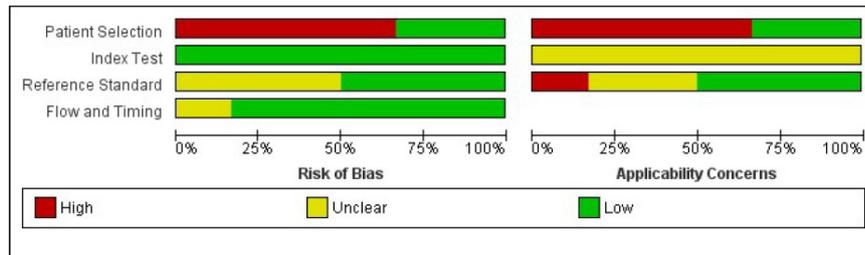| Author, year	Type of study	ultrasonography device	Population of study	Total population	Male percent	Operator	Cut-off for US grading	Reference standard test	Age of ultrasonography
Rosendahl, 1996 (16)	Cohort	NA	Universal screening		51	Radiologist	IIb	Clinical follow up for 3 years	1-3 days
Roovers, 2005 (18)	Retrospective cohort	Hitachi (EUB-405) with a linear transducer (7.5 or 5 MHz)	Universal screening	5170	NA	Radiologist technician	IIb	Follow up US in 8 months	First three months of life
Pillai, 2011 (15)	Cohort	Acuson™ linear 7 MHz probe	Selective screening	249	34.3	Orthopedic consultant	III	AP pelvic radiograph at 6 months	29-77 days
Kumar, 2016 (19)	Retrospective cohort	NA	Selective screening	662	NA	Radiologist	IIa	At least 2-year follow up	6 weeks
Gokharman, 2019 (14)	Cohort	Toshiba Xario, 7.5-MHz linear transducer, and Toshiba Aplio 300, 10-MHz linear transducer	Selective screening	360 (4 weeks), 819 (8 weeks)	47	Radiologist	IIc	US follow up at 12 weeks	4 and 8 weeks
Tan, 2020 (17)	Retrospective cohort	NA	Selective screening	160	NA	Radiologist	IIb	Pelvis radiograph at 1 year	<12 weeks

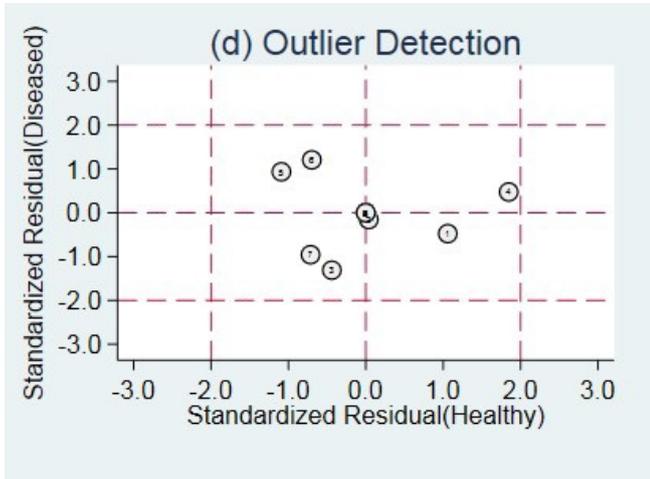Figure 2. Quality Assessment of Diagnostic Accuracy Studies-2 checklist.

Figure 3. Outlier detection analysis.

and/or radiologic follow-up. These follow-up tests were performed by various clinicians or radiologists; therefore, inter and intra-observer reliability was unclear in all studies.

Diagnostic performance
Outlier detection analysis demonstrated that none of the included studies was outlier [Figure 3]. The pooled sensitivity and specificity of the studies were obtained at 93% (95% CI: 0.57-0.99) and 97% (95% CI: 0.86-0.99), respectively [Figure 4]. The hierarchical summary ROC curve showed the AUC of 0.99 (95% CI: 0.97-0.99) [Figure 5]. The pooled positive and negative

likelihood ratios, as well as DORs, were reported as 28.4 (95% CI: 6.2-130.3), 0.07 (95% CI: 0.01-0.62), and 396 (95% CI: 25-6313), respectively. Subgroup analysis illustrated that the pooled sensitivity and specificity were 88% (95% CI: 0.19-1) and 91% (95% CI: 0.86-0.93), respectively, in ultrasonography test in the second month of life. The summary point of likelihood ratios in the left upper quadrant indicated that the Graf ultrasonography method could both confirm and exclude the DDH in neonates and is a suitable screening method [Figure 6].

Heterogeneity assessment
Cochrane's Q test and Higgins' I^2 statistics are generally used to evaluate the heterogeneity of the meta-analysis. Nevertheless, there are some considerations in their interpretation in diagnostic accuracy meta-analysis. The threshold effect is not considered in these tests, and assessment of the heterogeneity is performed with a single outcome variable (23). Consequently, there is no consensus on a single method to evaluate heterogeneity in diagnostic test studies (24). Therefore, all available tests should be used to assess heterogeneity (13).
According to the Cochran's Q test and Higgins' I^2 statistics, a notable heterogeneity was observed in both sensitivity and specificity between the studies (Cochrane Q test=69, df=6, *P=0.00*, I^2=91.3 for sensitivity, and Cochrane Q test=713.27, df=6, *P=0.00*, I^2=99.16 for specificity). On the other hand, the only single outcome variable in diagnostic test accuracy was DOR, and the heterogeneity assessment based on DOR pointed out low heterogeneity (Cochrane Q test=8.476, df=5, *P=0.132*, I^2=41, tau2: 4.085 CI (95%):0-52.737). However, heterogeneity is unavoidable in diagnostic
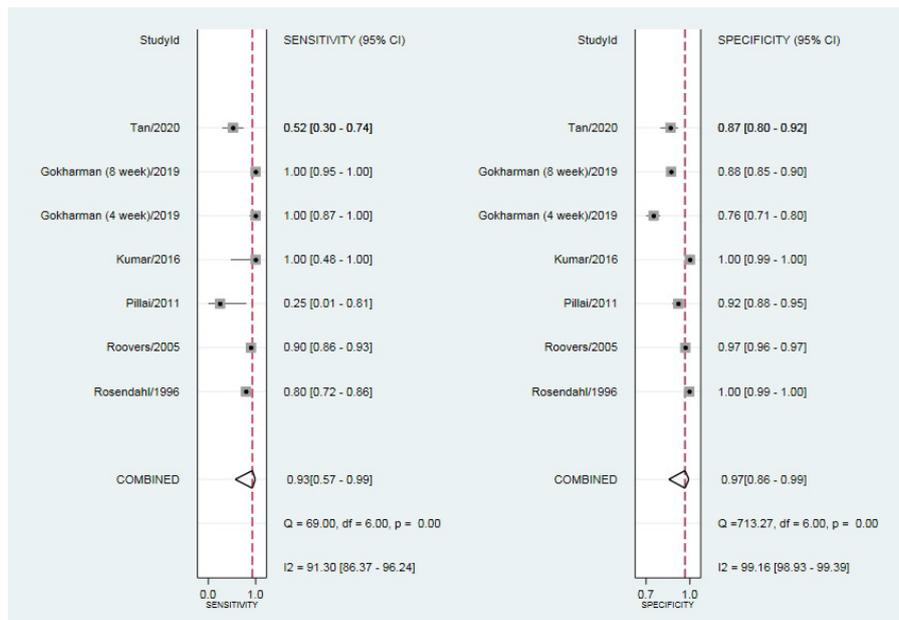

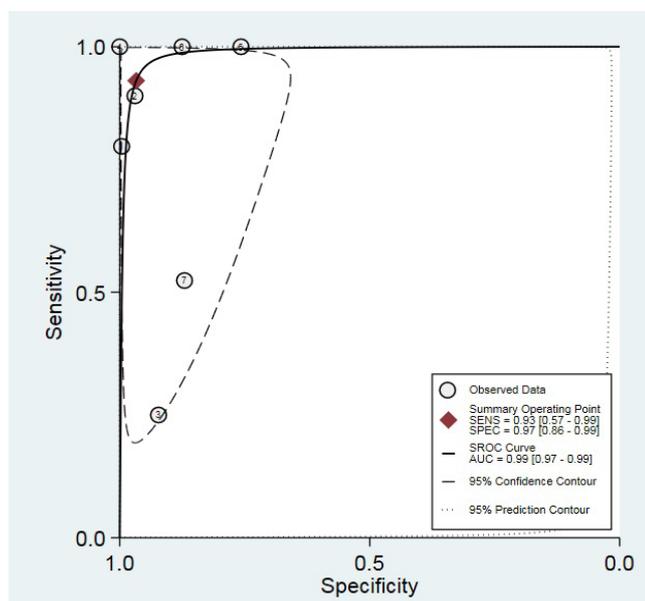**Figure 4. Pooled sensitivity and specificity of the studies.**

Figure 5. Hierarchical summary receiver-operating characteristic curve (HSROC).

test meta-analysis due to variable factors. The wide range of prediction regions indicated this issue. The threshold effect analysis revealed that the proportion of heterogeneity likely due to the threshold effect was 0.00, and the spearman correlation of sensitivities and false-positive rates resulted in r= -0.179 (13).

Publication bias

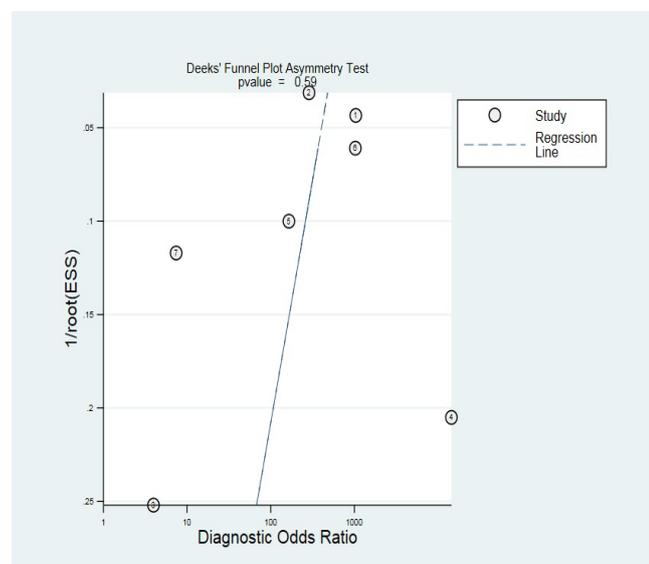Deeks' test displayed no significant publication bias among the studies (*P =0.59*) [Figure 7].

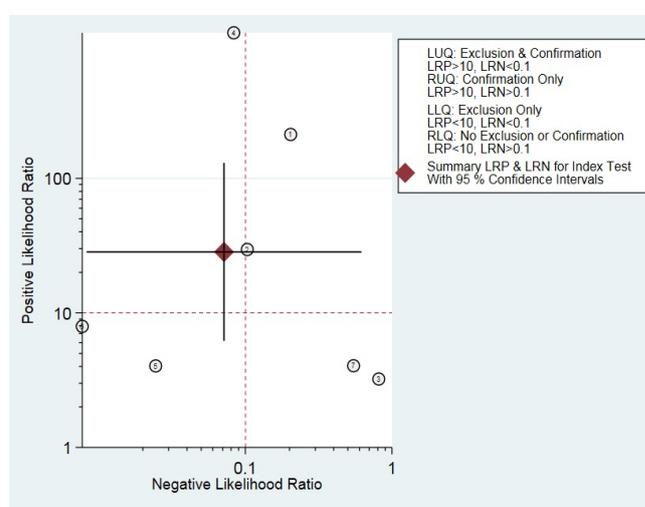**Figure 7. Deeks' funnel plot for publication bias assessment.**

Figure 6. Likelihood ratio scatter-gram.

Discussion

Graf's ultrasonography method has been long utilized for screening and diagnosis of DDH. Later, it was merged with Harcke's technique to obtain higher accuracy by evaluating the hip's dynamic properties. Although the reliability and validity of this method have been called into question, the findings of the present study, as well as the pooled sensitivity of 93% and specificity of 98%, demonstrated that this ultrasonography method is an acceptable modality for both confirmation and screening goals (25).

Timely diagnosis, particularly in children, significantly affects the treatment and quality of life. DDH, as a common congenital disability in newborns, can lead to a tremendous burden and debilitation if neglected. There is still some debate about the best screening program for DDH diagnosis, considering such issues as the accuracy of the test, accessibility, and cost. Although clinical screening is a low-cost and accessible method, the high false-negative ratio and variability among the examiners make it less acceptable as a screening method (26). In a similar vein, radiography is not preferred due to irradiation and the non-visibility of cartilages. Nonetheless, it is a useful modality for follow-up and evaluation of response to treatment in older children (6).

There is a dearth of evidence-based guidelines to discuss the best approach and timing for using ultrasonography in the screening of DDH. Apart from the Ortolani/ Barlow tests, two different ultrasonography programs of universal and selective screening have been used in screening policies worldwide, pointing to the variety of advantages and drawbacks (26). Due to the low incidence rate of DDH in patients without known risk factors, the performance of ultrasonography in all newborns may have higher costs, compared to routine clinical examinations. On the other hand, leaving children untreated may result in irreparable damages, a seven-fold increase in treatment expenses, and a

prolonged duration of the disease (1, 5).

Furthermore, it is noteworthy that there is no worldwide consensus on the risk factors for selective screening, particularly female gender, and this may significantly affect the outcomes of this program. As mentioned in the study conducted by Wilf-Miron et al., although the country's selected policy was selective screening, the executed pattern was closer to the universal policy (1). Clegg et al. reported that the costs of both policies are comparable when considering both treatment and ultrasonography expenses (27).

Thaler et al. assessed the screening programs in Austria and reported that the increased expenses of screening by addition of universal policy to the routine clinical examination are justifiable, decreasing the cost of treatment by about £60000 per year (8). On the other hand, in their study, Holen et al. emphasized the importance of clinical screening quality as a critical point in policy planning, pointing to the marginal benefit of universal screening in programs with high-quality clinical examinations (28). Along the same lines, Buonsenso et al. reported a significant number of cases benefited when using the universal program, while missed in the selective policy (29).

Laborie et al. reported that universal screening is not a cost-effective policy; nevertheless, they supported the idea that ultrasonography should be performed for all girls in high-incidence regions (30). Westacott et al. and Rosendahl et al. reported a similar late DDH presentation rate in both programs (3, 31). However, this finding was in line with the fact that universal screening significantly reduces the age at detection and treatment costs. Regarding the appropriate age for ultrasonography examination, although one consensus has reached to perform ultrasonography at the age of six weeks in the absence of risk factors, there are still different suggestions in the literature, ranging from ultrasonography at birth to six months of age (31).

In the absence of no generally acceptable consensus on the best week for performing ultrasonography, repeated tests would perform unnecessarily to reach the diagnosis, increasing the financial burden on the health care system. Although ultrasonography in older newborns shows better performance, the chance of on-time non-invasive treatment may be ruined. On the contrary, performing ultrasonography in the early days may result in unnecessary treatment and complications (32). Moreover, in immature hips, which include approximately 20% of all, a recheck examination might be necessary at older ages for confirmatory diagnosis, burdening the healthcare and increasing the costs (5).

Tan et al. concluded that performing ultrasonography before the fourth week is not reliable, showing a low correlation with 1-year radiography results (20). They indicated that the best result was obtained from the 5th weeks of age with no false result reported by per week analysis. Gokharman et al. pointed out that comparing the fourth and eighth-week ultrasonography test results, the eighth-week results were closer to the 12th-week follow-up ultrasonography as the reference test

(17). Laborie et al. put forward different arguments about postponing ultrasonography to six weeks of age. They believed that the concept of watchful waiting until the 6th week might result in delayed treatment, unreachable personalized decision management, increased costs, and missed neonates due to lack of parenthood compliance (30).

The wide range of ages in our included studies and the insufficient number of studies in each group prevents us from identifying the appropriate age. Nonetheless, a subgroup meta-analysis of three studies on neonates in the second month revealed a sensitivity and specificity of 88% and 91%, respectively. Due to the small number of studies, this result is not generalizable, and more age-specific studies are necessary to reach a reliable result. In a meta-analysis of common risk factors for DDH, it was proposed that early screening should be considered in newborns with following risk factors: breech presentation, female, left hip affected, first born and family history of DDH (33).

An influential factor in performing ultrasonography test and its results is a wide range of inter and intra-observer variabilities in the dysplasia metrics (34). Due to the existence of different methods and facilities for ultrasonography, as well as the level of expertise, these variabilities may become more sweeping in range. This issue also plays a significant role in the heterogeneity of the present study results. However, it is reported that the prerequisite for standardization of Graf's method is knowing the correct execution which is described by Graf and other experts, regardless of the interpreter's expertise (26). Mostofi et al. reported that this problem can be overcome by three-dimensional ultrasonography, demonstrating that novice operators with no or little experience can have a relatively similar performance to the experts (35).

As mentioned earlier, in terms of quality assessment, there is imperfect reporting in the published articles in the field of ultrasonography utilization in the diagnosis of DDH (22). This issue affected our review, and one of the notable limitations in the present study was heterogeneity in the included articles. Moreover, the reference standard of the current study was not a gold standard, which was set as follow up with different methods, including clinical, radiography, or ultrasonography itself. Therefore, it should be considered that true positive cases may cause overestimation and do not represent the actual true positives, including overtreatments. In a similar vein, based on the follow-up duration, the included studies may miss the late presenting DDH cases, representing after the end-points of the studies.

Furthermore, there were different cutoffs for taking into account Graf's classification as abnormal. Some studies used the threshold of orthopedic devices, while others considered classifications of IIa, IIb, IIc, or even III as abnormal. Regarding ultrasonography policy used in the studies, both universal and selective programs which also increased heterogeneity were included. Variability among ultrasonography devices and operators also affected the results.

To overcome the limitations of the present study, further cohorts with large populations and strong perspectives are needed. In addition, reaching a consensus on the best timing for ultrasonography is of great importance. In conclusion, it was found that Graf's method of ultrasonography test is a useful method with acceptable accuracy for screening the DDH in neonates; however, there are uncertainties about the best population and age for screening. Furthermore, more attention should be paid to the proper training of this method to reduce the number of operator errors.

Conflicts of interest: The authors report no conflict of interest concerning the materials or methods used in this study or the findings specified in this paper.

Disclosure: No financial or nonfinancial benefits have been received or will be received from any party related directly or indirectly to the subject of this article.

Mohammadreza Chavoshi MD[1]
Seyed Ali Mirshahvalad MD MPH[2]
Mehrzad Mahdizadeh MD[3]
Fatemeh Zamani MD[3]
1 Department of Radiology, Shariati Hospital, Tehran University of Medical Sciences, Tehran, Iran
2 Department of Nuclear Medicine, Shariati Hospital, Tehran University of Medical Sciences, Tehran, Iran
3 Department of Radiology, Children's Medical Center, Tehran University of Medical Sciences, Tehran, Iran

References

1. Wilf–Miron R, Kuint J, Peled R, Cohen A, Porath A. Utilization of ultrasonography to detect developmental dysplasia of the hip: when reality turns selective screening into universal use. BMC pediatrics. 2017;17(1):136.
2. Vafaee AR, Baghdadi T, Baghdadi A, Jamnani RK. DDH epidemiology revisited: do we need new strategies? Archives of Bone and Joint Surgery. 2017;5(6):440.
3. Westacott DJ, Butler D, Shears E, Cooke SJ, Gaffey A. Universal versus selective ultrasound screening for developmental dysplasia of the hip: a single-centre retrospective cohort study. Journal of Pediatric Orthopaedics B. 2018;27(5):387-90.
4. Zargarbashi RH, Bonaki HN, Zadegan SA, Baghdadi T, Nabian MH, Shirazi MR. Comparison of pediatric and general orthopedic surgeons' approaches in management of developmental dysplasia of the hip and flexible flatfoot: the road to clinical consensus. Archives of Bone and Joint Surgery. 2017;5(1):46.
5. De Pellegrin M, Moharamzadeh D, Fraschini G. Early diagnosis and treatment of DDH: a sonographic approach. Hip International. 2007;17(5_suppl):15-21.
6. Hansson G, Jacobsen S. Ultrasonography screening for developmental dysplasia of the hip joint. Acta Pædiatrica. 1997;86(9):913-5.
7. Alamdaran SA, Kazemi S, Parsa A, Moghadam MH, Feyzi A, Mardani R. Assessment of diagnostic value of single view dynamic technique in diagnosis of developmental dysplasia of hip: a comparison with static and dynamic ultrasond techniques. Archives of Bone and Joint Surgery. 2016;4(4):371.
8. Thaler M, Biedermann R, Lair J, Krismer M, Landauer F. Cost-effectiveness of universal ultrasound screening compared with clinical examination alone in the diagnosis and treatment of neonatal hip dysplasia in Austria. The Journal of bone and joint surgery British volume. 2011;93(8):1126-30.
9. Graf R. The diagnosis of congenital hip-joint dislocation by the ultrasonic Combound treatment. Archives of orthopaedic and traumatic surgery. 1980;97(2):117-33.
10. Woolacott NF, Puhan MA, Steurer J, Kleijnen J. Ultrasonography in screening for developmental dysplasia of the hip in newborns: systematic review. bmj. 2005;330(7505):1413.
11. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Systematic reviews. 2015;4(1):1.
12. Harbord RM, Whiting P. Metandi: meta-analysis of diagnostic accuracy using hierarchical logistic regression. The Stata Journal. 2009;9(2):211-29.
13. Kim KW, Lee J, Choi SH, Huh J, Park SH. Systematic review and meta-analysis of studies evaluating diagnostic test accuracy: a practical review for clinical researchers-part I. General guidance and tips. Korean journal of radiology. 2015;16(6):1175-87.
14. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. Journal of clinical epidemiology. 2005;58(9):882-93.
15. Doebler P, Holling H. Meta-analysis of diagnostic accuracy with mada. R Packag. 2015;1:15.
16. Dwamena B, Sylvester R, Carlos R. midas: Meta-analysis of diagnostic accuracy studies. Available at:)(Accessed on February 8, 2017) http://fmwww bc edu/repec/bocode/m/midas pdf View in Article. 2009.
17. Gokharman FD, Aydin S, Fatihoglu E, Ergun E, Kosar PN. Optimizing the time for developmental dysplasia of the hip screening: earlier or later? Ultrasound Quarterly. 2019;35(2):130-5.
18. Pillai A, Joseph J, McAuley A, Bramley D. Diagnostic accuracy of static graf technique of ultrasound evaluation of infant hips for developmental dysplasia. Archives of orthopaedic and trauma surgery. 2011;

131(1):53-8.

19. Rosendahl K, Markestad T, Lie R. Developmental dysplasia of the hip. A population-based comparison of ultrasound and clinical findings. Acta Paediatrica. 1996;85(1):64-9.

20. Tan SHS, Wu CH, Wong KL, Hui JH. Correlations between ultrasonographic and subsequent radiographic findings of developmental dysplasia of the hips. Ultrasonography. 2020;39(1):43.

21. Roovers E, Boere-Boonekamp MM, Castelein R, Zielhuis G, Kerkhoff T. Effectiveness of ultrasound screening for developmental dysplasia of the hip. Archives of Disease in Childhood-Fetal and Neonatal Edition. 2005;90(1):F25-F30.

22. Kishore Kumar R, Shah P, AN R, Rajan R. Diagnosing Developmental Dysplasia of Hip in Newborns Using Clinical Screen and Ultrasound of Hips—An Indian Experience. Journal of tropical pediatrics. 2016;62(3):241-5.

23. Lee J, Kim KW, Choi SH, Huh J, Park SH. Systematic review and meta-analysis of studies evaluating diagnostic test accuracy: a practical review for clinical researchers-part II. Statistical methods of meta-analysis. Korean journal of radiology. 2015; 16(6):1188-96.

24. McGrath TA, Alabousi M, Skidmore B, Korevaar DA, Bossuyt PM, Moher D, et al. Recommendations for reporting of systematic reviews and meta-analyses of diagnostic test accuracy: a systematic review. Systematic reviews. 2017;6(1):194.

25. Roposch A, Moreau NM, Uleryk E, Doria AS. Developmental dysplasia of the hip: quality of reporting of diagnostic accuracy for US. Radiology. 2006;241(3):854-60.

26. Pedrotti L, Crivellari I, Degrate A, De Rosa F, Ruggiero F, Mosconi M. Interpreting neonatal hip sonography: intraobserver and interobserver variability. Journal of Pediatric Orthopaedics B. 2020;29(3):214-8.

27. Clegg J, Bache C, Raut V. Financial justification for routine ultrasound screening of the neonatal hip. The Journal of Bone and Joint Surgery British volume. 1999;81(5):852-7.

28. Holen K, Tegnander A, Bredland T, Johansen O, Saether O, Eik-Nes S, et al. Universal or selective screening of the neonatal hip using ultrasound? A prospective, randomised trial of 15 529 newborn infants. The Journal of bone and joint surgery British volume. 2002;84(6):886-90.

29. Buonsenso D, Curatola A, Lazzareschi I, Panza G, Morello R, Marrocco R, et al. Developmental dysplasia of the hip: real world data from a retrospective analysis to evaluate the effectiveness of universal screening. Journal of ultrasound. 2020.

30. Laborie LB, Markestad TJ, Davidsen H, Brurås KR, Aukland SM, Bjørlykke JA, et al. Selective ultrasound screening for developmental hip dysplasia: effect on management and late detected cases. A prospective survey during 1991–2006. Pediatric radiology. 2014; 44(4):410-24.

31. Rosendahl K, Markestad T, Lie RT. Ultrasound screening for developmental dysplasia of the hip in the neonate: the effect on treatment rate and prevalence of late cases. Pediatrics. 1994;94(1):47-52.

32. Gulati V, Eseonu K, Sayani J, Ismail N, Uzoigwe C, Choudhury MZ, et al. Developmental dysplasia of the hip in the newborn: A systematic review. World journal of orthopedics. 2013;4(2):32.

33. Ortiz-Neira CL, Paolucci EO, Donnon T. A meta-analysis of common risk factors associated with the diagnosis of developmental dysplasia of the hip in newborns. European journal of radiology. 2012;81(3):e344-e51.

34. Quader N, Schaeffer EK, Hodgson AJ, Abugharbieh R, Mulpuri K. A systematic review and meta-analysis on the reproducibility of ultrasound-based metrics for assessing developmental dysplasia of the hip. Journal of Pediatric Orthopaedics. 2018;38(6):e305-e11.

35. Mostofi E, Chahal B, Zonoobi D, Hareendranathan A, Roshandeh KP, Dulai SK, et al. Reliability of 2D and 3D ultrasound for infant hip dysplasia in the hands of novice users. European Radiology. 2019;29(3):1489-95.