SYSTEMATIC REVIEW

Diagnostic Performance of Clinical Examination Versus Ultrasonography in the Detection of Developmental Dysplasia of Hip: A Systematic Review and Meta-Analysis

Mohammadreza Chavoshi, MD[1]; Ghazaleh Soltani, MD[2]; Shekoufe Shafiei Zargar, MD[3]; Cody Clayton Wyles, MD[4]; Hilal Maradit Kremers, MD[5]; Pouria Rouzrokh, MD[6]

Research performed at Mayo Clinic, Rochester, MN, USA

Received: 30 June 2021 *Accepted: 11 November 2021*

Abstract

Background: Developmental dysplasia of the hip (DDH) is a spectrum of diseases involving the femoroacetabular joint. Due to the controversies over the value of different strategies used for DDH screening, this systematic review and meta-analysis aimed to assess the diagnostic performance of standard physical examination maneuvers on the diagnosis of DDH, compared to the Graf ultrasonography (US) method.

Methods: PubMed, Web of Science, and SCOPUS databases were searched until the end of October 2020. Studies that (i) used the Ortolani test, Barlow test, or limited hip abduction (LHA) test to assess the risk of DDH in physical examination, (ii)used the Graf US method to examine DDH in sonography, and (iii) provided adequate data to extract the diagnostic performance were included. Pooled sensitivity and specificity were calculated for clinical examinations.

Results: A total of 25 studies (72,079 patients in total) were considered eligible to enter the present study. The pooled data of the Ortolani-Barlow test demonstrated a sensitivity of 36% (95% CI:0.25-0.48) and specificity of 98% (95% CI:0.93-0.99). Calculated pooled sensitivity and specificity for the limited hip abduction exam were obtained at 45% (95% CI:0.24-0.69) and 78% (95% CI:0.62-0.88) respectively. A separate analysis of the studies using both exams revealed a sensitivity of 57% (95% CI:0.30-0.82) and a specificity of 95% (95% CI:0.68-0.99).

Conclusion: Based on the results, the investigated clinical examinations have high specificity but low sensitivity to detect the DDH; therefore, they have limited application as a screening test. If obliged to rely on clinical examinations for screening, the combination of Ortolani-Barlow and LHA tests can provide more sensitivity than either of these tests performed independently.

Level of evidence: III

Keywords: Barlow, Developmental dysplasia of hip, Limited hip abduction, Ortolani, Ultrasonography

Introduction

Developmental dysplasia of the hip (DDH) is a spectrum of diseases involving the femoroacetabular joint. The pathology can be in the acetabulum, proximal femur, or their relation to each other (1, 2). The term "developmental" focuses on the nature of the disease as a dynamic pathology since birth. Disease prevalence and incidence are quite variable based on the disease definition, screening, and diagnostic methods, as well as geographic factors(3). The DDH treatment and subsequent disabilities can impose direct and indirect health costs on families and lead to a persistently painful hip for the child. Furthermore, there is an ongoing debate over the value of disease screening and diagnostic methods.

The DDH screening can be performed via a universal or a selective ultrasonographic (US) approach, in addition

Corresponding Author: Pouria Rouzrokh, Mayo Clinic, 200 First Street SW, Rochester, MN, USA
Email: rouzrokh.pouria@mayo.edu

to standard physical examination. Despite disagreements over the cost and hazards of overdiagnosis with universal US screening, a recent interdisciplinary expert panel recommended universal US screening in the early weeks by the Graf method, with high sensitivity and specificity and no radiation hazards (4). A recent meta-analysis has reported that the sensitivity and specificity of the Graf US method were 93% and 97%, respectively (5).

The clinical examination of the newborns is considered a primary, low-cost, and readily available screening method for DDH in almost all areas. Although it is routinely performed, the accuracy of the physical examination has been questioned. Different clinical examinations have been proposed to diagnose DDH; nonetheless, the two most widely used methods are Ortolani-Barlow and limited hip abduction (LHA) tests (6, 7). The present systematic review and meta-analysis aimed to assess the performance of these clinical examinations in the diagnosis of DDH while considering the Graf Us method as the reference test.

Materials and Methods

This study was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols (PRISMA-P) checklist (8).

Search Strategy

A systematic search was carried out on PubMed, Web of Science (Science and Social Science Citation Index), and SCOPUS databases until the end of March 2021. English articles which matched the following query in their titles, abstracts, or keywords were considered for our study: (pediatric* OR child* OR neonate* OR newborn* OR infant*) AND (ultrasound* OR sonograph*) AND ("physical exam*" OR "clinical exam*" OR Ortolani OR Barlow). We also looked at the references of our included articles (both primary and secondary studies) to ensure the maximum sensitivity of our search.

Study Selection

All original articles which met the following criteria were included in our final pool of studies

Use of Ortolani and Barlow or LHA test to assess the risk of DDH in physical examination.

Use of Graf method to examine DDH in sonography

Provision of adequate data to extract sensitivity and specificity of physical examinations

On the other hand, the exclusion criteria were as follows: abstracts with no full-text article available, unpublished studies, notes, letters, conference articles, duplicated studies, studies which used physical examinations other than the aforementioned ones, and studies which used sonography methods other than Graf. The titles and abstracts were first reviewed by two authors to check the eligibility of the study (GS, SS), and if both agreed, the full text of the studies were reviewed by two (and in case of discrepancy, three) of the authors. Studies that obtained the agreement of at least two reviewers were considered eligible for the meta-analyses.

Data Extraction

Two researchers independently extracted the required information from the included studies. Discrepancies were first discussed between the reviewers, and if needed, were referred to a third reviewer for a final decision. There is controversy over which cutoff should be used to categorize pathologic hips according to the Graf method. Since most studies have considered type IIb as the threshold, we used this cut-off in analysis to reduce the threshold effect. However, in some studies, it was not possible to extract the data based on this threshold; therefore, we used their own assigned thresholds, which were IIa, IIc, and III.

To assess the quality of the studies, the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool was utilized in Review Manager (version 5.2). For the performance of the meta-analysis, the studies were assigned to three distinct groups based on their reported method of physical examination: studies which used the Ortolani test, Barlow test, or both, studies which used LHA, and studies which used both tests (1-3). Pooled performance measures were then reported for each of these groups independently.

Statistical analysis

The hierarchical method was used to pool the diagnostic performance measures of the random-effect model, including sensitivity, specificity, and diagnostic odds ratio (DOR), for clinical examination and its subgroups. A bivariate model was then run to find the summary points for sensitivity and specificity, as well as their 95% confidence intervals (CI), considering within and between-study heterogeneity. The hierarchical summary receiver-operating characteristic (HSROC) curve and the area under the curve (AUC) were plotted using the HSROC model. Furthermore, 95% confidence and prediction regions were calculated to show the uncertainty degree of summarized sensitivity and specificity (9, 10). The summary receiver operating characteristic (ROC) plot was visually evaluated to explore the heterogeneity of studies.

The DOR, a routinely used single parameter of diagnostic accuracy in diagnostic test studies, was used to evaluate the heterogeneity of studies by calculating Higgins' I^2 statistics and Cochran's Q test. In addition, Cochran's Q test was calculated for pooled sensitivity and specificity. The linear correlation between sensitivity and the false-positive ratio was finally calculated to demonstrate the threshold effect, regarding $r \geq 0.6$ as significant (11). Finally, the Deeks' test was used to assess study publication bias, considering $P<0.10$ significant (12). All analyses were conducted by the "mada" package in R statistical analysis software (version 4.0.2, R Foundation for Statistical Computing, Vienna, Austria)(13).

Results

Study characteristics

A total of 2,151 studies were retrieved by searching the databases, out of which 589 articles were duplicated. The titles and abstracts of 1,562 studies were reviewed, and 75 articles were selected for full-text review. From this pool, 25 studies (72,079 patients in total) were considered eligible to enter our study (6, 14-37). A bibliography search was conducted on these 25 articles;

nonetheless, no further study was added. The detailed process of the study selection is displayed in Figure 1. The reason for the exclusion of the removed studies is explained in Supplement 1.

Among the 25 included studies, 16 provided data for the Ortolani-Barlow maneuver, 5 evaluated the LHA test, and the remaining four reported statistics for both tests. The detailed characteristics of the included studies are described in Table 1. The crude data and the calculated performance for each study are also presented in Table 2. The forest plot of the included studies is depicted in Figure 2.

The quality assessment of the studies according to the QUADAS-2 checklist is reported in Supplement 2, and the description of the QUADAS-2 critical appraisal checklist is provided in Supplement 3. Referral bias is a major concern in retrospective diagnostic test studies. Although the reference standard test for this study (US) is an operator-dependent procedure that needs inter and intra-observer variability assessment, studies with only one operator were regarded as low risk. Furthermore, the optimal time interval between clinical examination and US test is not definitely known, given the point that the more important issue here is that patient did not receive any treatment between the two tests; therefore, we considered the intervals of less than 12 weeks as acceptable.

Diagnostic performance

The pooled sensitivity and specificity of the studies for clinical examination were obtained at 40% (95% CI: 0.31-0.51) and 96% (95% CI:0.90-0.98), respectively. The hierarchical summary ROC curve for the clinical examination demonstrates an AUC of 0.71 (95% CI: 0.92-0.96). The pooled DOR was 16.45 (95% CI: 9.7-28.9) [Figure 3].

The pooled data of the Ortolani-Barlow test was calculated from 16 studies, demonstrating a sensitivity of 36% (95% CI:0.25-0.48) and specificity of 98% (95% CI:0.93-0.99). The LHA was analyzed in five studies, demonstrating a sensitivity and specificity of 45% (95% CI:0.24-0.69) and 78% (95% CI:0.62-0.88), respectively. The results of these two exams are compared in Figure 4. A separate analysis of the studies using both exams disclosed a sensitivity of 57% (95% CI:0.30-0.82) and specificity of 95% (95% CI:0.68-0.99). Nevertheless, since the number of studies in this group was low (n=4), a wide confidence interval was obtained.

Heterogeneity assessment

In general, the Cochrane's Q test and Higgins' I2

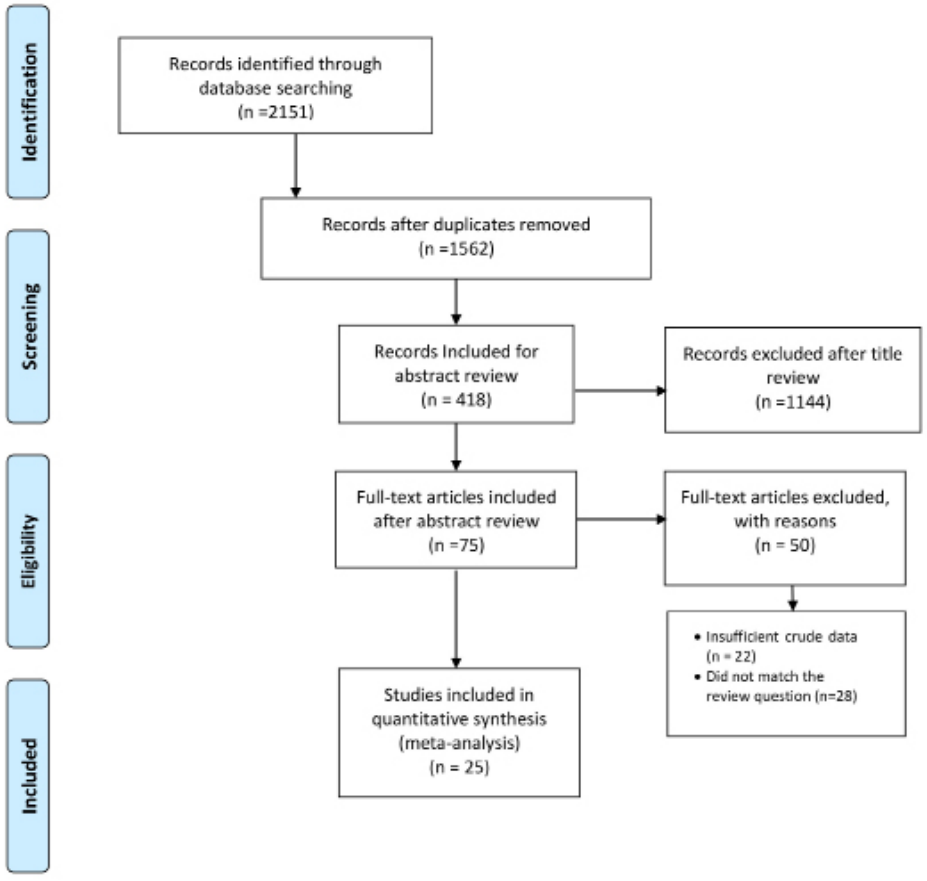

Figure 1. Flowchart of the study selection process for the systematic review.

Table 1. Characteristics of the included studies

Author, year	Patients	Population	US Age	Exam Age	Clinical Exam	Cut-off	US by:	Clinical Exam by:
Rosendahl, 1992	1503	All girls, only boys with identifiable risk factors	24-48 hours	24-48 hours	Ortolani-Barlow	2b	Radiologist	Pediatrician
Finnbogason, 1997	21	Abnormal clinical screen ± risk factors	Mean: 54 days	Mean: 30 days	Ortolani-Barlow	2c	Radiologist	Orthopedic Surgeon
Baronciani, 1997	7,082	All infants	4-6 w	1 week	Ortolani-Barlow	2b	27 months Experienced Doctor	Trained Doctors
Rosenberg, 1998	9199	All infants weighing above 1000g	24 hours	24 hours	Ortolani-Barlow.	2c	Pediatric Orthopedic Surgeon	Neonatologist/ Orthopedic Surgeon
Falliner, 1999	13096	All infants	1-4 days	1-4 days	Ortolani-Barlow	2c	Senior Orthopedist	Senior Orthopedist
Zenios, 2000	177	Suspicious instability + Risk factor	Within 6 weeks	within 6 weeks	Ortolani-Barlow,LHA	III	Radiologist	Orthopedic Surgeon
Castelein, 2001	683	Referred to the Orthopedic Department	mean: 173 days	90-310 days	LHA	2	Orthopedist	Orthopedist
Omeroglu, 2001	188	Young infants	3.7 months	3.7 months	LHA and Ortolani-Barlow	2a	Pediatric Radiologist	Pediatric Orthopedist
Jari, 2002	1107	Clinically unstable + Risk factors	1-9 weeks	N. A	LHA	2	Orthopedist	Orthopedist
Riboni, 2003	231	All infants	6-8 weeks	6-8 weeks	Ortolani-Barlow	2b	Radiologist	Orthopedist/ Pediatrician
Şenaran, 2004	464	Referred with LHA	within 4 months	within 4 months	LHA	2a	Orthopedic surgeon	Orthopedic surgeon
Finnbogason, 2008	1072	Risk factors + Clinically unstable/ suspicious	mean: 12.2 days	mean: 12.2 days	Ortolani, Barlow	2b	Pediatric Radiologist/ Trained Sonographer	Pediatric Orthopedist
Dogruel, 2008	7082	All infants	4-6 weeks	4-6 weeks	Ortolani, Barlow	2b	NA	Orthopedic Surgeon
Akgün, 2008	443	All infants	< 6 months	< 6 months	Ortolani, Barlow	2a	Orthopedic Surgeon	Pediatrician
Chen, 2010	2666	All infants	1 week	1 week	Ortolani, Barlow	2c	Pediatrician	Pediatrician/ Orthopedist
Sulaiman, 2011	30	Full term babies with breech presentation	one week	at birth	Ortolani, Barlow	NA	Radiologist	Trained Examiner
Choudry, 2013	5752	Abnormal clinical screen risk factors	1-2 weeks 6-9 weeks	1-2 weeks	LHA	3	Orthopedist	Orthopedic surgeon
Arti, 2013	11402	With risk factors	6 weeks	1,6 weeks	Ortolani, Barlow	2b	Experienced Sonographer	Orthopedic surgeon
Kumar, 2016	736	Term newborns	< 6 weeks	36 - 48 hours	Ortolani-Barlow	2b	Radiologist	NA
Anderton, 2018	105	Referred with asymmetrical skin creases of either the inguinal, adductor or gluteal folds	NA	3 months	Ortolani, Barlow, LHA,	3	NA	Orthopedic surgeon
Čustović, 2018	450	Neonates born at 37 to 42 weeks of gestation	24 hours	24 hours	LHA	2a	Orthopedic surgeon	Orthopedic surgeon
Gyurkovits, 2019	3272	Neonates born at gestational age ⩾37weeks	3 days	24 hours	Ortolani, Barlow	2c	Orthopedic surgeon	Orthopedic surgeon
D'Alessandr, 2019	318	Breech-born infants	NA	at birth	Ortolani, Barlow	2a	NA	Medical students, residents and attending physicians
Buonsenso, 2020	4000	All infants	6 weeks	1-3 days	Ortolani, Barlow	2a (-)	Pediatrician	NA
Omeroglu, 2020	1000	Infants with bilateral or unilateral dysplastic hips	Mean: 96.5 days	Mean: 96.5 days	Ortolani, Barlow, LHA	2a (-)	Senior Orthopedic surgeon	Orthopedics resident

Table 2. Reported performance for clinical examinations of developmental dysplasia of the hip in each of the included studies

Author, year	FN	TN	FP	TP	Sensitivity	Specificity	Positive Likelihood ratio	Negative Likelihood ratio
Rosendahl, 1992	47	2882	44	33	0.414 (0.313-0.522)	0.985 (0.980 -0.989)	27.203 (18.415-40.187)	0.595 (0.496- 0.715)
Finnbogason, 1997	1	28	10	3	0.700 (0.299-0.927)	0.731 (0.576-0.845)	2.600 (1.201-5.629)	0.411 (0.106- 1.587)
Baronciani, 1997	228	7514	3	12	0.052 (0.030- 0.088)	1.000 (0.999- 1.000)	111.811 (34.415- 363.262)	0.949 (0.921- 0.977)
Rosenberg, 1998	69	18250	3	76	0.524 (0.443 -0.603)	1.000 (0.999- 1.000)	2732.742 (947.8240- 7878.974)	0.476 (0.402_ 0.564)
Falliner, 1999	88	11688	1268	52	0.372 (0.297- 0.455)	0.902 (0.897- 0.907)	3.803 (3.050-4.742)	0.696 (0.613- 0.790)
Zenios, 2000	19	106	14	38	0.664 (0.535- 0.772)	0.880 (0.810- 0.927)	5.539 (3.305-9.284)	0.382 (0.264- 0.552)
Castelein, 2001	70	247	210	156	0.689 (0.626 -0.746)	0.540 (0.495- 0.586)	1.500 (1.314-1.712)	0.575 (0.465- 0.710)
Omeroglu, 2001	11	296	37	32	0.739 (0.594- 0.845)	0.888 (0.849 -0.917)	6.579 (4.640-9.327)	0.294 (0.179- 0.485)
Jari, 2002	88	816	118	85	0.491 (0.418- 0.565)	0.873 (0.850- 0.893)	3.877 (3.092-4.861)	0.582 (0.502- 0.675)
Riboni, 2003	48	8829	11	8	0.149 (0.079- 0.263)	0.999 (0.998- 0.999)	114.643 (49.126-267.537)	0.852 (0.764- 0.950)
Şenaran, 2004	12	266	155	31	0.716 (0.570- 0.828)	0.632 (0.584- 0.676)	1.943 (1.553-2.431)	0.450 (0.280- 0.723)
Finnbogason, 2008	10	866	175	21	0.672 (0.499- 0.808)	0.832 (0.808- 0.853)	3.989 (3.023-5.263)	0.395 (0.240- 0.648)
Dogruel, 2008	149	6502	372	59	0.285 (0.228- 0.349)	0.946 (0.940- 0.951)	5.254 (4.148-6.656)	0.756 (0.694- 0.824)
Akgün, 2008	56	299	53	35	0.386 (0.293- 0.488)	0.848 (0.807- 0.882)	2.546 (1.782-3.638)	0.724 (0.612- 0.856)
Chen, 2010	3	2643	14	6	0.650 (0.354- 0.863)	0.995 (0.991- 0.997)	119.152 (60.015-236.561)	0.352 (0.151- 0.819)
Sulaiman, 2011	2	23	1	4	0.643 (0.303- 0.882)	0.940 (0.777- 0.986)	10.714 (2.064-55.615)	0.380 (0.140- 1.032)
Choudry, 2013	130	4638	958	26	0.169 (0.118- 0.235)	0.829 (0.819- 0.838)	0.986 (0.693-1.401)	1.003 (0.934- 1.077)
Arti, 2013	240	10468	600	94	0.282 (0.237- 0.333)	0.946 (0.941- 0.950)	5.200 (4.310- 6.273)	0.759 (0.710- 0.812)
Kumar, 2016	0	546	188	2	0.833 (0.310- 0.982)	0.744 (0.711- 0.774)	3.249 (1.930- 5.470)	0.224 (0.018- 2.816)
Anderton, 2018	0	84	19	2	0.833 (0.310- 0.982)	0.812 (0.727-0.876)	4.444 (2.332-8.472)	0.205 (0.016- 2.580)
Čustović, 2018	75	308	40	27	0.267 (0.191- 0.360)	0.884 (0.846-0.913)	2.301 (1.494-3.542)	0.829 (0.734- 0.937)
Gyurkovits, 2019	56	3149	53	14	0.204 (0.127- 0.312)	0.983 (0.978- 0.987)	12.227 (7.193-20.783)	0.809 (0.719- 0.911)
D'Alessandr, 2019	27	281	4	6	0.191 (0.093- 0.352)	0.984 (0.962- 0.994)	12.150 (3.854-38.302)	0.822 (0.697- 0.968)
Buonsenso, 2020	32	3856	96	16	0.337 (0.221- 0.477)	0.976 (0.970- 0.980)	13.794 (8.887-21.409)	0.680 (0.557- 0.830)
Omeroglu, 2020	197	999	1	49	0.200 (0.155- 0.255)	0.999 (0.994- 1.000)	133.737 (26.509-674.698)	0.801 (0.752- 0.852)

statistics are used for heterogeneity assessment in meta-analyses; however, their implication in diagnostic test meta-analysis is uncertain. The threshold effect is not accounted for in these tests; moreover, the calculation of heterogeneity is based on the single outcome variables (like odds ratio)(38, 39). An alternative way to assess heterogeneity is to watch for the degree of closeness of observed study results to the SROC (38). There is no consensus on a specific method for heterogeneity assessments in diagnostic test accuracy studies, therefore, a combination of all available methods should be used (40, 10).

According to the Cochran's Q test, both sensitivity and specificity demonstrated considerable heterogeneity between the studies (Cochrane Q test = 437.6644, df = 24; *P<2e-16* for sensitivity, and Cochrane Q test = 8431.217, df = 24;*P<2e-16* for specificity). Furthermore, to calculate the Higgin's I^2 and Cochran's Q test by a single outcome variable, the DOR was used and the results showed high heterogeneity (Cochrane Q test = 46.026, df = 24, *P=0.004*, I2 = 47.86%, tau^2: 1.51 CI (95%):0.3-4.82). The threshold effect analysis revealed that the proportion of the heterogeneity due to this factor was considerable (r=0.6). In addition, the scattered pattern of each study concerning the SROC illustrated a high heterogeneity between the studies.

The multi-metaregression analysis was also conducted on the subgroups to find other possible sources of

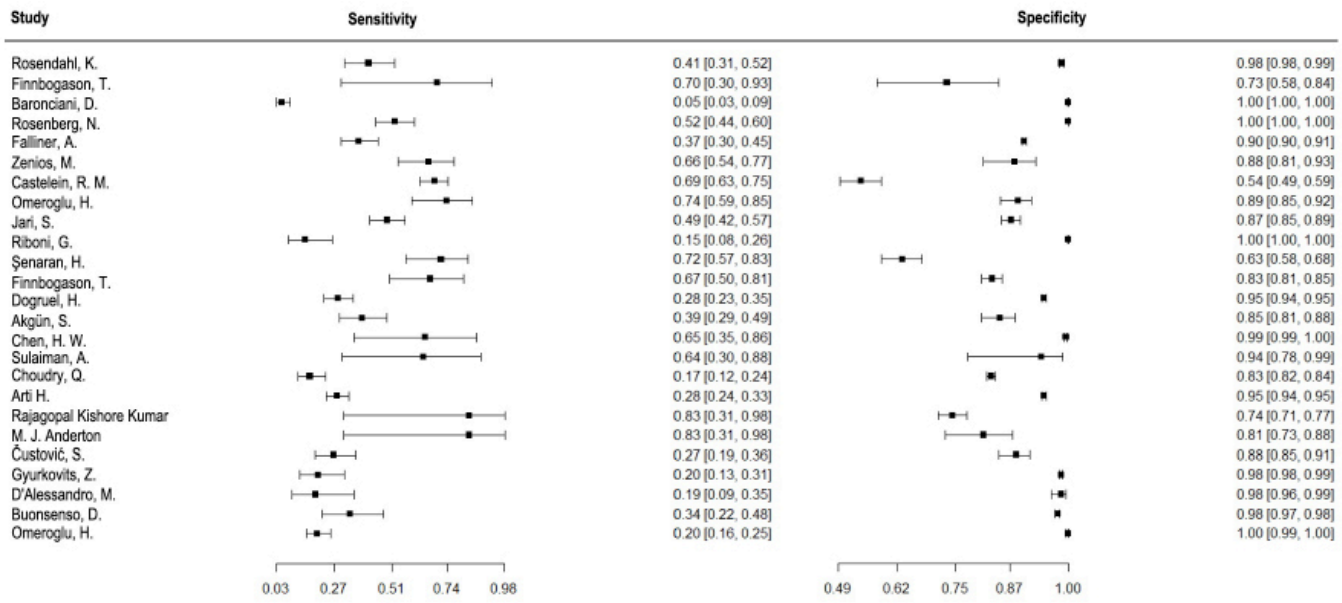

Figure 2. Forest plots of sensitivity and specificity of the studies. Horizontal lines represent 95% CIs of the individual studies.

heterogeneities. Since the minimum number of studies in each subgroup is recommended to be 10, only four factors could be analyzed: the operator of the US exam (radiologist vs. non-radiologist), the clinical examiner (orthopedic surgeons vs. others), examination week (1st week vs. after 1st week), and the population of the study (universal vs. selective screening) (1-4). The analysis demonstrated that only the clinical examiner was responsible for heterogeneity in specificity (*P=0.000*). Mild heterogeneity was also observed in specificity due to the population (*P=0.05*). None of the factors were responsible for heterogeneity in sensitivity (*P > 0.05*).

Publication bias

Deeks' funnels plot indicated a low publication bias with a *P-value* of 0.96 **[Supplement 4].**

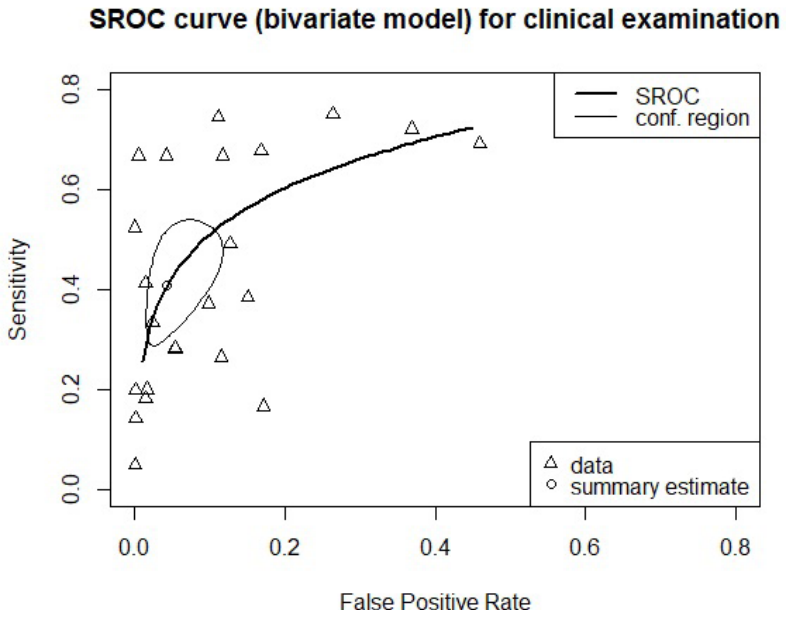

Figure 3. Hierarchical summary receiver-operating characteristic (HSROC) curve. The "data" points show performance for each study and the "summary estimate" points represent the pooled performance.

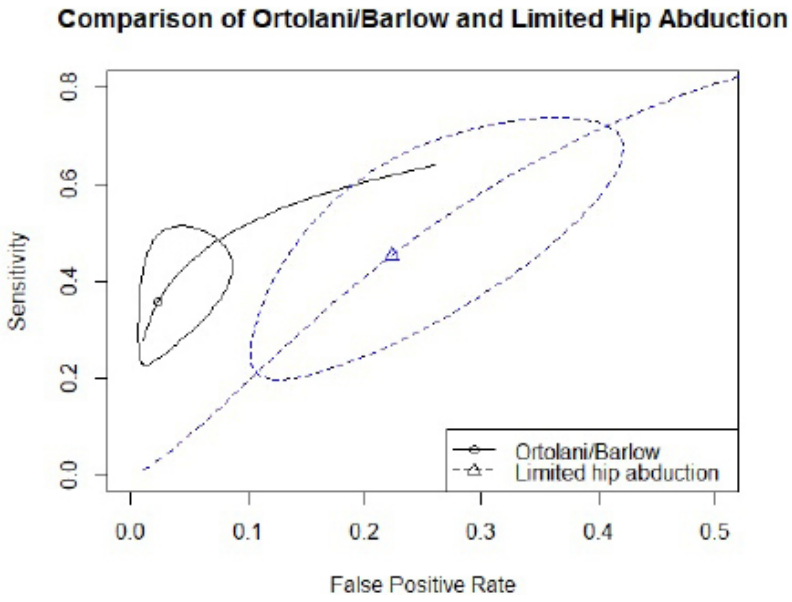

Figure 4. The comparison of the pooled data for Ortolani/Barlow and LHA.

Discussion

As evidenced by the results of this meta-analysis, clinical examination demonstrated high specificity but low sensitivity to detect the DDH; therefore, it has a limited application as a screening test. Indeed, this is the opposite of desired performance for ideal screening tests which should have high sensitivity, even if this compromises specificity. More advanced testing with high specificity can subsequently be relied upon for confirmatory testing. Our analysis showed that the combination of Ortolani-Barlow and LHA tests has the highest sensitivity, while the highest specificity pertains to the Ortolani-Barlow test. Moreover, it was found that the specialty of the examiner could influence the results of specificity.

Bilateral LHA could be a normal finding, especially in the first few weeks after birth; therefore, it is recommended to be performed after eight weeks (29). This could be the reason behind the lower specificity of the LHA test, in comparison with the Ortolani-Barlow test. However, a combination of both exams could increase the overall sensitivity. The screening method of DDH is a controversial issue in the countries. Various methods have been proposed, including clinical examination with or without ultrasonography (either universal or selective) and ultrasound exam alone(41). In addition, there are multiple methods for clinical examination and ultrasound (Graf, Terjesen, Harke). There is no consistent data about the value and effectiveness of each method in DDH patients. However, there is a strong agreement that clinical examination alone is suboptimal in DDH screening(4).

The DDH screening is of utmost importance since the disease could be treated less invasively and with better outcomes if diagnosed early. Although the Graph US method has gained a prominent place in DDH diagnosis, physical examination is still considered routine(42). In a decision-analytic model conducted by Mahan et al, screening all neonates for hip dysplasia with a physical examination followed by the selective US in high-risk patients (positive clinical exam, breech delivery, and positive family history) yielded the best strategy to prevent degenerative hips by the age of 60(43). However, the sole reliance on the clinical examination to identify high-risk patients results in missing cases given poor test sensitivity. Low sensitivity could be related to the inability of clinical exams in diagnosing the acetabular pathologies, while instability or dislocation of the femoral head is correctly diagnosed (4). Another reason could be the evolving nature of the disease. A normal physical exam early after birth does not exclude the possibility of DDH in older ages(44). Engesæter et al. reported that only 8% of patients who underwent total hip replacement had a history of hip instability at birth(45).

What is agreed upon is that the missed cases of DDH that go on to become hip dysplasia after the closure of the triradiate cartilage portend the worst prognosis for the hip along the spectrum of hip morphologies, which may require a periacetabular osteotomy to avoid total hip arthroplasty at a young age(46, 47). It should also be recognized that some clinicians exercise great caution in the serial performance of Ortolani-Barlow tests or their execution by multiple examiners. A great concern arises from the fact that these maneuvers place high stress on the acetabular cartilage when it is in its most plastic state, resulting in the risk of exacerbating borderline or true dysplasia or even developing it in a normal hip.

The present study has some limitations which must be addressed. Firstly, regarding the heterogeneity of the pooled studies, there was a considerable threshold effect due to the variation of defined cut-offs for labeling hips as pathologic in the Graf method. Some studies use type IIa as the cut-off for DDH, while others have used

types IIb or III as pathologic. Furthermore, the US is an operator-dependent method that could be affected by the experience of the operator. Although the multi-regression analysis revealed no heterogeneity due to the variation of operators, considering all the mentioned issues, the applicability of the reference test in the QUADAS-2 checklist was unclear for all studies. Secondly, the clinical examiners varied across our included studies, resulting in inter and intra-observer variations which affect their reported results. This was also confirmed by the multi-regression analysis, demonstrating that the examiner's experience could affect specificity in particular. This was the main reason why the applicability of the index test was assigned as high risk.

Furthermore, most of the included studies had assessed high-risk patients, and it is a potential source of population bias. Considering the purpose of this study which was the evaluation of the performance of clinical examinations in the general population, the studies that only focused on high-risk populations were assigned as unclear bias. Finally, there was no specified standard time interval between clinical exams and US exams to find the accuracy of the results.

As evidenced by the results of this study, it can be concluded that physical examination is not a suitable method for DDH screening due to its poor sensitivity, which is the most important characteristic of a screening test. Nonetheless, due to the high specificity, the treatment of positive cases should not be postponed until the Graph US approval. If obliged to rely on clinical examinations for screening, the combination of Ortolani-Barlow and LHA tests can provide more sensitivity than either of these tests performed independently. The totality of these results, along with other factors, such as cost, should be considered by practitioners and policymakers in determining best practices for DDH screening.

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

Mohammadreza Chavoshi MD[1]
Ghazaleh Soltani MD[2]
Shekoufe Shafiei Zargar MD[3]
Cody Clayton Wyles MD[4]
Hilal Maradit Kremers MD[5]
Pouria Rouzrokh MD[6]

1 Department of Radiology, Shariati Hospital, Tehran University of Medical Sciences, Tehran, Iran
2 Translational Ocular Research Center, Tehran University of Medical Sciences, Tehran, Iran
3 Student Scientific Research Center, Tehran University of Medical Sciences, Tehran, Iran
4 Department of Clinical Anatomy, Mayo Clinic, 200 First Street SW, Rochester, MN, 55905, USA
5 Departments of Health Science Research and rthopedics Surgery, Mayo Clinic, 200 First Street SW, Rochester, MN, 55905, USA
6 Department of Radiology, Radiology Informatics Laboratory, Mayo Clinic, 200 First Street SW, Rochester, MN, 55905, USA

References

1. Vaquero-Picado A, González-Morán G, Garay EG, Moraleda L. Developmental dysplasia of the hip: update of management. EFORT open reviews. 2019;4(9):548-56.
2. Schmitz MR, Murtha AS, Clohisy JC, Group AS. Developmental dysplasia of the hip in adolescents and young adults. JAAOS-Journal of the American Academy of Orthopaedic Surgeons. 2020;28(3):91-101.
3. Lehmann HP, Hinton R, Morello P, Santoli J. Developmental dysplasia of the hip practice guideline: technical report. Pediatrics. 2000;105(4):e57-e.
4. O'Beirne JG, Chlapoutakis K, Alshryda S, Aydingoz U, Baumann T, Casini C, et al. International interdisciplinary consensus meeting on the evaluation of developmental dysplasia of the hip. Ultraschall in der Medizin-European Journal of Ultrasound. 2019;40(04):454-64.
5. Chavoshi M, Mirshahvalad SA, Zamani F, Mahdizadeh M. Diagnostic Accuracy of Graf's Method Ultrasonography in Detecting Developmental Dysplasia of the Hip: A Meta-Analysis and Systematic Review. The Archives of Bone and Joint Surgery. 2021.
6. Jari S, Paton R, Srinivasan M. Unilateral limitation of abduction of the hip: A VALUABLE CLINICAL SIGN FOR DDH? The Journal of bone and joint surgery British volume. 2002;84(1):104-7.
7. Lipton GE, Guille JT, Altiok H, Bowen JR, Harcke HT. A reappraisal of the Ortolani examination in children with developmental dysplasia of the hip. Journal of Pediatric Orthopaedics. 2007;27(1):27-31.
8. 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.
9. Harbord RM, Whiting P. Metandi: meta-analysis of diagnostic accuracy using hierarchical logistic regression. The Stata Journal. 2009;9(2):211-29.
10. 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.
11. Devillé WL, Buntinx F, Bouter LM, Montori VM, De Vet HC, Van der Windt DA, et al. Conducting systematic

reviews of diagnostic studies: didactic guidelines. BMC medical research methodology. 2002;2(1):9.
12. 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.
13. Doebler P, Holling H. Meta-analysis of diagnostic accuracy with mada. R Packag. 2015;1:15.
14. Rosendahl K, Markestad T, Lie R. Congenital dislocation of the hip: a prospective study comparing ultrasound and clinical examination. Acta Paediatrica. 1992;81(2):177-81.
15. Finnbogason T, Jorulf H. Dynamic ultrasonography of the infant hip with suspected instability: A new technique. Acta Radiologica. 1997;38(2):206-9.
16. Baronciani D, Atti G, Andiloro F, Bartesaghi A, Gagliardi L, Passamonti C, et al. Screening for developmental dysplasia of the hip: from theory to practice. Pediatrics. 1997;99(2):e5-e.
17. Rosenberg N, Bialik V, Norman D, Blazer S. The importance of combined clinical and sonographic examination of instability of the neonatal hip. International orthopaedics. 1998;22(3):185-8.
18. Falliner A, Hahne HJ, Hassenpflug J. Sonographic hip screening and early management of developmental dysplasia of the hip. Journal of pediatric orthopedics Part B. 1999;8(2):112-7.
19. Zenios M, Wilson B, Galasko C. The effect of selective ultrasound screening on late presenting DDH. Journal of pediatric orthopedics Part B. 2000;9(4):244-7.
20. Castelein RM, Korte J. Limited hip abduction in the infant. Journal of Pediatric Orthopaedics. 2001;21(5):668-70.
21. Ömerog H, Koparal S. The role of clinical examination and risk factors in the diagnosis of developmental dysplasia of the hip: a prospective study in 188 referred young infants. Archives of orthopaedic and trauma surgery. 2001;121(1-2):7-11.
22. Riboni G, Bellini A, Serantoni S, Rognoni E, Bisanti L. Ultrasound screening for developmental dysplasia of the hip. Pediatric radiology. 2003;33(7):475-81.
23. Şenaran H, Özdemir HM, Ögün TC, Kapicioglu MS. Value of limited hip abduction in developmental dysplasia of the hip. Pediatrics international. 2004;46(4):456-8.
24. Finnbogason T, Jorulf H, Söderman E, Rehnberg L. Anterior dynamic ultrasound and Graf's examination in neonatal hip instability. Acta Radiologica. 2008;49(2):204-11.
25. Dogruel H, Atalar H, Yavuz O, Sayli U. Clinical examination versus ultrasonography in detecting developmental dysplasia of the hip. International orthopaedics. 2008;32(3):415-9.
26. Akgün S, Bakar C, Budakoğlu Iİ, Tuncay C, Cemil T. Is clinical examination reliable in diagnosis of developmental dysplasia of the hip? Gazi Medical Journal. 2008;19(2).
27. Chen H-W, Chang C-H, Tsai S-T, Liu W-J, Chua C, Chen Y-Y, et al. Natural progression of hip dysplasia in newborns: a reflection of hip ultrasonographic screenings in newborn nurseries. Journal of Pediatric Orthopaedics B. 2010;19(5):418-23.
28. Sulaiman A, Yusof Z, Munajat I, Lee N, Zaki N. Developmental dysplasia of hip screening using Ortolani and Barlow testing on breech delivered neonates. Malaysian Orthopaedic Journal. 2011;5(3):13.
29. Choudry Q, Goyal R, Paton R. Is limitation of hip abduction a useful clinical sign in the diagnosis of developmental dysplasia of the hip? Archives of disease in childhood. 2013;98(11):862-6.
30. Arti H, Mehdinasab SA, Arti S. Comparing results of clinical versus ultrasonographic examination in developmental dysplasia of hip. Journal of research in medical sciences: the official journal of Isfahan University of Medical Sciences. 2013;18(12):1051.
31. 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.
32. Anderton M, Hastie G, Paton RW. The positive predictive value of asymmetrical skin creases in the diagnosis of pathological developmental dysplasia of the hip. Bone Joint J. 2018;100(5):675-9.
33. Čustović S, Šadić S, Vujadinović A, Hrustić A, Jašarević M, Čustović A, et al. The predictive value of the clinical sign of limited hip abduction for developmental dysplasia of the hip (DDH). Med Glas (Zenica). 2018;15(2):174-8.
34. Gyurkovits Z, Sohár G, Baricsa A, Németh G, Orvos H, Dubs B. Early detection of developmental dysplasia of hip by ultrasound. HIP International. 2019:1120700019879687.
35. D'Alessandro M, Dow K. Investigating the need for routine ultrasound screening to detect developmental dysplasia of the hip in infants born with breech presentation. Paediatrics & child health. 2019;24(2):e88-e93.
36. Buonsenso D, Curatola A, Lazzareschi I, Panza G, Morello R, Marrocco R, Valentini P, Cota F, Rendeli C. Developmental dysplasia of the hip: real world data from a retrospective analysis to evaluate the effectiveness of universal screening. Journal of ultrasound. 2021;24(4):403-10.37.
37. Ömeroglu H, Tatlici E, Köse N. Significance of asymmetry of groin and thigh skin creases in developmental dysplasia of the hip revisited: results of a comparative study. Journal of Pediatric Orthopaedics. 2020;40(8):e761-e5.
38. 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.
39. Bossuyt P, Davenport C, Deeks J, Hyde C, Leeflang M, Scholten R. Chapter 11: Interpreting results and drawing conclusions. Cochrane handbook for systematic reviews of diagnostic test accuracy version 09 The Cochrane Collaboration. 2013:1-30.
40. 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.
41. Shorter D, Hong T, Osborn DA. Cochrane Review: Screening programmes for developmental dysplasia of the hip in newborn infants. Evidence-Based Child Health: A Cochrane Review Journal. 2013;8(1):11-54.
42. Shipman SA, Helfand M, Moyer VA, Yawn BP. Screening for developmental dysplasia of the hip: a systematic literature review for the US Preventive Services Task Force. Pediatrics. 2006;117(3):e557-e76.
43. Mahan ST, Katz JN, Kim Y-J. To screen or not to screen? A decision analysis of the utility of screening for developmental dysplasia of the hip. The Journal of Bone and Joint Surgery American volume. 2009;91(7):1705.
44. Vitale MG, Skaggs DL. Developmental dysplasia of the hip from six months to four years of age. JAAOS-Journal of the American Academy of Orthopaedic Surgeons. 2001;9(6):401-11.
45. Engesæter IØ, Lie SA, Lehmann TG, Furnes O, Vollset SE, Engesæter LB. Neonatal hip instability and risk of total hip replacement in young adulthood. Acta orthopaedica. 2008;79(3):321-6.
46. Wyles CC, Heidenreich MJ, Jeng J, Larson DR, Trousdale RT, Sierra RJ. The John Charnley Award: redefining the natural history of osteoarthritis in patients with hip dysplasia and impingement. Clinical Orthopaedics and Related Research®. 2017;475(2):336-50.
47. Wyles CC, Vargas JS, Heidenreich MJ, Mara KC, Peters CL, Clohisy JC, et al. Natural history of the dysplastic hip following modern periacetabular osteotomy. JBJS. 2019;101(10):932-8.