VALIDITY AND DIAGNOSTIC BIAS IN THE CLINICAL SCREENING FOR CONGENITAL DYSPLASIA OF THE HIP

C. JIMÉNEZ1, M. DELGADO-RODRÍGUEZ2,3,5, M. LÓPEZ-MORATALLA1, M. SILLERO4, R. GÁLVEZ-VARGAS2

To assess the validity of the clinical maneuvers and signs in the screening for congenital dysplasia of the hip and the presence of a diagnostic bias in this screening, a random sample of 261 newborns was studied at a tertiary hospital. All newborns were clinically and sonographically studied in the first 48 hours after birth. Hips were classified according to Graf’s scheme. Sensitivity, specificity and predictive values were estimated. Nineteen dysplasias and 100 physiologically immature hips were detected. Sensitivity and specificity for the Ortolani-Barlow maneuver were 26% and 84% respectively. A higher sensitivity was observed when all clinical maneuvers/signs were considered; however, specificity decreased, so the positive predictive value reached a similar figure to that of Ortolani-Barlow, 5%. A higher positive rate with the Ortolani-Barlow procedure was observed when newborns had risk factors for congenital disease of the hip (family history and female sex) or belonged to groups that could be considered as high-risk (low Apgar score, younger mothers). Other procedures apart from clinical examination should also be recommended as an effective screening procedure to decrease the number of cases missed and to avoid diagnostic bias in future epidemiologic research on this condition.

Keywords: clinical screening ; hip ; congenital dysplasia.
Mots-clés: dépistage clinique ; hanche ; dysplasie congénitale.

INTRODUCTION

After the contributions of Ortolani, Palmén, von Rosen, and Barlow, neonatal screening for congenital dislocation of the hip became widespread in many countries in the 1960s (10). Congenital dysplasia is less severe than congenital dislocation, but it may not be a negligible condition. Several authors have called attention to the fact that many adult cases presumed to be idiopathic osteoarthritis of the hip are actually conditions secondary to preexisting deformities, especially adult hip dysplasia (1, 33). It is not established that neonatal dysplasia leads to the adult dysplastic hip, although it might be a previous stage in the natural course of the disease. Thus neonatal screening should also be aimed at discerning less advanced stages of congenital dislocation of the hip.

The efficacy of neonatal screening for the disorder has been seriously questioned because of the high percentage of false negatives (20). Although neonatal screening is widely applied, its validity has not been assessed very often, especially regarding congenital dysplasia of the hip. The following study was undertaken at the University of Granada Departments of Preventive Medicine and Orthopedics and Surgery as part of the

1 Department of Surgery and Orthopedics.
2 Department of Preventive Medicine, School of Medicine.
3 Unit of Clinical Epidemiology, General Hospital «Virgen de la Nieves», Granada, Spain.
4 Provincial Office for Health, Jaén,
5 Present Address : Division of Preventive Medicine and Public Health, School of Medicine, University of Cantabria, Santander, Spain.

Correspondence : M. Delgado-Rodriguez.
normal clinical service, primarily to determine the effectiveness of clinical screening, namely the sensitivity, specificity and predictive values of the clinical neonatal screening for congenital dysplasia of the hip (CDH). Secondly, our goal was to ascertain the presence of diagnostic bias in the clinical screening (i.e. a higher rate of CDH diagnosis when a risk factor for CDH is present). A risk factor for CDH can favour CDH diagnosis in several ways (e.g. by increasing the awareness of a physician about the disease, this can certainly improve the performance of a clinical exam). Suppose that at a clinic, pediatricians, aware of the epidemiology of CDH, perform a much more thorough clinical exam of newborns in those with relatives affected (a known risk factor for the disease). A future analysis of these cases will show that family history is a stronger risk factor for the disease than it truly is. In this situation one must have in mind that family history favors diagnosis and its presence in CDH is inflated. This leads to an overestimation of a risk factor.

METHODS

This study was based at the University of Granada Hospital, a tertiary 800-bed hospital which is the referral centre for a population of 450,000. Data were gathered for the period January 1990 to February 1991.

A random sample of 261 newborns was drawn from all deliveries at our hospital. These newborns were examined within 24 hours of birth following the most recent guidelines for the detection of congenital dislocation of the hip given by the Standing Medical Advisory Committee and the Standing Nursing and Midwifery Advisory Committee Working Party (28). The baby was in the supine position with the feet toward the examiner. The Ortolani-Barlow maneuver was performed by pediatricians shortly after birth (first 24 hours). The remaining maneuvers/signs recommended (leg posture, limb shortening, asymmetry of thighs, flattening of the buttocks, limitation of abduction, hip instability) were not routinely investigated by pediatricians. These maneuvers were completed by one of us (CJ).

Newborns were also examined in the first 48 hours of life by the Graf technique of ultrasound (US) (12-15). A Thompson’s linear scanner (model Logic-7) with a 7.5-MHz transducer was used. All infants had a check-up after 3 months and those who required further observation or therapy were followed up at 6-week intervals. After US and a follow-up of 6 months of those who showed US abnormalities, 12 out of the initial sample of 261 babies were classified as having CDH and 53 as having maturity deficit, thus leaving 196 controls.

Congenital dysplasia of the hip was diagnosed following the criteria developed by Graf (12, 15): a bony roof angle (α) of 50-59° and a cartilage roof angle (β) of ≥ 55° after 3 months from birth (hips 2b), or an α of 43-50° and a β of 70-77° from birth (hips 2c). Twelve infants (with 19 hips affected) were diagnosed. This criterion should also be accomplished at 6 months of age.

Physiologically immature hips (hips 2a-) were diagnosed only when the angle α yielded a figure of 50-59° and β was ≥ 55° during the first 3 months after birth. A baby who kept the same angle figures after 6 months of life was considered to be in the dysplasia group. Fifty-three babies (with 100 hips affected) were included in this category.

The ultrasound studies were performed without knowledge of the preceding clinical findings nor of any data concerning risk factors for CDH. These observers were blind as regards to any data which could influence interpretation of the test.

Clinical findings were compared to US (reference or gold standard). The unit of analysis was the hip (522 hips were studied). Sensitivity (proportion of true positives in the pathologic hips), specificity (proportion of true negatives in the healthy hips), positive predictive value (proportion of true positives in all positive values) negative predictive value (proportion of true negatives in all negative values), and accuracy (proportion of true, either positive or negative, values) were estimated (25).

Two general analyses were performed. First, the validity of the Ortolani-Barlow maneuver was assessed. Secondly, all the clinical signs recommended by the Standing Medical Advisory Committee were assessed together as a parallel test: any positive result in any maneuver/sign would be considered as a positive result at screening. On the contrary, screening would be considered negative only when all tests were negative.

To assess the presence of diagnostic bias, the sensitivity of the Ortolani-Barlow and the proportion of false positives in normal hips were stratified according to several variables: known risk factors for CDH and indicators of high-risk in newborns. This was not performed for the remaining maneuvers because they are not routinely (in all newborns) done at our centre and, in addition, they were carried out by one of our
team without him knowing any characteristics of the mother or of the newborn.

The exact binomial 95% confidence intervals were estimated when a sample was less than 200. The statistical tests applied were the Fisher’s exact test and $\chi^2$.

RESULTS

Five positive Ortolani-Barlow tests were found in 19 hips with CDH (sensitivity = 26.3%), while 348 were negative in the 413 normal hips (specificity = 84.3%). With these figures, it easy to estimate predictive values and accuracy. All these parameters are summarized in table I. The sensitivity of the Ortolani-Barlow maneuver is low: it only identifies about one quarter of all cases (almost 75% of all cases are missed). When all clinical procedures were combined (sensitivity = 63.2%) 12 positives were obtained in hips affected and 242 negatives in normal hips (specificity = 58.4%). With all procedures sensitivity increases, but specificity decreases (as expected), thus yielding a positive predictive value similar to that of Ortolani-Barlow. The positive predictive value indicates the ability of the test (either Ortolani-Barlow only or all procedures) to correctly identify a presumed CDH. A figure of about 5% means that 19 out of 20 positive clinical maneuvers are false positives.

On the contrary, the negative predictive value denotes the ability of the clinical exam to correctly confirm the absence of CDH after a negative clinical exam. Figures higher than 95% imply that after a negative clinical exam the physician can feel reasonably sure that CDH does not exist.

However, the negative predictive value must be interpreted with respect to the prevalence of the condition. If the prevalence is low (as it is for CDH), a closer-to-100% figure should be desired.

In table II the positive rate of Ortolani-Barlow for CDH, physiologically immature and normal hips is stratified by several factors. The main points to be highlighted are the similarity between the positive rates for CDH and maturity deficit and the trend for obtaining more positive results in females (non-significant), in newborns with a positive family history (two well-known risk factors for CHD), in newborns from young mothers and in newborns with lower Apgar scores (two indicators of high risk for other conditions). This trend was not observed for breech presentation, side of hip, and premature rupture of membranes. In premature newborns a non-significant lower positive rate was appreciated.

DISCUSSION

The examination of neonates for congenital dislocation of the hip has become routine in health care. However, the rate of conditions diagnosed after the newborn period is quite similar to that reported before the introduction of screening. Neonatal screening has not solved the problem of congenital disease of the hip; many cases are still missed (17, 19, 21). This suggests two possibilities: a) sensitivity of clinical screening is far from an acceptable figure; b) two distinct entities exist, neonatal (detectable at or shortly after birth) and postneonatal, with epidemiological differences (4).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Ortolani-Barlow</th>
<th>All procedures*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Percent 95% CI</td>
<td>Percent 95% CI</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>26.3 (9.2-51.2)</td>
<td>63.2 (38.4-83.7)</td>
</tr>
<tr>
<td>Specificity</td>
<td>84.3 (80.7-87.8)</td>
<td>58.4 (53.6-63.1)</td>
</tr>
<tr>
<td>Positive Predictive Value</td>
<td>5.9 (1.9-13.2)</td>
<td>5.4 (2.6-8.2)</td>
</tr>
<tr>
<td>Negative Predictive Value</td>
<td>96.6 (95.1-98.5)</td>
<td>97.7 (95.9-99.5)</td>
</tr>
<tr>
<td>Accuracy</td>
<td>82.2 (78.9-85.4)</td>
<td>58.5 (54.3-62.8)</td>
</tr>
</tbody>
</table>

* See text
<table>
<thead>
<tr>
<th></th>
<th>DYSPLASIA</th>
<th>MATUREY DEFICIT</th>
<th>CONTROLS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+/n*</td>
<td>Sb (%)</td>
<td>P</td>
</tr>
<tr>
<td>OVERALL</td>
<td>5/19 = 26.3</td>
<td>33/100 = 33.0</td>
<td>65/413 = 15.7</td>
</tr>
<tr>
<td>VARIABLES ASSOCIATED WITH CONGENITAL DISEASE OF THE HIP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hip</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td>2/11 = 18.2</td>
<td>12/49 = 24.5</td>
<td>31/206 = 15.0</td>
</tr>
<tr>
<td>Right</td>
<td>3/8 = 37.5</td>
<td>21/51 = 41.2</td>
<td>0.082</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1/5 = 20.0</td>
<td>11/42 = 26.2</td>
<td>33/223 = 14.8</td>
</tr>
<tr>
<td>Female</td>
<td>4/14 = 28.6</td>
<td>22/58 = 37.9</td>
<td>32/190 = 16.8</td>
</tr>
<tr>
<td>Family history</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0/2 = 0.0</td>
<td>7/17 = 41.2</td>
<td>15/63 = 23.8</td>
</tr>
<tr>
<td>No</td>
<td>5/17 = 29.4</td>
<td>26/83 = 31.3</td>
<td>50/350 = 14.3</td>
</tr>
<tr>
<td>Breech presentation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0/4 = 0.0</td>
<td>2/6 = 33.3</td>
<td>1/14 = 7.1</td>
</tr>
<tr>
<td>No</td>
<td>5/15 = 33.3</td>
<td>31/94 = 33.3</td>
<td>64/399 = 16.0</td>
</tr>
<tr>
<td>INDICATORS OF HIGH-RISK DELIVERY/NEWBORN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prematurity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>-/-</td>
<td>-</td>
<td>1/5 = 20.0</td>
</tr>
<tr>
<td>No</td>
<td>5/19 = 26.3</td>
<td>32/95 = 33.7</td>
<td>62/388 = 16.0</td>
</tr>
<tr>
<td>Age of mother</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 20 yr</td>
<td>-/-</td>
<td>-</td>
<td>2/4 = 50.0</td>
</tr>
<tr>
<td>&gt; 20</td>
<td>5/19 = 26.3</td>
<td>31/96 = 32.3</td>
<td>54/371 = 14.6</td>
</tr>
<tr>
<td>Premature rupture of membranes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0/1 = 0.0</td>
<td>7/23 = 30.4</td>
<td>12/80 = 15.0</td>
</tr>
<tr>
<td>No</td>
<td>5/18 = 27.8</td>
<td>26/77 = 33.8</td>
<td>53/333 = 15.9</td>
</tr>
<tr>
<td>Apgar score 1 min.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 8</td>
<td>3/5 = 60.0</td>
<td>0.091</td>
<td>10/33 = 30.3</td>
</tr>
<tr>
<td>&gt; 8</td>
<td>2/14 = 14.3</td>
<td>23/67 = 34.3</td>
<td>43/315 = 15.8</td>
</tr>
</tbody>
</table>

* n : number of subjects

b S indicates sensitivity (positive rate in affected hips) ;

c Sp indicates specificity ; 1-Sp is the (false) positive rate in normal hips.

Only P-values under 0.10 are shown.

It can be argued that our gold standard, sonography, is not adequate. This is a major point. Boal and Schwenkter (3) found neither false-negative nor false-positive results among infants with hip dysplasia. Inconsistencies between the clinical exam and ultrasound can occur and they can be solved by x rays, as recommended by the agreement achieved in Paris (9). Unfortunately, we did not apply this procedure ; however, in a series of 111 dysplasias (diagnosed after referral by the pediatrician to our hospital after 2 years), sonography gave similar results to x ray (results not shown). Other authors have established that hip sonography permits an accurate and clinically relevant evaluation of the hip in the first days after birth (11, 26, 29-32); however, this point of view is not shared by others (6, 16, 22-24, 27). These latter studies highlight that ultrasound can give false positives, especially if used shortly after birth. We have tried to avoid it by repeating the ultrasound until 6 months after birth; however, 6 months might not be enough to rule out the spontaneous „healing” of an abnormal hip. Other authors (8), although they consider dynamic sonography a good procedure, do not recommend ultrasound because of problems of reliability.
Our sample size for assessing sensitivity was small. These results can be seen at best to be suggestive and not conclusive. On the other hand, the sample size was adequate for estimating specificity; it does allow us to find out whether there is a trend in obtaining positive results at screening, given certain variables. The lack of statistical significance in comparing positive rates does not preclude the existence of diagnostic biases. Significance is not necessary (although its presence lends further support) in the assessment of errors. Statistical significance mainly depends upon sample size, whereas errors rely on the absolute difference among estimates.

The sensitivity yielded by our results for the Ortolani-Barlow maneuver, 26.3% (almost 75% of the existent CDH will remain undiagnosed if only this maneuver is applied) is somewhat higher than the 10% obtained by Burger et al. (5) but lower than the figure reported by Tönnis et al. (30) who found absence of palpatory signs in 65.6% of dysplasias (i.e., a sensitivity of 34.4%), and the 75% of Venbrocks et al. (32). The low sensitivity found for clinical exam might be a consequence of the high sensitivity of sonography (leading to false positive results), as several Belgian and French authors have suggested (22, 23, 27). It is also very suggestive that the positive rate for physiologically immature hips is somewhat higher than for dysplasia. If a clinical maneuver does not distinguish between a disease and a normal condition and the sensitivity is low (many cases would remain unnoticed), a different procedure (such as sonography) ought to be applied.

The specificity of the Ortolani-Barlow maneuver found by Burger et al. (5) was 99%, whereas ours is clearly lower (84.3%). This means that clinical exam yields a false positive result in about one-sixth of the non-diseased hips.

The positive predictive value found by Burger et al. (5), 22%, is more favorable than our 5.9%. Our positive predictive value reached a similar figure when all clinical screening signs were combined, although sensitivity was much higher, but still the proportion of false positives remains high (low specificity) and thus imposes the use of a confirmatory procedure. These findings strongly suggest ultrasound should be used routinely for screening of CDH. The conclusions reached at the Conférence de Consensus held in Paris outline the use of ultrasound for diagnosing congenital dislocation (not dysplasia) of the hip when either clinical signs and/or risk factors are present (9), and so do Rombouts and Rombouts-Godin (23). The low sensitivity and specificity of clinical examination for CDH diagnosis in our study suggest the use of other procedures to detect it.

Our results also suggest that diagnostic biases may be present. Since sonography is not used for all newborns and the positive rate of the Ortolani-Barlow maneuver is higher in infants with two classic risk factors for the disease (family history and female sex), the disease will be more easily diagnosed in these babies. Future epidemiologic studies will find higher (biased) associations between CDH and these factors. The same facts could not be observed in breech presentation, probably due to the small number of cases. In this assessment of a likely diagnostic bias we included variables which a priori could influence medical attention: Apgar score, age of mother, premature rupture of membranes. Once again, in two of them a higher positive rate was found. The lower positive rate observed in preterm neonates confirms the greater difficulty of diagnosing the disease in these newborns.

The likely existence of a diagnostic bias poses a strong recommendation for future epidemiologic studies on CDH. They should be based on a population of newborns routinely screened by methods not influenced by previous knowledge of CDH epidemiology. Again, the same conclusion is suggested, that sonography or other (objective) procedures should be performed.

The clinical exam has a low sensitivity and positive predictive value. This suggests the use of other procedures to lower the number of missed CDH cases and to avoid diagnostic biases in future epidemiologic research on this condition.

Acknowledgements

To David Peiris for reviewing the English version of this manuscript.
REFERENCES

SAMENVATTING


Het doel van de studie is de betrouwbaarheid na te gaan van de klinische tekens, gebruikt bij de diagnostiek van aangeboren heupdysplasie en de leemten van deze tekens aan te tonen. Tweehonderdeenenzestig pasgeboren werden onderzocht en gevolgd op een ziekenhuis in de periferie. Al de babies werden klinisch en echografisch onderzocht tijdens de eerste 48 uren. De heupen werden geclasseerd volgens Graf. Sensibiliteit, specifiekheid en predictieve waarden werden bepaald. Negentien dysplasieën en honderd fysiologische immature heupen werden ontdekt. De sensitiviteit en specifiekheid van het manoeuvre van Ortolani-Barlow waren respectievelijk 26% en 84%. Wanneer al de klinische manoeuvres en tekens in acht genomen werden, was de sensitiviteit veel groter, maar de specifiekheid verminderde, zodat de positieve predictieve waarde het niveau van het teken van Ortolani-Barlow, 5% bereikte. Een hoger aantal positieve tekens van Ortolani-Barlow werd genoteerd wanneer de pasgeboren een hogere risicofaktoren hadden voor congenitale heupdysplasie (familiale anamnese en vrouwelijk geslacht), of zich in de groep hoger risico situeerden (lage Apgar score, jongere moeders). In conclusie: andere technieken dan het klinisch onderzoek worden ook aanbevolen bij de screening, om het aantal miskende gevallen te verminderen en diagnostische fouten te voorkomen in verdere epidemiologische studies van deze aandoening.

RÉSUMÉ


Le but de cette étude est de définir la fiabilité des manœuvres et signes cliniques pour le diagnostic de la luxation et de la dysplasie congénitale de hanche et de définir les lacunes de cette technique.

Un échantillon de 261 nouveau-nés fut examiné dans un hôpital périphérique. Tous les nouveau-nés furent examinés cliniquement et par échographie au cours des 48 premières heures après la naissance. Les hanche furent classées suivant le schéma de Graf. La sensibilité, la spécificité et la valeur prédictive furent évaluées. Dix-neuf hanche dysplasiques et 100 hanche physiologiquement immatures furent découvertes. La sensibilité et la spécificité de la manœuvre d'Ortolani-Barlow étaient respectivement de 26% et de 84%. Une sensibilité plus élevée fut notée lorsque tous les signes cliniques étaient pris en considération, avec une diminution de la spécificité de telle sorte que la valeur prédictive positive atteignait un niveau similaire à celui du test d'Ortolani-Barlow, soit 5%. Un nombre plus important de tests d'Ortolani-Barlow positifs était noté lorsquels nouveau-nés présentaient des facteurs de risque pour dysplasie congénitale (histoire familiale et sexe féminin) ou se situaient dans des groupes qui pouvaient être considérés comme à risque (faible score d'Apgar, jeunes mères).

D'autres techniques que l'examen clinique doivent être recommandées pour un dépistage efficace, diminuant le nombre des faux négatifs et évitant les erreurs de diagnostic au cours des études épidémiologiques futures.