Systematic review and meta-analysis of performance of second-trimester nasal bone assessment in detection of fetuses with Down syndrome

M. MORENO-CID*, A. RUBIO-LORENTE*, M. J. RODRÍGUEZ*, G. BUENO-PACHECO*, J. M. TENÍAS†, C. ROMÁN-ORTIZ† and À. ARIAS†

*Department of Obstetrics and Gynecology, Hospital General La Mancha Centro, Alcázar de San Juan, Spain; †Research Support Unit, Hospital General La Mancha Centro, Alcázar de San Juan, Spain

KEYWORDS: absent nasal bone; hypoplasia; pregnancy; sensitivity; specificity; trisomy 21; ultrasonography

ABSTRACT

Objective To review systematically the literature on diagnostic tests and performance of second-trimester sonographic assessment of nasal bone (NB) in identifying fetuses affected by Down syndrome.

Methods A search of studies involving screening tests for NB evaluation and measurements was carried out in the main international bibliographic databases (MEDLINE, EMBASE and CINAHL). Those considered to be relevant were then subjected to critical reading, following Critical Appraisal Skills Programme (CASP) criteria, by at least three independent observers. All data were extracted and tabulated by two independent investigators. A statistical synthesis of sensitivity, specificity and likelihood ratios was performed using specific software (Meta-DiSc).

Results From an initial list of 852 articles referring to ultrasound markers for Down syndrome, 207 relevant papers were selected. Following exclusions, 21 studies were included in the quantitative synthesis. The pooled estimates of positive and negative likelihood ratios were 40.08 (95% CI, 18.10–88.76) and 0.71 (95% CI, 0.64–0.79), respectively, for absent NB and 15.15 (95% CI, 8.15–28.16) and 0.47 (95% CI, 0.34–0.64), respectively, for hypoplastic NB. No relevant differences were found between the various means of defining nasal hypoplasia (multiples of the median (MoM) or percentiles). The biparietal diameter/nasal bone length (BPD/NBL) ratio showed somewhat higher sensitivity but lower specificity with a threshold effect.

Conclusions NB absence or hypoplasia show high specificity and low but acceptable sensitivity in identifying fetuses with Down syndrome. Screening performance is better with NB measurements as a function of MoM or percentiles rather than as the BPD/NBL ratio. Classification of women into various risk groups for Down syndrome does not affect diagnostic performance.

INTRODUCTION

Chromosomal abnormalities are present in 0.1–0.2% of all newborns1. Of these, trisomy 21 is the most frequent chromosomal alteration. Its prevalence is approximately 1 per 660 live births in the absence of any type of prenatal intervention2. Furthermore, it is associated with significant morbidity and carries a high economic and psychosocial cost3,4. Over the past decades, screening tests have been developed to improve detection rates for this syndrome at earlier stages in pregnancy5–8.

Hypoplasia or agenesis of the nasal bone (NB) is one of the most recently described ultrasound markers for Down syndrome. It is caused by incomplete calcification of the bone, which is often associated with trisomy 21 along with other abnormalities in bone growth and development9. Since 2002, when Cicero et al.10 published their results on NB abnormalities in fetuses with Down syndrome, numerous studies have investigated the association between absence of NB and this aneuploidy, with contradictory results obtained7,11–17.

NB detection becomes easier as the fetus grows, so that the NB is clearly visible in the second trimester, even in fetuses with trisomy 21. These studies also describe a more frequent association between NB hypoplasia and Down syndrome during this period of pregnancy12–19.

Because the size of the NB in fetuses without Down syndrome depends on gestational age, different meth-
ods have been proposed to define nasal hypoplasia in the second trimester of pregnancy, for example, NB measurements < 10th or < 2.5th percentile, nasal bone length (NBL) < 2.5 mm or a biparietal diameter (BPD)/NBL ratio of ≥ 10 or ≥ 11. In most cases, these definitions have been established arbitrarily with no prospective analysis of their screening performance, thus limiting their usefulness for estimating the risk of trisomy 21.

Other limitations of these studies include bias in terms of diagnostic criteria, lack of clarity in inclusion criteria and a lack of data on the presence of other associated ultrasound abnormalities. With this in mind, we believe a meta-analysis assessing the performance of NB measurements in Down syndrome screening is justified, especially given the evidence currently available.

METHODS

Diagnostic studies in the major international bibliographic databases (MEDLINE, EMBASE and CINAHL) were searched for online, with the final search conducted in October 2012. The references included in the selected articles were also reviewed to search for related citations.

Comprehensive search criteria were used to identify articles that included Down syndrome and ultrasound findings. These were combined with methodological filters developed by Haynes et al. to search for diagnostic studies20,21. The thesauri for MEDLINE (MeSH) and EMBASE (EMTREE) were also used. For the remaining databases, free text searches with truncations were used. From the studies thus identified, those analyzing the screening performance of NB measurements in the detection of Down syndrome were selected.

Diagnostic studies of Down syndrome and ultrasound tests using NB measurements were included in the analysis. The search was not restricted with regard to date or language of publication. The resulting search lists included the title and/or abstract (for most articles), which were used to carry out an initial identification of relevant documents.

Two independent researchers participated in this initial stage. An article was considered to be relevant if at least one of the observers considered it relevant. The agreement between observers was calculated. The full text of all articles considered to be relevant was then retrieved.

Studies considered to be relevant were subjected to critical reading by a group of at least three evaluators who used Critical Appraisal Skills Programme (CASP)22 criteria for diagnostic studies. For a study to be selected, it had to withstand the removal questions on the evaluation forms. Studies that were considered to be both relevant and methodologically correct were then examined by two independent observers who extracted the following data on the ultrasound findings analyzed: sensitivity (absolute and relative frequencies); specificity (absolute and relative frequencies); and likelihood ratios (LRs).

Statistical analysis

The meta-analysis was performed taking into consideration three types of indicators: NB absence; NB hypoplasia; and the BPD/NBL ratio.

Studies that examined NB distinguished between two subgroups: those which used a threshold below a multiples of the median (MoM) as the level of nasal hypoplasia; and those which used percentiles or direct measures of its length (e.g. mm).

For both nasal hypoplasia and the BPD/NBL ratio, the possible presence of a threshold effect was evaluated with the aid of graphical methods (summary receiver–operating characteristics curves), as well as with a statistical method calculating the Spearman correlation coefficient between sensitivity and specificity.

The data were analyzed using the software program ‘Meta-DiSc’, developed by the Clinical Biostatistics Unit of the Ramón y Cajal Hospital, Madrid23, as this program was considered suitable for meta-analysis of the accuracy of test data.

RESULTS

From an initial list of 852 articles, two independent observers selected 207 as being potentially relevant for the study of ultrasound markers. The agreement between both observers was high, with a kappa index value of 0.87.

Of the 207 articles selected, 70 were excluded for the reasons detailed in Figure 1. Of the 137 remaining articles, 24 analyzing the usefulness of NB assessment in diagnosing Down syndrome were chosen. Finally, 21 studies were selected for the quantitative synthesis of this systematic review. Of these, 13 analyzed NB absence as an ultrasound marker13,14,16,17,19,28–32,34,37,38. 15 measured NB hypoplasia or absence15–17,19,25–30,33,35–38 and eight assessed the BPD/NBL ratio14,17,19,24,29–32 (Table S1).

The quality of 15 of the studies was acceptable (medium or high)13–15,17,19,24,26–30,34–36,38 (Table S2), according to the criteria outlined in the Osteba Critical Reading guidelines39. Interobserver agreement was assessed in only three of these studies34,30,34.

In 14 articles, the study population consisted of pregnant women at high risk for Down syndrome, defined as those who had been referred for a comprehensive ultrasound scan either after a previously positive combined screening result or because of advanced maternal age or other risk factors for Down syndrome13–15,17,19,24,25,27,30,32,35–38. In six studies, the study population consisted of pregnant women at low risk for Down syndrome from the unscreened population16,26,28,29,31,33. In one study, the population’s risk for Down syndrome was not specified34.

Screening performance indicators are shown in Table 1. The results for NB absence and NB hypoplasia showed low sensitivity, especially for the former, but high specificity. No significant differences were found between
the different variants used to define nasal hypoplasia (MoM or percentiles), although the sensitivity was somewhat higher for studies that defined hypoplasia in terms of percentiles. Two studies that were not included in either subgroup were those of Lefebvre et al.\(^35,\)\(^4\) in which the criteria for hypoplasia were not defined, and Cicero et al.\(^15,\) in which direct measurements (2.5 mm) were used as criteria rather than MoM or percentiles.

The BPD/NBL ratio showed slightly higher levels of sensitivity, but with lower specificity. In addition, a threshold effect was observed for this indicator (Figure 2). Based on this finding, the screening performance of the BPD/NBL ratio for various cut-off points was assessed (Table 2). The study of Tran et al.\(^32\) was excluded from the subgroup analysis because it used two cut-off points – 14 and 18 – that were not used in any other studies.

The subgroup analysis showed a slightly higher sensitivity in low-quality studies and in those which included ultrasound scans from the latter part of the second trimester (Table 3). The group risk for Down syndrome was not associated with significant differences between validity indicators.

Only seven studies were performed on a single, identifiable ethnic group (four white and three Asian). It was impossible to detect differences in the absence of the NB (only one study in Asians) or in the BPD/NBL ratio (only one study in Caucasians). As for studies dealing

Table 1 Meta-analysis of diagnostic studies using second-trimester nasal bone (NB) ultrasound assessment to detect Down syndrome

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Hypoplasia of the NB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td>NB absence</td>
</tr>
<tr>
<td></td>
<td>All studies</td>
</tr>
<tr>
<td></td>
<td>Based on MoM</td>
</tr>
<tr>
<td></td>
<td>Based on percentiles</td>
</tr>
<tr>
<td></td>
<td>BPD/NBL ratio</td>
</tr>
<tr>
<td>Number of studies</td>
<td>13 (13)</td>
</tr>
<tr>
<td>Sensitivity (95% CI)</td>
<td>(0.280–0.396)</td>
</tr>
<tr>
<td>Specificity (95% CI)</td>
<td>(0.991–0.994)</td>
</tr>
<tr>
<td>LR+ (95% CI)</td>
<td>(18.10–88.76)</td>
</tr>
<tr>
<td>LR− (95% CI)</td>
<td>(0.71)</td>
</tr>
<tr>
<td>Presence of heterogeneity</td>
<td>Sp/LR+</td>
</tr>
<tr>
<td>Threshold effect</td>
<td>No</td>
</tr>
</tbody>
</table>

BPD, biparietal diameter; LR+, positive likelihood ratio; LR−, negative likelihood ratio; MoM, multiples of the median; NBL, nasal bone length; Sen, sensitivity; Sp, specificity.
Table 2 Meta-analysis of various cut-off points in biparietal diameter/nasal bone length (BPD/NBL) ratio for detection of Down syndrome in the second trimester

<table>
<thead>
<tr>
<th>Variable</th>
<th>BPD/NBL ratio cut-off point</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt; 9</td>
</tr>
<tr>
<td>Number of studies</td>
<td>4</td>
</tr>
<tr>
<td>Sensitivity (95% CI)</td>
<td>0.717 (0.577–0.832)</td>
</tr>
<tr>
<td>Specificity (95% CI)</td>
<td>0.595 (0.579–0.611)</td>
</tr>
<tr>
<td>LR+ (95% CI)</td>
<td>2.19 (1.29–3.71)</td>
</tr>
<tr>
<td>LR− (95% CI)</td>
<td>0.51 (0.27–0.99)</td>
</tr>
</tbody>
</table>

LR+, positive likelihood ratio; LR−, negative likelihood ratio; Sen, sensitivity; Sp, specificity.

Table 3 Analysis of subgroups according to different study variables (prior risk for Down syndrome, methodological quality of study and last week of pregnancy in which ultrasound scan was performed)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Absence of NB</th>
<th>Hypoplasia of NB</th>
<th>BPD/NBL ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sen</td>
<td>Sp</td>
<td>LR+</td>
</tr>
<tr>
<td>Risk for DS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High risk</td>
<td>0.359</td>
<td>0.991</td>
<td>35.49</td>
</tr>
<tr>
<td>Low risk</td>
<td>0.304</td>
<td>0.993</td>
<td>52.00</td>
</tr>
<tr>
<td>Quality of study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>0.346</td>
<td>0.997</td>
<td>134.00</td>
</tr>
<tr>
<td>Medium/High</td>
<td>0.332</td>
<td>0.990</td>
<td>24.62</td>
</tr>
<tr>
<td>Last week of US</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 24 weeks</td>
<td>0.316</td>
<td>0.991</td>
<td>33.00</td>
</tr>
<tr>
<td>≥ 24 weeks</td>
<td>0.385</td>
<td>0.998</td>
<td>60.71</td>
</tr>
</tbody>
</table>

BPD/NBL, biparietal diameter/nasal bone length; DS, Down syndrome; LR+, positive likelihood ratio; LR−, negative likelihood ratio; NB, nasal bone; Sen, sensitivity; Sp, specificity; US, ultrasound.

with the detection of nasal hypoplasia, we analyzed three studies conducted on Caucasians and two on Asian subjects. The sensitivity was somewhat higher in studies conducted on Asian subjects (0.700 vs 0.600), but the specificity was much lower (0.722 vs 0.980).

**DISCUSSION**

In our meta-analysis, the results for NB absence and hypoplasia show high specificity with acceptable sensitivity. Above all, this allows the correct diagnosis of healthy fetuses with a low rate of false positives which, in turn, contributes to decreased maternal anxiety when using this marker in Down syndrome screening.

In our study we obtained overall positive and negative LRs of 40.08 and 0.71, respectively, for the absence of NB and 15.15 and 0.47, respectively, for hypoplasia (Table 1), all of which are significantly lower than the results recently published by Agathokleus et al. (66.75 and 0.71 for NB absence and 23.27 and 0.46 for hypoplasia). These differences are most probably a result of the larger number of studies analyzed in our research (21 vs 12).

The published rates for NB absence or NB hypoplasia in fetuses with trisomy 21 vary widely, ranging from 26% to 77%, with specificity between 80% and 99%13–17,19,31. The determining factor influencing this variability is the study design, with lower rates being found in prospective studies conducted in real time17 as opposed to retrospective studies involving a review of ultrasound images13,14. In addition, some publications include exceedingly small NBs in the NB absence group15,19.

Another factor that may play a role in this variability is the choice of screening subjects, with some studies focusing on high-risk women and others including women at low risk for carrying a fetus with Down syndrome. Regarding this question, when validity indicators of NB absence or hypoplasia are published in diagnostic studies, one of the limitations most commonly mentioned in the discussion is that the results are associated with high-risk populations and that prospective studies of the same features should be carried out in low-risk populations14,30–32. Our meta-analysis showed no significant differences between studies conducted in different risk groups, which means that the LRs calculated in the various studies can be used interchangeably.

Another aspect in which the studies vary is the method used to estimate NB measurements. Basically, studies can be classified into those that use percentiles and those that use the MoM based on the BPD/NBL ratio. Indeed, different researchers apply various methods of measurement in order to minimize the influence of gestational age on this marker, thereby permitting the
calculation of individual risk based on the degree of hypoplasia.

According to studies conducted on normal fetuses, on average, the NB is between 4.3 and 4.7 mm at 15 weeks, reaching 7.5–8.2 mm at 22 weeks. This means that NB measurements in the 2.5\textsuperscript{th} percentile would be approximately 2.8 mm at 15 weeks and 5.6 mm at 22 weeks\textsuperscript{33,41,42}. In several studies, 2.5 mm is taken as the lower limit of NB measurement in the second trimester and is considered to be an absolute value independent of gestational age\textsuperscript{15}.

Although supporters of each NB measurement method – with percentiles or the MoM – argue that the strategy used in their study has the best diagnostic performance found in the literature, our meta-analysis found no significant differences in the validity indicators of the two methods, although the sensitivity was somewhat higher in studies that defined hypoplasia in terms of percentiles.

Although hypoplasia is defined by different cut-off points when using the MoM or percentiles, we found no threshold effect (i.e. a change in sensitivity and specificity for the detection of Down syndrome depending on the definition). There may be several reasons for this. First, it is very difficult to know what the actual cut-off point is when using the MoM; likewise, the exact measurement of the fetal NB remains unknown when percentiles are used. It is possible that the differences between the methods for estimating nasal hypoplasia are lower than expected, so that in the end they present no threshold effect.

In principle, the BPD/NBL ratio should be a more stable and homogeneous measurement within each cut-off point because it standardizes NB size with the BPD size. In order to choose the optimal cut-off when using the ratio, various authors recommend the one that obtains an acceptable detection rate (ranging between 10 and 11) with high specificity\textsuperscript{14,24,30,31}. According to our meta-analysis a suitable cut-off point would be a ratio of > 11 because with 86\% specificity it detects 57\% of Down syndrome cases. Nevertheless, even at the most suitable cut-off point, the validity markers are more accurate when assessing NB size as a function of MoM or percentiles.

It is worth noting that the sensitivity is somewhat higher both in low-quality studies and in those that include ultrasound scans from the latter weeks of the second trimester. With increasing gestational age it is easier to identify and measure the NB, so that the level of expertise required for measurement is lower and the sensitivity of the test is higher.

The finding that lower-quality studies show better sensitivity may be due to several factors. The lack of independence in the data collection may influence the test results. There may also be an inadequate spectrum of cases, with, for example, a clear differentiation between affected and unaffected cases or poor reproducibility of findings.

With regard to ethnicity-related differences, these data may indicate a greater ability to detect nasal hypoplasia at the expense of a higher rate of false positives (hypoplasia that does not actually correspond with Down syndrome).

Of all the articles selected, only three (14.3\%) analyzed interobserver agreement\textsuperscript{14,30,34}. It is well known that reliability assessment is an important aspect in evaluating studies on diagnostic tests; indeed, it is one of the items used to assess the quality of screening tests.

While the most important ultrasound marker of the midface hypoplasia typical of Down syndrome is absence or hypoplasia of the NB, the presence of subcutaneous edema in the frontonasal area of fetuses with Down syndrome gives rise to a separate series of ultrasound markers that have only been described in the past few years and have yet to be studied fully. These include both prenasal thickness (PT)\textsuperscript{13–45} and frontonasal fold thickness (FNF)\textsuperscript{46}.

Recently, some authors have studied the validity of the PT/NBL or FNF/NBL ratios as a screening method for Down syndrome. Despite the good results obtained to date, with a detection rate of 75–100\% and a false-positive rate of 1–5\%\textsuperscript{43–46}, the small number of studies published, most of which with prior knowledge of the fetal karyotype, calls for the corroboration of these results with adequately designed prospective studies.

Among the limitations of our study we would highlight the possible presence of publication bias, although in order to minimize this we have conducted an exhaustive search of relevant documents in all the main information sources available. Also, to avoid selection of unsuitable documents, an independent selection of articles was carried out by two observers who showed a high level of agreement. Assessment of study quality and data extraction was also performed in duplicate to avoid errors in data tabulation and analysis.

We believe that the determination of both intra- and interobserver agreement for the most significant ultrasound markers constitutes a highly interesting line of research, justified by the absence of this type of analysis currently available in the literature.

ACKNOWLEDGMENT

This work is partially supported by grants from the Fundación de Investigación Sanitaria de Castilla La Mancha (FISCAM). Ref: PI-2010/052.

REFERENCES


Nasal bone in Down syndrome


SUPPORTING INFORMATION ON THE INTERNET

The following supporting information may be found in the online version of this article:
- Table S1 Studies analyzed and markers studied
- Table S2 Characteristics of the studies included