Universal or selective screening of the neonatal hip using ultrasound?

A PROSPECTIVE, RANDOMISED TRIAL OF 15 529 NEWBORN INFANTS

K. J. Holen, A. Tegnander, T. Bredland, O. J. Johansen, O. D. Sæther, S. H. Eik-Nes, T. Terjesen
From the University Hospital of Trondheim, Norway

The aim of this study was to evaluate whether universal (all neonates) or selective (neonates belonging to the risk groups) ultrasound screening of the hips should be recommended at birth.

We carried out a prospective, randomised trial between 1988 and 1992, including all newborn infants at our hospital. A total of 15 529 infants was randomised to either clinical screening and ultrasound examination of all hips or clinical screening of all hips and ultrasound examination only of those at risk. The effect of the screening was assessed by the rate of late detection of congenital or developmental hip dysplasia in the two groups.

During follow-up of between six and 11 years, only one late-detected hip dysplasia was seen in the universal group, compared with five in the subjective group, representing a rate of 0.13 and 0.65 per 1000, respectively. The difference in late detection between the two groups was not statistically significant (p = 0.22).

When clinical screening is of high quality, as in our study, the effect of an additional ultrasound examination, measured as late-presenting hip dysplasia, is marginal. Under such circumstances, we consider that universal ultrasound screening is not necessary, but recommend selective ultrasound screening for neonates with abnormal or suspicious clinical findings and those with risk factors for hip dysplasia.

Received 24 January 2001; Accepted after revision 26 February 2002

The Ortolani and Barlow clinical tests for neonatal hip instability (NHI) can, in inexperienced hands, be unreliable and result in high numbers of false-positive and false-negative results. The former lead to overdiagnosis and overtreatment, whereas the latter may result in the late detection of hip dysplasia or dislocation.

During the last decade ultrasound has become widely used in the diagnosis of disorders of the hip in infants, especially in the detection of NHI. There is no consensus as to whether ultrasonography should be undertaken in all neonates or only in those with risk factors for neonatal or developmental hip dysplasia (HD). In Austria, universal ultrasound screening of the hip in newborn infants has been established. In other regions a selective screening policy is preferred. To solve this problem, randomised clinical trials (RCT) are required, as have been suggested by several authors. So far, only one prospective, randomised trial addressing the efficacy of ultrasound screening in neonates has been published.

Our aim was to determine whether universal ultrasound screening of all newborn infants is rational and desirable or whether a selective screening policy is preferred. We have evaluated this by a prospective, randomised trial in which the rate of late-detected HD was used as the outcome measure.

Patients and Methods

Between 1988 and 1992, all parents at the University Hospital of Trondheim were invited to have their newborn children participate in a non-selected RCT addressing the possible benefit of ultrasonographic screening of the neonatal hip. The study protocol was approved by the Ethical Committee at the University Hospital of Trondheim.

Sample size calculations. Our working hypothesis was that there is no difference in the incidence of late-detected HD between a group in which all newborn infants are examined by ultrasound and a group in which ultrasound examination is undertaken for clinical reasons only.

Between 1981 and 1985, Bredland and Terjesen found an incidence of three per 1000 of late-detected cases of HD in our county in Norway. The goal of additional ultrasound screening was to bring this incidence to below 0.5 per 1000, and to see if it would be significantly lower in a
group in which all neonates were examined by ultrasound as compared with that in which only those at risk for HD were examined.

The sample size was calculated according to Fleiss. With an alpha value of 0.05 and a beta value of 0.10 and a power of 0.90 in the study, it was calculated that we needed at least 12 194 infants (6097 infants in each group) to prove a significant difference in late-detected cases of HD in the two study groups. A five-year study was planned based on the yearly number of births at our hospital.

**Inclusion and exclusion criteria, and risk factor registration.** The parents were given written information about the study and most gave their written consent to participate. Babies of parents who refused consent were registered, but not included. Those with a permanent address outside our county were not included in the study because they could not be easily followed up. Newborn infants with very low birth-weight were included and examined later when their general health had improved. The information form included a short explanation of congenital HD and a questionnaire. The form was distributed and collected by the midwives. The parents were asked for any history of HD in the close family of the child, and also to specify the family relationship to those with a positive history of HD. In addition, all parents confirming a family history on the questionnaire were asked directly about this before randomisation, to rule out disorders of the hip other than HD. Other risk factors were registered from the birth protocol (e.g. NHI, doubtful clinical findings and foot deformities, including postural deformities). Breech position was defined as breech position at the time of delivery.

**The randomisation and study groups.** On the second day after birth, the orthopaedic surgeons evaluated the answers on the questionnaires and the results of the paediatricians’ examinations. After including the newborn infants according to the protocol, we used the birth protocol to carry out the randomisation by the method of random sampling numbers. From a table of random sampling numbers between 00 and 99, with a total of 2500 numbers, the infants were given numbers consecutively according to the birth protocol. When all the 2500 numbers were used, we started from the beginning of the table again. Those with numbers between 00 and 49 were assigned to group 1 and those with numbers between 50 and 99 were assigned to group 2. They were registered in the birth protocol by the midwives immediately after delivery. This was not influenced by the clinician doing the randomisation.

In group 1 all hips were examined by ultrasound in addition to the clinical screening by the paediatricians (universal ultrasound screening). In group 2 all hips were clinically screened by the paediatricians, but only those infants with risk factors for HD had an ultrasound examination (selective ultrasound screening). The risk factors considered to be an indication for ultrasound examination were NHI, doubtful clinical findings, HD in the family, breech position, and foot deformities. The term ‘doubtful clinical findings’ was used for hips with possible instability (by the Barlow test); ‘clicks’ were not included in this term.

Throughout the study period, the randomisation and examinations went on continuously.

**Clinical and ultrasound examinations.** All newborn infants were clinically examined on the first day of life by a senior paediatrician. From September 1989 almost all clinical examinations were carried out by the same experienced paediatrician (OJJ). This included the Ortolani and Barlow tests. These were also carried out by the orthopaedic surgeons at the time of the ultrasound examination.

Examination by ultrasound was usually done on the third day after birth by the orthopaedic surgeons who were involved in the study (TT, AT, TB, KJH). The ultrasound method described by Terjesen et al and Holen et al was used. This is mainly based on measurement of the percentage cover of the femoral head by the acetabular roof (femoral head cover). In addition, a subjective, dynamic evaluation of hip stability and acetabular anatomy was undertaken. We used a 5 MHz linear transducer (Sonoline; Siemens, Erlangen, Germany) to obtain one longitudinal and one transverse scan. The measurement of cover of the femoral head was determined from the longitudinal scan. In a previous interobserver study the 95% confidence interval of femoral head cover was 8% indicating that the method is sufficiently reliable in screening of the hip in newborn infants. Values of 47% in boys and 44% in girls for femoral head cover were defined as borderline by ultrasound.

**Neonatal hip instability.** Infants who needed treatment were treated with a Frejka pillow after the ultrasound examination for four months. The treatment protocol was changed during the study. During the first three years all newborn infants with clinical instability and femoral head cover below the borderline value, were treated immediately. During the last two years treatment has been delayed for two weeks. All those with clinical instability and abnormal ultrasound findings had a second examination at two weeks of age, and only those with persisting clinical instability and low values for femoral head cover, were treated. Assessment of the latter by ultrasonography and radiography was the most important measurement in the evaluation of the hips during follow-up.

**Registration of late-detected HD.** Late-detected HD was defined as that diagnosed after one month of age, including dislocation, subluxation, and acetabular dysplasia, based on the classification of Terjesen, Randén and Tangerud. Although our hospital is the only hospital treating patients with late-detected HD in our county, some infants included in the study could have been treated for late HD elsewhere. Therefore, information was requested from all hospitals in Norway involved in treating late-detected HD, asking if they had treated any children born in Trondheim during the study period.
If treatment was started during follow-up in children followed from birth, they were not registered as being detected late.

**Data registration and statistics.** For all newborn infants, pre-, peri-, and postnatal data were prospectively registered in a database at the National Centre for Fetal Medicine. For those who had an ultrasound examination all results from the registration form were added to this database.

We used Student’s *t*-test, the chi-squared test and Fisher’s exact test in the statistical analysis. *p* values below 0.05 were considered to be significant. The relative risk of detection of late HD was calculated as the rate in the ultrasound group divided by the rate in the control group, and precision was given with the 95% confidence interval.

**Results**

Of the 15,939 live births in the five-year period, 410 newborn infants (2.6%) were not included (Fig. 1). The remaining 15,529 infants were randomised. The trial profile is shown in Figure 1.

A total of 7,840 newborn infants was randomised to group 1 (universal ultrasound screening), whereas 7,689 were randomised to group 2 (selective ultrasound screening). Because of risk factors, a total of 872 newborn infants in group 2 (11.3%) had an ultrasound examination.

In group 1, 7,489 of the 7,840 newborns who had been randomised, were examined by ultrasound. The remaining 351 newborns missed ultrasound examination for various reasons. Some had been transferred to the neonatal intensive care unit on the first day of life because of prematurity or other health problems, and some died neonatally. Others were sent home or to other hospitals without the nursing staff informing us that they were ready for ultrasound examination. We did not have complete data regarding these 351 infants and they were omitted from Table I.

There were no significant differences between the two study groups concerning gender and birth rank (Table I) or the mean birth weight (3,490 g in group 1 and 3,472 in group 2). Concerning the risk factors NHI, breech position, foot deformities and family history for HD, the groups were also similar (Table I). About 15% of the infants had more than one risk factor for HD.
Table 1. Details of the randomised neonates and the distribution of risk factors between group 1 (ultrasound and clinical examination) and group 2 (clinical examination), by number and percentage

<table>
<thead>
<tr>
<th></th>
<th>Group 1*</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>7489</td>
<td>7689</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girls</td>
<td>3673 (49)</td>
<td>3752 (48.8)</td>
</tr>
<tr>
<td>Boys</td>
<td>3816 (51)</td>
<td>3937 (51.2)</td>
</tr>
<tr>
<td>Birth rank</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3400 (46.6)</td>
<td>3598 (46.8)</td>
</tr>
<tr>
<td>2</td>
<td>2644 (35.3)</td>
<td>2722 (35.4)</td>
</tr>
<tr>
<td>3 or more</td>
<td>1355 (18.1)</td>
<td>1369 (17.8)</td>
</tr>
<tr>
<td>Risk factors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NHI</td>
<td>73</td>
<td>66</td>
</tr>
<tr>
<td>Breech position</td>
<td>332</td>
<td>331</td>
</tr>
<tr>
<td>Family history of HD</td>
<td>351</td>
<td>338</td>
</tr>
<tr>
<td>Foot deformity</td>
<td>44</td>
<td>40</td>
</tr>
</tbody>
</table>

*351 infants, not examined by ultrasound, have been omitted from this table because of incomplete data

The rate of treatment with the Frejka pillow was 0.96% in group 1 and 0.86% in group 2. This difference between the two groups was not statistically significant. The overall five-year treatment rate was 0.92%. Because of persistent dysplasia after the treatment period three infants (two in group 1 and one in group 2) needed additional treatment with a plaster cast or abduction orthosis. Thereafter, the hips developed normally. Avascular necrosis of the femoral head occurred in one patient (a girl in group 2). Apart from these complications, the hips of all children who were treated developed normally.

During the follow-up period of six to 11 years (mean 8.5), one child in group 1 developed late-detected HD, representing a rate of 0.13 per 1000 births. Limited hip abduction was detected at three months of age by a physiotherapist who was treating her for metatarsus adductus. Because of the foot deformity this infant should have been routinely followed up by ultrasound but the protocol was for some reason not adhered to. Ultrasound and radiography showed left-sided acetabular dysplasia. She was treated by an abduction splint and the hip developed normally. In the subgroup of 351 children who were not examined by ultrasound at birth, there were no cases of late-detected HD.

Five infants in group 2 have presented with late-detected HD, a rate of 0.65 per 1000. They were all girls with no other risk factors for HD. One had bilateral dislocation, diagnosed at the age of 11 months. She needed tenotomy of adductor longus, closed reduction and a hip spica cast. So far, the hips have developed satisfactorily. Of the other four, two had unilateral subluxation and two unilateral acetabular dysplasia. Their ages at diagnosis were between five and six months. These four patients were treated by an abduction splint, and their hips developed normally. One of the girls with subluxation was detected at another hospital because the parents had moved to a neighbouring county. Apart from this patient, no other infants with late-detected HD were reported from other orthopaedic departments in Norway.

The relative risk for late-detected HD in the universal ultrasound group was 0.21 (95% CI 0.03 to 1.45), and the difference in late-detected HD between the two randomised groups was not statistically significant (Fisher’s exact test, \( p = 0.22 \)).

**Discussion**

This study met all the requirements of an RCT. The two study groups were similar concerning the number of infants, gender, birth-weight and rank, NHI and risk factors. With no breaks in the study during the five-year period, we assumed that no significant bias in the design of the study occurred.

Despite the scarcity of RCTs and the lack of scientific evidence of the efficacy of ultrasonography in the examination of the hip in newborn infants, the method has been widely accepted as an improvement on the clinical examination of the newborn hip.\(^5,6\) Universal ultrasound hip screening has been established in Austria\(^7\) and several authors have suggested RCTs to determine whether ultrasound screening should be universal (all newborn infants) or selective (risk groups).\(^6,9\)

In an RCT of 11 925 infants Rosendahl et al\(^10\) found no statistically significant difference in the incidence of late-presenting HD (including subluxation, dislocation and acetabular dysplasia) between those managed by clinical examination and those who had an additional ultrasound examination. Based on this, and our study, which to our knowledge are the only RCTs so far reported, it seems reasonable to conclude that the requirement of a dedicated and experienced clinical examiner is sufficient to give adequate screening of hips in newborn infants. The advantage offered by ultrasound screening becomes evident when such screening is compared with screening by less experienced examiners.\(^18\)

It seems possible that a few children with normal hips at birth may develop late dysplasia or dislocations.\(^4\) It has also been reported that some hips with normal clinical findings but abnormal sonography do not resolve spontaneously.\(^19,20\) Therefore, an incidence of late HD up to 0.5 per 1000 should be acceptable for clinical screening.

Previous reports from our hospital and from other parts of Norway have shown a relatively stable incidence of late-detected HD of approximately two to three per 1000, including frank dislocation, subluxation, and acetabular dysplasia.\(^3,18,21\) In the years before this study, Bredland and Terjesen\(^11\) found an incidence of three per 1000 at our hospital. The goal of our study was a reduction of late-detected cases to below 0.5 per 1000 in group 1, which would be significantly lower than the expected incidence in group 2. The number of infants with late-detected HD in group 2 was only five, and thus much lower than expected. This low incidence in the selective ultrasound screening
group was the main reason for the lack of a statistically significant difference between the two groups. Two reasons were probably responsible for the relatively good results in group 2, namely, the augmentation of the clinical screening by using ultrasound in neonates with risk factors for HD, and the special focus on the clinical screening among the involved paediatricians brought about by this study. Others have reported a similar tendency towards better results when attention is focused on a special problem by a clinical study.  

Our results in group 1 indicate that ultrasound screening could have the potential to eradicate late-presenting HD. The only case of late-detected HD in this group should have been avoided if the protocol had been followed. Our results are supported by those of Marks, Clegg and Al-Chalabi, who found no cases of late-detected HD in a group of 14,050 infants screened neonatally by ultrasound. Rosendahl et al found no cases of late-detected dislocation in their universal ultrasound group, one of late dislocation in the selective ultrasound group and two of late dislocation in the clinical screening group. With additional ultrasound screening late detection of subluxation or acetabular dysplasia may occur, but it seems that the most serious examples of late dislocation will be avoided.

Although we do not recommend universal ultrasound screening, some benefits of ultrasonography should be emphasised. First, direct visualisation and dynamic assessment augment the evaluation of the hips, provided that the examiner has sufficient knowledge and experience with ultrasound. Clinical instability can be confirmed or disproved and false-positive tests can be revealed, thus avoiding the trend towards over-diagnosis and unnecessary treatment. Furthermore, abnormalities of the hip can be detected by ultrasound in newborn infants with normal clinical findings. This experience, although not particularly focused on in this report, is the reason why we recommend selective ultrasound screening of infants with clinical instability, suspicious findings, and other risk factors.

It has been reported that prematurity could increase the risk of late HD. In an ultrasonographic evaluation of premature hips, Gardiner, Clarke and Dunn found no differences in Graf’s angles in preterm and term infants, indicating that prematurity alone should not represent a higher risk for late HD. In our study, 351 infants randomised to ultrasound examination were not examined. Most of these were transferred to the neonatal intensive-care unit because of prematurity. They had normal findings on clinical screening at birth and were followed closely by paediatricians during their first year of life because of their prematurity. No cases of late HD have been found in these children and they should therefore not influence our results.

We conclude that if the neonatal clinical screening of the hip is of high quality as in our study, universal ultrasound screening is not needed. A selective screening policy should be recommended.

No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the subject of this article.

References