The deformity index as a predictor of final radiological outcome in Perthes’ disease

D. Nelson, M. Zenios, K. Ward, M. Ramachandran, D. G. Little
From the Children’s Hospital at Westmead, Westmead, Australia

The deformity index is a new radiological measurement of the degree of deformity of the femoral head in unilateral Perthes’ disease. Its values represent a continuous outcome measure of deformity incorporating changes in femoral epiphyseal height and width compared with the unaffected side. The sphericity of the femoral head in 30 radiographs (ten normal and 20 from patients with Perthes’ disease) were rated blindly as normal, mild, moderate or severe by three observers. Further blinded measurements of the deformity index were made on two further occasions with intervals of one month.

There was good agreement between the deformity index score and the subjective grading of deformity. Intra- and interobserver agreement for the deformity index was high. The intraobserver intraclass correlation coefficient for each observer was 0.98, 0.99 and 0.97, respectively, while the interobserver intraclass correlation coefficient was 0.98 for the first and 0.97 for the second set of calculations.

We also reviewed retrospectively 96 radiographs of children with Perthes’ disease, who were part of a multicentre trial which followed them to skeletal maturity. We found that the deformity index at two years correlated well with the Stulberg grading at skeletal maturity. A deformity index value above 0.3 was associated with the development of an aspherical femoral head. Using a deformity index value of 0.3 to divide groups for risk gives a sensitivity of 80% and specificity of 81% for predicting a Stulberg grade of III or IV.

We conclude that the deformity index at two years is a valid and reliable radiological outcome measure in unilateral Perthes’ disease.

The long-term outcome in Perthes’ disease is related to the deformity of the femoral head and its congruency with the acetabulum. The grading system of choice for determining the long-term clinical outcome is that of Stulberg et al., which is recorded at skeletal maturity. While the lateral pillar classification is strongly predictive of outcome, it does not necessarily predict the outcome after surgery since it is usually assigned when treatment begins. Its purpose is to group patients by risk so that interventions can be compared in groups with a similar extent of the disease. There is no satisfactory system which allows assessment of outcome before skeletal maturity. Since follow-up of at least 15 years is required to assess outcome, it is not feasible to use the Stulberg grading to evaluate new forms of treatment for Perthes’ disease. We therefore designed a new radiological measurement, the deformity index, which is measured on plain anteroposterior (AP) radiographs. It incorporates changes in the femoral epiphyseal height and width in relation to the normal contralateral hip.

Our aim in this study was to assess the validity of the deformity index at two years as a surrogate for the measurement of radiological outcome at maturity. Intra- and interobserver reliability were also assessed.

Patients and Methods

The rationale for using the unaffected hip as a template for deformity was based on a number of factors. Measurements of sphericity such as Mose circles rely on the determination of the centre of the hip which in young patients is not always feasible. Also, bone age delay in Perthes’ disease may be associated with a variable delay in the development of the ossific nucleus. We assumed that delay in ossification would be similar in both the affected and unaffected hips. Finally, we noted that the characteristic deformity on the AP radiograph is flattening of the femoral head and lateral extrusion. The medial aspect of the femoral head and calcar are minimally affected. We decided to use the medial femoral head and calcar of the opposite hip as a reference point to allow direct
comparison of the flattening and lateral extrusion of the epiphysis.

The method for calculating the deformity index from an AP radiograph of the pelvis is shown in Figure 1. Using commonly available image-analysis software (Adobe Photoshop CS2 version 9.0; Adobe Systems, San Jose, California) a horizontally flipped image of the affected hip is overlaid as a semi-transparent image on to that of the unaffected hip, using the medial epiphysis and calcar to adjust for rotation and translation. The maximal orthogonal differences in height and width between the affected and normal epiphyses are measured. These data are summed and divided by the width of the normal growth plate. This last step controls for magnification differences on radiographs.

Initially, we randomly selected radiographs of ten normal hips and 20 from patients with Perthes’ disease to determine if numerical data produced by calculation of the deformity index matched subjective ratings of deformity of the hip. Each of three surgeons (DGL, DN and one who is not an author) assigned a category of normal, mild, moderate or severe to the radiographs. They measured the deformity index on two further occasions with an interval of one month. They were blinded to the previous results and in the subsequent reviews the order of the radiographs was randomly changed. Values for the deformity index were plotted against the subjective rating of normal, mild, moderate and severe. The intra- and interobserver reliability was determined by comparing the independent measurements of the normal pelvic radiographs and those from patients with Perthes’ disease.

In order to determine the predictive utility of the deformity index as a surrogate outcome measure, we used 96 radiographs of children with unilateral Perthes’ disease with follow-up to skeletal maturity from a recent large multicentre trial conducted by Herring et al.\textsuperscript{5,6} These radio-

\[ DI = \frac{h+w}{d} \]

Fig. 1a

\( h \) for height and \( w \) for width. The width of the normal physis is represented as \( d \). b) Diagram representing the normal hip (solid line) and the deformed hip (dashed line) and the measurements and formula for calculation of the deformity index (DI).

**Table I.** The Stulberg grades as used by Herring et al\textsuperscript{5,6}

<table>
<thead>
<tr>
<th>Stulberg grade</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Normal hip</td>
</tr>
<tr>
<td>II</td>
<td>The femoral head is within 2 mm of a concentric circle on AP* and frog-lateral radiographs</td>
</tr>
<tr>
<td>III</td>
<td>A non-spherical femoral head with an ovoid, mushroom or umbrella shape. The femoral head falls outside of a 2 mm concentric circle on AP and/or frog-lateral radiographs</td>
</tr>
<tr>
<td>IV</td>
<td>A femoral head with flattening of $\geq 1$ cm in a weight-bearing area on AP and/or frog-lateral radiograph</td>
</tr>
<tr>
<td>V</td>
<td>A femoral head with collapse, usually central, within a round acetabulum</td>
</tr>
</tbody>
</table>

* AP, anteroposterior
graphs had been graded using the Stulberg classification (Table I). From these, 32 were randomly selected for assessment from each of the groups which had been rated as Stulberg grades II, III and IV at maturity. The deformity index was measured by the first author (DN) on radiographs taken two years after presentation and at maturity. These data were compared with the corresponding Stulberg grade to determine the predictive value of the deformity index as a surrogate outcome measure. These data were analysed according to treatment subgroups of surgery (17 cases), bracing (67 cases) or non-containment (12 cases) (either a range of motion programme or no treatment).

Statistical analysis. This was performed using SPSS 11.5.1 (SPSS Inc., Chicago, Illinois). The mean deformity index at two years was compared with the Stulberg grade using analysis of variance (ANOVA), with a p-value ≤ 0.05 being considered statistically significant. A receiver operating characteristic curve measured the predictive value of the deformity index at two years and was applied to determine a cut-off value to separate patients likely to be graded as Stulberg grade III or grade IV (aspherical) compared with Stulberg grade II (spherical). The intraclass correlation coefficient was used to calculate intra- and interobserver reliability; a value greater than 0.75 was considered to be acceptable.

Results

Content validity. The initial study showed that mean measurements of the deformity index reflected our subjective grading of deformity. All ten normal radiographs had a deformity index of less than 0.1 and were ranked as the lowest ten scores. Mild and moderate hips had a degree of overlap, but moderate hips ranked higher in terms of the deformity index score. The hips classed as severe almost uniformly scored the highest deformity index (Fig. 2).

Repeatability. Figure 3 shows the mean versus differences for each of the three surgeons. The limits of agreement were identical for each of the surgeons and ranged between -0.1 and +0.1 indicating good intraobserver repeatability. There was a tendency for the measurement error to increase slightly with increasing measurement values but this was not significant. The mean difference, measurement error and 95% error range for each surgeon are shown in Table II. The intraobserver intraclass correlation coefficient for each of the observers was 0.98, 0.99 and 0.97, respectively. The first set of measurements by each of the surgeons had a measurement error of 0.005 and an error range of 0.1. Identical results were obtained for the second set of calculations of the deformity index. These results indicated acceptable interobserver repeatability.

Criterion validity. The mean age at diagnosis of the 96 children whose radiographs were reviewed was 8.0 years (5.8 to 11.0). The mean age of the patients whose radiographs were classified according to the Stulberg system was 16.4 years (10.2 to 20.5). The receiver operating characteristic curve in Figure 4 measured the predictive value of the deformity index at two years and decided a cut-off value to separate hips likely to be graded as Stulberg grade III or grade IV (aspherical) compared with Stulberg grade II (spherical). The intraclass correlation coefficient was used to calculate intra- and interobserver reliability; a value greater than 0.75 was considered to be acceptable. The area under the curve was greater than 0.5 (Z-test, p < 0.0001), thereby predicting which femoral heads would be spherical at skeletal maturity.
From the receiver operating characteristic curve and tabulated data, patients with a deformity index at two years of 0.3 or less were likely to have femoral heads which were spherical (Stulberg grade II) at maturity and those with a deformity index of 0.31 or greater to have aspherical femoral heads at maturity (Fig. 5). The sensitivity of this cut-off for the deformity index was 80% (95% confidence interval (CI) 70 to 90) and the specificity was 81% (95% CI 67 to 96). The calculated likelihood ratio was 4.3.

Figure 5 shows the mean deformity index at two years and at maturity for hips with Stulberg grades I or II, III and IV. There was a significant difference in remodelling in the three groups (ANOVA, p < 0.01). There was a significant trend for the deformity index at two years and maturity to increase as the Stulberg grade increased (ANOVA, p < 0.0001). Paired t-tests compared the deformity index values at two years or maturity for each grade and demonstrated a significant difference in grades I or II and III (p < 0.005), but not for grade IV (paired t-tests, p = 0.451). This trend was further demonstrated when subgroups were analysed for treatment.

For the surgically-treated group, six patients had a deformity index < 0.3 at two years. Two were Stulberg grade II and four Stulberg grade III at maturity. Of 11 surgical patients with a deformity index > 0.3 at two years, eight were Stulberg grade IV and three Stulberg grade III. Surgically-treated patients did not change their deformity index values over time, indicating that any improvement must have occurred before the radiograph was taken at two years.

For children treated in a brace, there were 29 with a deformity index < 0.3 at two years, 23 of Stulberg grade II and six Stulberg grade III. In the group of 38 children with a deformity index > 0.3 at two years, 12 were Stulberg grade IV, 19 Stulberg grade III and seven Stulberg grade II. In children treated by a brace, significant improvements in the deformity index were noted from two years to maturity (paired t-test, p < 0.05).

The deformity index at two years for the group which received no active treatment was poor. In this group, there were two patients with a deformity index < 0.3, both of whom were graded as Stulberg grade III at maturity. The remaining ten patients had a deformity index at

### Table II. The mean difference and measurement error for the deformity index in the two measurements by the surgeons in the study

<table>
<thead>
<tr>
<th>Surgeon</th>
<th>Mean difference between two measurements</th>
<th>Measurement error</th>
<th>95% error range</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.01</td>
<td>0.04</td>
<td>0.07</td>
</tr>
<tr>
<td>2</td>
<td>-0.01</td>
<td>0.04</td>
<td>0.07</td>
</tr>
<tr>
<td>3</td>
<td>-0.02</td>
<td>0.05</td>
<td>0.09</td>
</tr>
</tbody>
</table>
two years of > 0.3. These hips were all Stulberg grade IV at maturity. The deformity index did not change from the poor score at two years for patients in the non-containment group.

Discussion
The deformity index was created to give a continuous outcome measure of deformity of the femoral head which would predict the Stulberg outcome. We were interested in having a grading system in the healing phase which would reliably predict radiological outcome.

The Herring study⁶ evaluated bracing and osteotomy, either femoral varus or Salter. Other treatment options for Perthes’ disease are emerging, including pharmacological treatment,³,⁴ acetabular augmentation¹¹ and distraction techniques.¹² The healing phase associated with formation of new bone normally starts around two years after presentation.¹³ Our study has suggested that an objective measurement of deformity of the femoral head by the deformity index at that time predicts the outcome at skeletal maturity. The deformity index may therefore be a useful research tool in the assessment of treatment in the early stages of the disease. This contrasts with the lateral pillar classification, which, although an early marker of prognosis, has not been shown to be altered by intervention.

The two-dimensional deformity index incorporates changes in the femoral head in height and width and is thus expected to adequately quantify changes in the shape of the femoral head. We acknowledge that increased accuracy may be obtained by a three-dimensional measurement using either an AP and lateral radiograph in combination, or more complex imaging such as MRI or CT. However, obtaining such images on a repeated basis for clinical trials may not be feasible. Stulberg grades correlate with the long-term outcome,⁴ demonstrating the impact of the shape of the femoral head on the long-term function and the onset of osteoarthritis (OA). Hips of Stulberg grade I and grade II essentially have a round head and a good outcome. Those of grades III, IV and V have an aspherical head and have a poorer outcome. However, the aspherical but congruent hips of grades III and IV fared generally well in the long-term compared with hips of grade V. The specific criteria for each Stulberg grade were further clarified by Herring et al¹⁰ (Table I). For our study, radiographs were taken of hips of Stulberg grades II, III and IV. A sufficient number of grade I or grade V hips were not available.

We have shown that significant remodelling occurs in hips of Stulberg grades II and III, but not in those of grade IV. Thus it appears that at two years, those which would go on to become grade III have a similar deformity index to those which will become grade IV. The difference is related to the remodelling of the femoral head. Hips of Stulberg grade IV appear to define themselves as those which do not remodel, that is, do not significantly lower their deformity index from that at the two-year point to that at skeletal maturity, whereas the hips of Stulberg grade III remodel as indicated by their reduction in the deformity index. It can therefore be accepted that the deformity index measured at two years can identify those patients who will end up with spherical or aspherical femoral heads. It cannot, however, predict those with a final Stulberg grading of III or IV.

Data were also analysed according to whether surgery, bracing or physiotherapy only was undertaken. The deformity index at two years predicted outcome in those subgroups. Although the numbers were small, the deformity index showed differences in remodelling among the groups. Surgically-treated or untreated patients did not improve after two years. Braced patients continued to remodel between two years and maturity.

The disadvantage of the deformity index is that it can be used only for unilateral cases. The prevalence of bilateral involvement ranges between 8% and 24%.¹⁴,¹⁵ However, since such patients may have some underlying diagnosis, it is possible that they are inappropriate candidates for clinical trials.

The deformity index fared well in terms of repeatability compared with other grading systems of Perthes’ disease such as the Catterall, Herring and Salter-Thompson radiological classifications, which have all been used to predict prognosis in the active phase of the disease.¹⁶-¹⁸

In summary, the deformity index has good validity and excellent intra- and interobserver repeatability. We believe that the deformity index at two years can reliably predict spherical and aspherical femoral heads. The use of a continuous outcome measure enables trials of a feasible size to be adequately powered to test the hypothesis that new treatments may reduce the progression of deformity of the femoral head. Trials with promising results at two years based on the deformity index could be expanded and extended to meet the Stulberg outcome.

Supplementary Material
A further opinion by Dr A. C. Offiah is available with the electronic version of this article on our website at www.jbjs.org.uk.

We wish to thank Dr T. Herring for allowing us to have access to radiographs from the clinical trial of the Perthes’ study group and S. Dulai for her participation in the interobserver error study.

The author or one or more of the authors have received or will receive benefits for personal or professional use from a commercial party related directly or indirectly to the subject of this article.

References