Untreated acetabular dysplasia following treatment for developmental dysplasia of the hip (DDH) leads to early degenerative joint disease. Clinicians must accurately and reliably recognise dysplasia in order to intervene appropriately with secondary acetabular or femoral procedures. This study sought early predictors of residual dysplasia in order to establish empirically-based indications for treatment. DDH treated by closed or open reduction alone was reviewed. Residual hip dysplasia was defined according to the Severin classification at skeletal maturity. Future hip replacement in a subset of these patients was compared with the Severin classification. Serial measurements of acetabular development and subluxation of the femoral head were collected, as were the age at reduction, type of reduction, and Tonnis grade prior to reduction. These variables were used to predict the Severin classification.

The mean age at reduction in 72 hips was 16 months (1 to 46). On the final radiograph, 47 hips (65%) were classified as Severin I/II, and 25 as Severin III/IV (35%). At 40 years after reduction, five of 43 hips (21%) had had a total hip replacement (THR). The Severin grade was predictive for THR. Early measurements of the acetabular index (AI) were predictive for Severin grade. For example, an AI of 35˚ or more at two years after reduction was associated with an 80% probability of becoming a Severin grade III/IV hip.

This study links early acetabular remodelling, residual dysplasia at skeletal maturity and the long-term risk of THR. It presents evidence describing the diagnostic value of early predictors of residual dysplasia, and therefore, of the long-term risk of degenerative change.

The aim of treatment in developmental dysplasia of the hip (DDH) is to achieve concentric reduction, by closed or open methods, and to maintain this reduction through childhood and adolescence. If a concentric, stable reduction is maintained, the acetabulum has the potential to recover and resume normal growth and development.1-6 When the interaction between the natural remodelling forces of the hip and primary treatment fails, persistent residual acetabular dysplasia results. Dysplasia threatens long-term function by increasing the chance of early degenerative joint disease.7-13 In an attempt to avoid these problems, surgeons may perform secondary acetabular and/or femoral procedures to establish a more anatomically normal relationship between the acetabulum and femoral head. Although the long-term effectiveness of these procedures has yet to be firmly established, the use of these procedures is commonplace.

The characteristics of the residual hip dysplasia which requires surgical intervention has yet to be defined. Research has concentrated on factors associated with residual dysplasia8,11,12,14-28 and on the anatomical response to surgery.29-39 These papers often recommend guidelines for the use of secondary procedures, but these guidelines are usually too vague to be of practical use. Exceptions include Kim et al,21 Marafioti and Westin,38 and Wenger and Frick.40 More important than this lack of specificity is the fact that indications based on the retrospective study of surgical patients are biased because the indications, and therefore the outcomes of surgery, are in part pre-determined by the decision to operate. What is needed, is a set of predictors based on a non-surgical series where the outcomes are not confounded by other events and the complete process of remodelling can be observed and analyzed.3,8,14,35,41

The success or failure of remodelling cannot be completely evaluated for decades. Unfortunately, by the time degenerative joint disease is radiologically apparent, the opportunity for...
preventative measures has been lost, and salvage or reconstructive procedures are the only options. Therefore, we must rely on the evaluation of intermediate outcomes. We propose a model where residual dysplasia at skeletal maturity is treated as the intermediate outcome for predicting long-term outcome. The purpose of this study was to find early and reliable predictors of residual dysplasia at skeletal maturity, and to discuss the implications for secondary treatment.

Patients and Methods
There is much evidence that residual acetabular dysplasia and subluxation lead to early degenerative joint disease. It has been shown that the Severin classification, a simultaneous evaluation of acetabular dysplasia, femoral head deformity and subluxation, correlates well with long-term radiographic, clinical and functional outcome. The first step of this study involved a further validation of the Severin classification at maturity as a surrogate outcome for the long-term risk of degenerative joint disease.

The second step was to create a diagnostic test for residual dysplasia, using the Severin classification as the reference standard. This involved not only quantifying the relationship between early clinical and radiographic variables and the Severin classification (as has most commonly been done), but by using these variables to predict the classification. Predictor variables included the age, type and concentricity of reduction, and serial indicators of acetabular development. These variables have been used in various combinations by authors discussing residual dysplasia and the need for, and success of, secondary procedures. The predictor variables then constitute the diagnostic test for residual dysplasia. The predicted results of the diagnostic test were then compared with the observed Severin classification, and the accuracy of the test described in terms of sensitivity, specificity, and false positive and negative rates. The implications of using the results of this diagnostic test to indicate secondary procedures will be discussed, focusing on issues of over- and under-treatment.

Lastly, the cessation of remodelling is important to the decision to perform a secondary procedure. It has been proposed that remaining remodelling potential is necessary to reshape the acetabulum after acetabular osteotomy, and that the younger a child at the time of osteotomy, the less complicated the procedure and recovery. Therefore, it is important to know not only the predicted long-term outcome of the hip without surgery, but also the age before which the operation should take place in order to maximise success. We determined when the acetabulum ceased measurable remodelling, and combined this information with the diagnostic test results in order to propose guidelines for treatment.

Patient selection. The study was approved by our institutional review board. We reviewed the records of 196 children with DDH. Only patients with idiopathic DDH were included. In order to evaluate “natural” acetabular remodelling after reduction, we excluded eight patients who underwent secondary operative procedures (femoral or acetabular) and 21 with severe growth disturbance of the proximal femur (Bucholz and Ogden Type III) to exclude these potential influences on acetabular growth. A further 24 patients were excluded because pre-reduction radiographs were not available. Of the remaining 143, 81 patients had at least a seven-year post-reduction follow-up, with 58 having an open triradiate cartilage and, therefore, a measurable acetabular index or acetabular floor thickness. The study therefore included clinical and radiographic data from 58 patients (72 hips).

The mean age at reduction was 16 months (1 to 46). There were 45 female (78%), and 13 male patients. Unilateral dislocation was seen in 44 patients (76%) and 14 had bilateral dislocations. The left hip was involved in 28 of the 44 unilateral cases (64%).

Treatment. Closed reduction (48 hips) was achieved by gentle manipulation under general anaesthesia, applying traction with the hip and knee flexed while the greater trochanter was pushed anteriorly. The children then wore a hip-spica plaster cast for three months. Concentric reduction of the femoral head was thereafter maintained with an abduction brace at night and when sleeping during the day. Open reduction (24 hips) was indicated when a congruent, stable, closed reduction under anaesthesia and arthrographic control was not obtainable, or when extreme abduction was necessary to maintain reduction. Open reduction was performed through an anteromedial approach followed by immobilisation in a spica for three months. Thereafter, the patients wore a brace all the time for two months, and then at night and when sleeping during the day for a period of one or two years. A single surgeon supervised the care of all patients treated with open reduction (SLW).

Radiographic evaluation. All patients had radiographs before reduction and at skeletal maturity. Three patients had two radiographs available between the pre-reduction and final follow-up evaluations, four had three, eight had four, 19 had five and 24 had six radiographs; a total of 405 radiographs in 58 patients. The average time between films was 18 months (0 to 90). The mean age at the time of the radiograph which was used to evaluate the Severin classification was 14 years (SD = 3.06; 9 to 21 years).

All measurements were made by a single observer (MDM) who was not involved in the clinical care and blind to the Severin classification and long-term outcome. After all pre-maturity radiographs had been measured, two of us (JA, MDM), also not involved in the care of the patients’ examined the radiograph at maturity, and came to agreement concerning the Severin classification. Grading as Severin I or II was classified as “negative” for residual dysplasia, and these hips were considered to have had successful primary treatment. Severin III or IV hips were classified as “positive” for residual dysplasia and considered as...
failures of treatment. The following variables were evaluated as possible predictors of the Severin classification: age at reduction, type of reduction, pre-reduction Tonnis grade of displacement, and serial measurements of the acetabular index (measured using the superolateral aspect of the acetabulum), acetabular floor thickness, and the Smith lateral and superior centering ratios.

Acetabular remodelling was evaluated by observing the change in the acetabular index and acetabular floor thickness over time in relation to the Severin classification. Cessation of remodelling was defined as absence of significant change in the radiographic measurements over time.

**Long-term outcome.** The subset of patients who underwent closed reduction are being followed prospectively in a separate study. These patients are now between 32 and 59 years old and many have developed radiographic evidence of osteoarthritis. Additionally, some patients in this cohort have undergone femoral and/or pelvic osteotomy or total hip arthroplasty. We used total hip arthroplasty as a surrogate for the presence of degenerative joint disease. The patients who had an open reduction are also being followed in a separate long-term study and because of their young age (all currently under 30 years) were not included in the risk model of degenerative joint disease.

**Statistical analyses**

*Radiographic data adjustment and estimation.* The radiographs were taken periodically and not all patients had radiographs at all follow-up periods of interest. Since missing data can lead to biased estimates of risk (due to the use of incomplete data or the exclusion of cases with incomplete data), values for missing data are commonly imputed.

To partly correct this source of bias, all analyses were performed on a data set that included both observed and imputed values for the acetabular index, acetabular floor thickness, and lateral and superior centering ratios. Using a linear spline function, we estimated missing values of given predictors based on the magnitude and pattern of each individual patient’s existing measurements. Simultaneously, the function adjusted these values to estimate what would have been observed if all radiographs had been taken at the same time relative to the reduction, namely, six, 12 and 18 months, and two to seven years post-reduction. This function creates a complete series of data for each patient, with measurements corresponding to a standardized follow-up protocol. An example of the outcome of the linear spine function is given in Figure 1.

**Inferential statistics**

*Evaluation of Severin classification as a surrogate outcome.* Logistic regression was used to quantify the risk of having a total hip replacement given the Severin classification at maturity after controlling for the effect of age. Alpha was set at 0.05.

*Prediction of Severin classification at skeletal maturity.* Univariable and multivariable methods were used to evaluate the relations among the pre- and post-reduction risk factors, and between the risk factors and the outcomes Severin classification I/II versus III/IV. The relationship between the Severin classification, grade of hip displacement and type of reduction was assessed using Fisher’s exact test. The independent Student t-test was used to assess differences in the mean age at reduction, acetabular index, acetabular floor thickness, and centering ratios between Severin classifications at each time period. Univariable relationships were considered significant at alpha of 0.05. Multivariable logistic regression was used to develop models predicting the Severin classification. The age at reduction, type of reduction, Smith centering ratios, acetabular index and acetabular floor thickness were entered as risk factors. Models were created for each follow-up period in order to provide information concerning risk at several points during the clinical course. Statistical significance was determined through the use of the likelihood ratio test, and model fit was evaluated using the Hosmer-Lemeshow goodness-of-fit test. The models and the predictors within those models were considered statistically significant using alpha of 0.01. Based on these results, we created risk profiles for each time period. The profile gives the criteria (cut-off value) for low risk (<20% probability of failure), intermediate (<60% probability of failure) and high risk (>79% probability of failure).

*Acetabular remodelling.* Repeated measures analysis of variance was used to examine the change over time in the acetabular index and floor thickness for Severin I/II versus III/IV hips. Post-hoc Helmert contrasts were used to determine the point after reduction when significant change in the acetabular measurements ceased, indicating the cessation of measurable remodelling. Due to the multiple tests involved, and a desire to find the clinical as well as statistically significant end of remodelling, a stringent alpha of 0.005 was used.

---

**Graph showing a random patient’s observed acetabular index measurements and the interpolated values over time after reduction.**

---

**Fig. 1**

Graph showing a random patient’s observed acetabular index measurements and the interpolated values over time after reduction.
Results

Sample characteristics. At skeletal maturity, 47 of 72 hips (65%) were classified as either Severin I or II and 25 as Severin III or IV (35%). Prior to reduction, 23 hips (32%) were classified as Tonnis grade II, 24 (33%) as grade III, and 25 (35%) as grade IV. The mean pre-reduction acetabular index in affected hips was $40.4 \pm 7.6^\circ$ (20 to 54) and $25.7 \pm 6.2^\circ$ (12 to 50) in the contralateral, unaffected hips. The mean pre-reduction acetabular floor thickness was $7.7 \pm 1.8$ mm (5 to 12) in the affected hips and $7.4 \pm 1.4$ mm (5 to 12) in the contralateral, unaffected hips.

Of the 72 hips, 48 (67%) were treated by closed and 24 (33%) by open reduction. The mean age at reduction in the closed group was 18 months ($SD = 10.8; 1$ to 46 months) and eight months ($SD = 6.8; 2$ to 24 months) in the open reduction group. Of the 72 hips, 32 (44%) were adequately reduced as demonstrated by normal lateral (0.60 to 0.85) and superior centering ratios (0.10 to 0.20)$^5$ at six months post-reduction. At seven years post-reduction, all hips had achieved or maintained normal superior ratios, and 49 hips (68%) had normal lateral centering ratios.

The mean acetabular index prior to closure of the tri-radiate cartilage was $20.8 \pm 5.2^\circ$ (10 to 30) in the affected hips and $15.4 \pm 2.5^\circ$ (11 to 42) in unaffected hips. At two years post-reduction six hips (9%) had an acetabular index less than $20^\circ$; 19 (38%) at four years post-reduction; and 35 (52%) at seven years post-reduction. The mean floor thickness just prior to closure of the cartilage was $13.4 \pm 1.9$ mm (8 to 17) in the affected hips and $11.9 \pm 1.6$ mm (7 to 15) in the unaffected hips.

Current status was available for 36 of the 40 patients who had a closed reduction (43 of 48 hips). Three patients could not be located and one had been institutionalised. At a mean age of 43 years (32 to 59), 29 patients (34 hips) had had no further procedures since the index reduction. Additional procedures were performed in nine hips (21% of 43 patients): one patient had bilateral Chiari osteotomies at the age of 34, and five patients (seven hips) underwent total hip replacement at a mean age of 48 years (45 to 50).

Severin classification at maturity and risk of total hip replacement. The results of multivariable logistic regression indicated that a Severin classification of III or IV at skeletal maturity is associated with a significantly higher risk of total hip replacement than a classification of I or II (odds ratio = 2.376, 95% CI = 1.125, 5.016). The probability of having a total hip replacement was 7% for Severin I/II hips, 29% for Severin III hips, and 49% for Severin IV hips. The current age of the patient, or the age at which the total hip replacement was performed, was not related to the Severin classification at maturity ($p < 0.50$), nor was it related to the risk of requiring a total hip replacement ($p < 0.08$).

Risk factors and Severin classification at maturity. The relationship between risk factors and the Severin classification is summarised in Tables I and II. Patients with residual dysplasia (Severin III/IV) tended to be significantly older at
reduction (average 21 versus 13 months, p < 0.004). There was no significant relationship between the type of reduction and Severin classification; seven of 24 hips treated with open reduction were dysplastic (29%) compared with 18 of 48 hips treated by closed reduction (38%, p < 0.61). In hips reduced before 13 months, four (22%) treated with open reduction were dysplastic, compared with two of nine hips (22%) treated by closed reduction (p < 1.00). Likewise, of those hips treated between the ages of 13 and 24 months, three of six treated by open reduction and seven of 25 (28%) treated with closed reduction were dysplastic (p < 0.36). All 14 hips reduced after the age of 24 months underwent closed reduction, and of these, nine were dysplastic (64%). By examining the effect of age only, we found six (22%) of those hips reduced at or before the age of 12 months, ten (32%) of those hips reduced between 13 and 24 months, and nine (64%) of those reduced after 24 months of age had residual dysplasia (p < 0.04). Therefore, the prevalence of residual dysplasia increases with increasing age at reduction regardless of the type of reduction.

The Tonnis grade of dislocation was not significantly associated with the Severin classification (p < 0.36). The lateral centering ratios were not associated with Severin classification (all p values > 0.49); however, at each time period from two years post-reduction to the end of the series, a larger percentage of Severin III/IV hips had abnormal superior centering ratios than Severin I/II hips (p values < 0.04). The mean acetabular index and acetabular floor thickness measurements by Severin classification are given in Table II.

Small clinical, but highly significant, differences were seen early in the series on these measurements of acetabular remodelling between Severin classes. By one year post-reduction, there was a significant difference in the mean acetabular index, acetabular floor thickness, and abnormality of the superior centering ratio (defined as ≤0.10) at each time period. For example, the six-month post-reduction model included the age at reduction, and the six-month post-reduction values of the acetabular index, acetabular floor thickness, and the superior centering ratio. The six-year post-reduction model included the same variables but used their values at six years post-reduction. Interactions between the age at reduction and the other predictors were also included.

At six months post-reduction, the only significant predictor of Severin classification was the continuous variable age at reduction ($\beta = 0.07$, $p < 0.005$), therefore, the older the patient at treatment, the greater the probability of residual

<table>
<thead>
<tr>
<th>Time after reduction</th>
<th>Acetabular index (mean ± SD)</th>
<th>Acetabular floor thickness (mean ± SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Severin I/II (n = 44)</td>
<td>Severin III/IV (n = 23)</td>
</tr>
<tr>
<td>Pre-reduction</td>
<td>39.73 ± 7.69</td>
<td>41.61 ± 7.45</td>
</tr>
<tr>
<td>6 mths</td>
<td>32.14 ± 6.12</td>
<td>35.16 ± 6.28</td>
</tr>
<tr>
<td>12 mths</td>
<td>28.43 ± 5.22</td>
<td>32.75 ± 4.83</td>
</tr>
<tr>
<td>18 mths</td>
<td>26.54 ± 4.79</td>
<td>31.69 ± 4.30</td>
</tr>
<tr>
<td>2 yrs</td>
<td>25.28 ± 4.25</td>
<td>30.61 ± 3.71</td>
</tr>
<tr>
<td>3 yrs</td>
<td>23.08 ± 4.22</td>
<td>28.57 ± 3.57</td>
</tr>
<tr>
<td>4 yrs</td>
<td>21.72 ± 4.11†</td>
<td>27.34 ± 3.83</td>
</tr>
<tr>
<td>5 yrs</td>
<td>20.53 ± 4.09</td>
<td>26.38 ± 3.81</td>
</tr>
<tr>
<td>6 yrs</td>
<td>19.48 ± 4.06</td>
<td>25.71 ± 3.62</td>
</tr>
<tr>
<td>7 yrs</td>
<td>18.52 ± 4.27</td>
<td>25.13 ± 3.80</td>
</tr>
</tbody>
</table>

* = indicates point in time when significant change in acetabular index over time ceases (end of remodelling)
† = indicates point in time when significant change in acetabular floor thickness ceases (end of remodelling)
Age at reduction was also the strongest predictor in the 12- and 18-month models. One can locate the patient’s age at reduction on the x-axis and the corresponding risk of Severin III/IV on the y-axis (Fig. 2).

From two to seven years post-reduction, the strongest models included only the acetabular index ($\beta > 0.31$), therefore, the larger the acetabular index the greater the chance of failure. Using Figure 3, one can again locate the patient’s acetabular index on the x-axis and the corresponding risk of Severin III/IV on the y-axis. The relationship is shown on the graph for the acetabular index at two, four, and six years post-reduction. Because of the high correlations between the acetabular index and the other predictor variables, to include these variables in the models with the acetabular index was statistically redundant; this was confirmed by the lack of significant difference between the likelihood ratio statistics of the univariable and multivariable models. Likewise, the c-statistics indicate the onenvariable models discriminated between Severin classifications as well as the multivariable models. Age at reduction alone is associated with satisfactory discrimination between Severin classes (c statistic = 0.704), while the later models using the acetabular index (c statistics range 0.842 to 0.875) demonstrate quite adequate discriminative accuracy.

**Diagnostic tests for residual dysplasia.** These models can be used as diagnostic tests for residual dysplasia, where the Severin classification at maturity is taken as the reference standard. The profiles consist of the indicators age at reduction and acetabular index values at each follow-up period. Profiles corresponding to low ($< 20\%$), intermediate ($< 60\%$) and high ($\geq 80\%$) probability of residual dysplasia at two, four and six years after reduction are given in Tables III and IV. These profiles are premised on the 35% prevalence of residual dysplasia noted in this sample.

The profiles are used as follows. If a patient presents with an indicator equal to or greater than the criterion values ('cut off values') given in the table, the test for residual dysplasia would be considered positive. Lesser values than those listed would indicate a negative test. The sensitivity, specificity, false positive and negative rates and likelihood ratios were derived using an unbiased jack-knifing approach, which avoids the bias caused by using the same cases to estimate the predictive accuracy of the model. Examination of the false positive and negative rates over

### Table III. Tests using age at reduction to diagnose residual dysplasia

<table>
<thead>
<tr>
<th>Cut-off value</th>
<th>Minimum probability of dysplasia (Severin III/IV)</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>False + (%)</th>
<th>False - (%)</th>
<th>Diagnostic accuracy (%)</th>
<th>Likelihood ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 mths +</td>
<td>0.20</td>
<td>0.84</td>
<td>0.30</td>
<td>61</td>
<td>22</td>
<td>49</td>
<td>1.2</td>
</tr>
<tr>
<td>18 mths +</td>
<td>0.36</td>
<td>0.60</td>
<td>0.66</td>
<td>52</td>
<td>24</td>
<td>60</td>
<td>1.8</td>
</tr>
<tr>
<td>32 mths +</td>
<td>0.60</td>
<td>0.20</td>
<td>0.96</td>
<td>29</td>
<td>31</td>
<td>69</td>
<td>5.0</td>
</tr>
</tbody>
</table>

* total % correct (true positive + true negative)

### Table IV. Tests using the acetabular index to diagnose residual dysplasia at 2, 4 and 6 years after reduction

<table>
<thead>
<tr>
<th>Years after reduction</th>
<th>Cut-off value</th>
<th>Minimum probability of dysplasia (Severin III/IV)</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>False + (%)</th>
<th>False - (%)</th>
<th>Diagnostic accuracy (%)</th>
<th>Likelihood ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>26+</td>
<td>0.20</td>
<td>0.84</td>
<td>0.60</td>
<td>48</td>
<td>13</td>
<td>68</td>
<td>2.1</td>
</tr>
<tr>
<td></td>
<td>32+</td>
<td>0.60</td>
<td>0.44</td>
<td>0.92</td>
<td>27</td>
<td>25</td>
<td>75</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td>35+</td>
<td>0.80</td>
<td>0.16</td>
<td>0.94</td>
<td>43</td>
<td>30</td>
<td>67</td>
<td>2.7</td>
</tr>
<tr>
<td>4</td>
<td>23+</td>
<td>0.20</td>
<td>0.92</td>
<td>0.59</td>
<td>45</td>
<td>7</td>
<td>70</td>
<td>2.2</td>
</tr>
<tr>
<td></td>
<td>28+</td>
<td>0.60</td>
<td>0.52</td>
<td>0.91</td>
<td>24</td>
<td>22</td>
<td>78</td>
<td>5.8</td>
</tr>
<tr>
<td></td>
<td>30+</td>
<td>0.80</td>
<td>0.20</td>
<td>0.96</td>
<td>29</td>
<td>31</td>
<td>69</td>
<td>5.0</td>
</tr>
<tr>
<td>6</td>
<td>21+</td>
<td>0.20</td>
<td>0.83</td>
<td>0.59</td>
<td>49</td>
<td>13</td>
<td>67</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>26+</td>
<td>0.60</td>
<td>0.65</td>
<td>0.93</td>
<td>17</td>
<td>16</td>
<td>84</td>
<td>9.3</td>
</tr>
<tr>
<td></td>
<td>28+</td>
<td>0.80</td>
<td>0.30</td>
<td>0.98</td>
<td>13</td>
<td>27</td>
<td>75</td>
<td>15.0</td>
</tr>
</tbody>
</table>

* total % correct (true positive + true negative)
time shows that the chance of misdiagnosing residual dysplasia decreases as the child ages and the acetabulum begins to remodel.

Table III summarises the properties of diagnostic tests based on the age at reduction and is interpreted in the following manner. For illustrative purposes, the results for three ages at which the test could be considered positive are given (six, 18 and 32 months), as well as the associated probability of failure. For example, patients who have primary reduction at the age of 32 months or older have, at minimum, a 60% probability of developing residual dysplasia. These probabilities come from the regression equations in Table V. The consequences of using each of these three ages as the cut-off for a positive test are also given. If all patients treated after the age of 32 months are considered to be ‘positive’ for residual dysplasia, the test result would match the observed Severin classification in 69% of the cases (percent correct), 29% would be misdiagnosed as not having residual dysplasia (false negative rate) and 31% would be misdiagnosed as having dysplasia (false positive rate). The likelihood ratio\(^5\) of 5.0 for this cut-off value means that residual dysplasia is five times more likely in patients treated after the age of 32 months than earlier.

Table V. Results of logistic regression at each follow-up period

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Intercept</th>
<th>Regression coefficient</th>
<th>p value</th>
<th>Adjusted odds ratio</th>
<th>95% CI</th>
<th>c statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at reduction</td>
<td>-1.8286</td>
<td>0.0707</td>
<td>0.0056</td>
<td>1.073</td>
<td>1.024, 1.133</td>
<td>0.703</td>
</tr>
<tr>
<td>AI at 2 yrs after treatment</td>
<td>-9.2462</td>
<td>0.3059</td>
<td>0.0002</td>
<td>1.358</td>
<td>1.185, 1.605</td>
<td>0.842</td>
</tr>
<tr>
<td>AI at 3 yrs after treatment</td>
<td>-9.6311</td>
<td>0.3468</td>
<td>0.0002</td>
<td>1.415</td>
<td>1.218, 1.708</td>
<td>0.854</td>
</tr>
<tr>
<td>AI at 4 yrs after treatment</td>
<td>-10.0777</td>
<td>0.3804</td>
<td>0.0002</td>
<td>1.463</td>
<td>1.246, 1.796</td>
<td>0.886</td>
</tr>
<tr>
<td>AI at 5 yrs after treatment</td>
<td>-10.9117</td>
<td>0.3935</td>
<td>0.0002</td>
<td>1.482</td>
<td>1.259, 1.830</td>
<td>0.870</td>
</tr>
<tr>
<td>AI at 6 yrs after treatment</td>
<td>-10.2208</td>
<td>0.4186</td>
<td>0.0001</td>
<td>1.520</td>
<td>1.272, 1.921</td>
<td>0.874</td>
</tr>
<tr>
<td>AI at 7 yrs after treatment</td>
<td>-9.1297</td>
<td>0.3645</td>
<td>0.0001</td>
<td>1.469</td>
<td>1.249, 1.813</td>
<td>0.875</td>
</tr>
</tbody>
</table>

Properties of tests using the acetabular index at two, four and six years post-reduction are given in Table IV. For example, if the acetabular index at two years after reduction is 35° or larger, the patient has at least an 80% probability of developing residual dysplasia. If 35° is considered the cut-off for a positive test, the associated diagnostic accuracy is 67%, indicating that 33% of the patients would be misdiagnosed.

**Patterns of acetabular development by Severin classification.** Changes in the acetabular index and acetabular floor thickness over time was compared for Severin I/II and Severin III/IV hips (Figs 4 and 5, Table II). Repeated measures analysis of variance demonstrated significantly different patterns of acetabular development for these two groups (p < 0.0001). In hips with residual dysplasia, the acetabular index ceased improvement between four and five years after reduction, whereas the acetabular index in hips with a good result at maturity continued to improve until six years post-reduction. The acetabular floor thickness over time was also different in Severin I/II hips and Severin III/IV hips (p < 0.0004). Cessation of remodelling was noted between five and six years after reduction in the Severin III/IV hips, however, the acetabular floor thickness continued to

---

*Graph showing acetabular index related to Severin classification and years after reduction.*

*Graph showing acetabular floor thickness related to Severin classification and years after reduction.*
decrease significantly until the end of the series in the Severin I/II hips.

**Discussion**

This study sought, as Brougham et al\(^1\) recommended, to develop an “early and reliable radiological guide for the prediction of subsequent acetabular dysplasia”\(^2\) to “enable early rather than late surgery to be performed on acetabula with a poor prognosis and unnecessary acetabular operations to be avoided”. This need has been acknowledged in numerous papers concerning the treatment of residual hip dysplasia after initial open or closed reduction. However, this study is unique in addressing the multiple issues involved, namely, the diagnosis of residual hip dysplasia and the early prediction, timing and appropriateness of surgery.

Looking at the Severin classification and the prevalence of total hip arthroplasty, we found that residual hip dysplasia at skeletal maturity resulted in a higher risk of early degenerative joint disease in a subset of patients who were all followed up to a mean age of 44 years. This increased risk was also noted by Malvitz and Weinstein,\(^3\) who found a strong concurrent association between the Severin classification and degenerative joint disease: 46% of the Severin III/IV hips had severe degenerative changes compared with only 3% of the Severin I/II hips. Severin class, in turn, was also significantly related to function as measured by the Iowa Hip Rating, where 25% of Severin I and II hips had excellent and 58% good ratings, compared with a fair rating in 80% of Severin III hips, and poor ratings in 81% of Severin IV hips. We believe these studies provide adequate evidence to suggest the use of the Severin classification at maturity as a surrogate for long-term radiographic and functional outcomes.

Using the Severin classification as the reference standard for residual hip dysplasia, we assessed several measurements of acetabular development and femoral head subluxation as early diagnostic tests. Before treatment and up to two years after treatment, the age at reduction is of greater diagnostic value with regard to the future Severin classification than any of the radiographic parameters tested. This makes sense if one considers that there was no significant difference between the acetabular index in Severin I/II hips and III/IV hips prior to reduction, but that the age of the patients did differ significantly. This could indicate that the age of the patient, in conjunction with the adequacy of the reduction, determines subsequent remodelling. It is not until two years post-reduction that there is enough variability in the acetabular index between the two outcomes for the acetabular index to be independently predictive. When we consider the results of previous work, these relationships are not surprising,\(^4\) but confirming the use of age and acetabular index as early predictors is a step forward. The risk profiles indicate that patients at the age of 32 months and older have at least a 60% probability of developing residual acetabular dysplasia. This suggests a different initial treatment plan than that formulated for the child who presents at age six months with a 20% probability of residual dysplasia, or at 18 months with a 36% probability of residual hip dysplasia.

As the process of remodelling proceeds over time, the diagnostic tests proposed here become more accurate, perhaps reflecting cessation of remodelling. The literature contains disparate reports concerning the spontaneous remodelling potential of the acetabulum\(^5,6,16,18,20,27,44,49,59-62\) and this disagreement adds to the variability in treatment philosophy. The lower limit of remodelling which has been reported is two years of age\(^6,25,26\) or two years after reduction\(^27\) and the upper limit is 11 years of age\(^18\). The findings from this study fall within this range, but indicate that hips at risk of long-term dysplasia cease measurable remodelling earlier after reduction than hips with a better prognosis. In our series we found that the acetabular index stopped improving between four and five years post-reduction in hips which were eventually classified as Severin III/IV, whereas those eventually classified as Severin I or II continued to improve up to six years after reduction. Although all hips showed a decrease in the acetabular index over time, this series demonstrated a significant difference between the mean acetabular index in Severin I/II and Severin III/IV hips at one year after reduction. Some authors\(^7,23,35,41\) believe this initial improvement could be predictive of the future appearance of the hip.

The acetabular floor thickness changed over time and was strongly associated with the final Severin classification. This thickness stopped increasing in Severin I/II hips between six and seven years after reduction, but continued to increase until skeletal maturity in Severin III/IV hips. We believe that if the hip is well reduced at an early age, the normal growth of the acetabulum and the triradiate cartilage may be restored and this parameter should develop in approximately the same way as that of the unaffected hip. It will be necessary to evaluate acetabular floor thickness more extensively in normal children to establish its size and variability in order to set benchmarks for comparison with dysplastic hips.\(^15,47,48\)

Acetabular or femoral osteotomies are commonly performed when the joint does not develop satisfactorily after initial reduction.\(^6,23,29,35,36,41,63-72\) The indications and timing of such procedures are still a matter of debate. Salter\(^23\) is of the opinion that all hips treated at or after the age of 18 months should undergo acetabuloplasty concurrent with open reduction.\(^26\) Others believe the opposite, stating that early, routine, pelvic osteotomy to correct acetabular dysplasia may be unnecessary in many children.\(^18,44\) Other authors have observed the natural development of the hip and resolution of dysplasia without additional intervention.\(^6,8,16,18,20,44\) Most clinicians undertake secondary treatment when faced with subluxation or dislocation. Other published indications for secondary procedures include age at reduction,\(^6,11,25,26,30,34,68,73,74\) incongruity, intra-operative stability, anteversion, poor cover and residual acetabular dysplasia.\(^14,18,23,27,29,31-33,35,37,38,74,75\)
and the shape or size of the teardrop figure\textsuperscript{15,32} or the superolateral aspect of the acetabulum.\textsuperscript{22} As presented, these criteria are not specific enough to use as treatment guidelines, especially considering the widely-varying interpretations of such terms as acetabular dysplasia, joint incongruency, and lack of progressive development.

A second problem with the existing literature is the fact that the majority of studies describe the characteristics of patients who have undergone secondary procedures, but without empirical study of the necessity for these procedures. As stated before, conclusions from these studies are biased by the decision-making process used by the surgeon and parent. Other characteristics of the literature which limit its use in determining the need for surgery include lack of adequate follow-up, use of incomplete datasets, predictors for one point in time only, and the use of analyses which describe the difference in mean values (e.g. analysis of variance) but not the relative utility of variables as predictors of outcomes (e.g. discriminant analysis or multivariable regression), while simultaneously controlling for the effect of confounding variables. The most adequate work to date from a methodological point of view is that of Kim et al\textsuperscript{22} who evaluated annual radiographs taken over a mean of 13 years in a cohort of patients treated by closed reduction. They concluded that the shape of the superolateral aspect of the acetabulum and the centre-head distance discrepancy (CHDD) sufficiently predicts residual dysplasia as defined by the Severin classification. Their prediction rule states an upward sloping sourcil and a CHDD of \( \geq 6\% \) in a child at four to five years of age requires osteotomy to prevent permanent residual dysplasia. This appears to be a rule which functions well as it incorporates both the shape of the superior acetabulum and lateral displacement of the femoral head. However, since the CHDD reflects the displacement of the affected femoral head relative to the contralateral, normal hip, its use as a predictor is limited to unilateral cases of DDH. Also, a different rule could have emerged if the predictor variables had been tested simultaneously; it is possible that the CHDD behaves very similarly to the lateral centering ratios, and the slope of the superolateral aspect of the acetabulum is highly related to the acetabular index. Lastly, these predictions are specific to hips at four to five years after reduction.

It is beyond the scope of this study to recommend specific treatments for patients deemed at risk. However, the risk profiles and the time remaining for the remodelling process presented here can inform the decision to perform a secondary procedure. Ultimately, a decision based on the risk profiles from these models must incorporate the surgeon’s perception of outcome and the trade-off between over- and under-treatment. Surgeons who believe, like Wenger and Frick,\textsuperscript{40} that the long-term risk of early degenerative joint disease outweighs the risk of unnecessary surgery will wish to perform secondary procedures when the chance of falsely predicting no residual dysplasia (the false negative rate) is low. In this study, the lowest false negative rate was 7\%, associated with operating on all children with an acetabular index of 22° or more at four years after reduction. Conversely, surgeons who consider over-treatment more serious than under-treatment will wish to intervene at the point where the risk of falsely diagnosing residual hip dysplasia (false positive rate) is lowest. Therefore, surgeons with this philosophy will wait for the acetabulum to remodel, taking into account that false positive rates drop below 20\% only five years after reduction.

If the most conservative estimates of the cessation of acetabular remodelling are assumed correct, and remodelling is complete by the age of two, then surgeons who wish to intervene prior to the end of remodelling should perform secondary procedures in all children presenting at 18 months or older, or when the acetabular index is still about 30°. These two indications result in the lowest misclassification rates possible within that time frame. However, these rates are high (52\% to 74\%), demonstrating the difficulty involved in making accurate predictions at this early age. Based on the current study, clinicians who prefer to intervene while remodelling potential remains should be prepared to operate by four years after reduction. If future research shows that better results occur when secondary procedures are performed early (two to three years of age) rather than later (six to eight years of age), models predicting the need for surgery will have to be more sensitive and specific at those ages than the ones derived here.

On the other hand, if remodelling proceeds to the upper limits of published estimates, clinicians who wait until remodelling is finished will benefit from the increased accuracy of these models as the time after reduction increases. Up to four years after reduction, operating on hips with an acetabular index of 28° or more is associated with both the lowest false positive and overall misclassification rates, while operating on hips with an acetabular index of 23° or more results in the lowest false negative rate. When waiting for up to seven years after reduction, the lowest false positive rate occurs when operating at 28° at six to seven years after reduction; the lowest false negative rates occur at 22° and five years post-reduction or at 20° and seven years after reduction. Of all models, the lowest overall misclassification rate is associated with an acetabular index of 25° or greater at seven years after reduction.

By following the acetabular index over time and using the algorithms proposed here, it is possible to estimate the probability of intermediate- and long-term residual dysplasia based on measurements taken from very early radiographs. This ability to quantify the probability of a poor result increases confidence in the choice of treatment. Surgeons now have a range of therapeutic options to offer the patient with DDH. Therefore, future research should focus not only on the appropriate indications for secondary procedures, but on the procedures themselves. Current work\textsuperscript{76-80} has focussed on better defining acetabular insufficiency in order to improve the planning and performance of secondary procedures.
Other essential work includes study of the long-term outcome after secondary procedures. Studies to date have found that these procedures improve the Severin classification and other indicators of hip status in the short term. However, it is not known whether a hip that improves to a certain classification without surgery. Additionally, work needs to be done to examine the advantages and disadvantages of early procedures and to address more specifically the issue of timing in relation to prognostic factors such as remaining remodelling potential. These questions, like many others, will be most efficiently addressed through collaborative, prospective, controlled studies.

The authors would like to thank Ignacio V. Ponseti for allowing us to review his patients for this study. This work has been partially supported by Conselleria de Salud de la Comunidad Autonomia de Madrid and the St. Giles Foundation, New York City.

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
