Radiological hip shape and patient-reported outcome measures in healed Perthes’ disease

Aims
This study aimed to evaluate the relationship between hip shape and mid-term function in Perthes’ disease. It also explored whether the modified three-group Stulberg classification can offer similar prognostic information to the five-group system.

Methods
A total of 136 individuals aged 12 years or older who had Perthes’ disease in childhood completed the Patient-Reported Outcomes Measurement Information System (PROMIS) Mobility score (function), Nonarthritic Hip Score (NAHS) (function), EuroQol five-dimension five-level questionnaire (EQ-5D-5L) score (quality of life), and the numeric rating scale for pain (NRS). The Stulberg class of the participants’ hip radiographs were evaluated by three fellowship-trained paediatric orthopaedic surgeons. Hip shape and Stulberg class were compared to PROM scores.

Results
A spherical hip was associated with the highest function and quality of life, and lowest pain. Conversely, aspherical hips exhibited the lowest functional scores and highest pain. The association between worsening Stulberg class (i.e. greater deviation from sphericity) and worse outcome persisted after adjustment for age and sex in relation to PROMIS (predicted mean difference -1.77 (95% confidence interval (CI) -2.70 to -0.83)), NAHS (-5.68 (95% CI -8.45 to -2.90)), and NRS (0.61 (95% CI 0.14 to 1.08)), but not EQ-5D-5L (-0.03 (95% CI -0.72 to 0.11)).

Conclusion
Patient-reported outcomes identify lower function, quality of life, and higher pain in aspherical hips. The magnitude of symptoms deteriorated with time. Hip sphericity (i.e. the modified three-group classification of spherical, oval, and aspherical) appeared to offer similar levels of detail to the five-group Stulberg classification.

Cite this article: Bone Joint J 2023;105-B(6):711–716.

Introduction
The outcomes of Perthes’ disease are usually described in terms of radiological appearance. However, the orthopaedic literature increasingly recommends the use of ‘core outcome sets’ as a minimum set of outcome domains that should be reported in high-quality studies. The development of core outcomes involves multiple stakeholders, including patients and families, to define which outcomes are important. Radiological appearance, however, is almost invariably ranked as less important than functional outcomes. The core outcome set for Perthes’ disease consists of 14 measures encompassing various areas which impact the quality of life, as well as the radiological outcome.

Although it is plausible that radiological appearance would be related to patient-reported outcome measures in Perthes’ disease, this has not yet been demonstrated. Hailer and Penno investigated 61 patients with 28 years’ follow-up and concluded that long-term patient-reported outcome measures (PROMs) exhibited moderate-to-strong correlations with radiological measures of sphericity, femoral head enlargement, and femoral neck growth inhibition. Femoral head sphericity also showed a moderate correlation with the Harris Hip Score, in which hip function was shown to be worse in aspherical hips.

In this study, we have compared the radiological hip shape with mid-term functional outcomes.
Table I. Disease-specific inclusion and exclusion criteria.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Inclusion criteria</th>
<th>Exclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis of Perthes’ disease made while skeletally immature</td>
<td><strong>Inclusion criteria</strong></td>
<td><strong>Exclusion criteria</strong></td>
</tr>
<tr>
<td>Any of the following radiological features within the femoral epiphysis</td>
<td>Flattening</td>
<td>Able to understand the study documentation</td>
</tr>
<tr>
<td>Sclerosis</td>
<td>Collapse</td>
<td>Treatment for developmental hip dysplasia (not including double nappies)</td>
</tr>
<tr>
<td>Fragmentation</td>
<td>Reossification</td>
<td>Chemotherapy for malignancy</td>
</tr>
<tr>
<td>Collapse</td>
<td>Features may be evident on plain radiographs, or MRI</td>
<td>Sickle cell anaemia</td>
</tr>
<tr>
<td></td>
<td>Resident within England, Scotland, or Wales</td>
<td>MED or SED</td>
</tr>
<tr>
<td></td>
<td>Able to understand the study documentation</td>
<td>Coagulopathy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gaucher’s disease</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Same-sided hip fracture</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hypothyroidism</td>
</tr>
<tr>
<td>MED, multiple epiphyseal disease; SED, spondyloepiphyseal dysplasia.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table II. Radiological features of participants’ hip radiographs.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Participants, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stulberg class</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>19</td>
</tr>
<tr>
<td>II</td>
<td>27</td>
</tr>
<tr>
<td>III</td>
<td>37</td>
</tr>
<tr>
<td>IV</td>
<td>33</td>
</tr>
<tr>
<td>V</td>
<td>20</td>
</tr>
<tr>
<td>Hip shape*</td>
<td></td>
</tr>
<tr>
<td>Spherical</td>
<td>46</td>
</tr>
<tr>
<td>Ovoid</td>
<td>37</td>
</tr>
<tr>
<td>Aspherical</td>
<td>53</td>
</tr>
</tbody>
</table>

*Per the three-group Stulberg classification.

T scores, with normative data suggesting a score of 50 is the population mean with a standard deviation (SD) of 10. The minimal clinically important difference (MCID) of PROMIS Mobility is reported to be 4.2. Nonarthritic Hip Score. The Nonarthritic Hip Score (NAHS) is a measure of hip function intended for use in younger patients without arthritic problems or degenerative joint disease, and is widely used in the assessment of Perthes’ disease. The score is adjusted and between 0 and 100 where 0 represents a hip without meaningful function, and 100 represents a perfectly functioning hip. The NAHS MCID has been reported to be 8.7.

EQ-5D-5L. The EQ-5D-5L is a generic health-related quality of life measure. The questionnaire has five domains relating to activities of daily living with five levels of answer within each domain. Respondents are scored from no problems (score = 1.0) to extreme impairment on all five dimensions (value = -0.594). The MCID for EQ-5D-5L is 0.32.

Numeric Rating Scale for pain. This is a unidimensional measure of pain intensity. Among the various versions, the most commonly used is the selection of a whole number between integers of 0 to 10. A score of 0 represents ‘no pain’ while a score of 10 represents ‘worst imaginable pain’. On average, a reduction of one point or reduction of 15% represents a MCID for the NRS.

Radiological outcome. The Stulberg classification is the gold-standard measure of hip shape. We used an established classification tree to ensure greater consistency in the descriptions. Radiographs of the hip in the residual/healed stage in late childhood/adolescence were assessed by three fellowship-trained consultant paediatric orthopaedic surgeons (see Acknowledgements). The surgeons independently assessed each image, with the decision of the third surgeon used to resolve discrepancies between the other two by majority vote. We used both the traditional five-group Stulberg classification and the modified three-group classification.

Age groups of participants. At the point of PROMs completion, participants were divided into three age categories: 12 to 16 years (adolescents), 17 to 25 years (young adults), and ≥ 26 years (adults).

Statistical analysis. Kruskal-Wallis one-way analysis of variance was used to compare non-normally distributed continuous variables. As the outcome data were right-skewed, log-linked generalized linear models were used to determine whether
Stulberg class is independently associated with outcome after adjustment for age and sex. Age was included as a continuous variable within these models. Statistical analyses were performed using SPSS v. 26.0 (IBM, USA) and StataIC v. 15 (StataCorp, USA), with p < 0.05 used as the threshold for statistical significance.

**Results**

Of the 856 patients invited to participate, 300 returned questionnaires and of these, 291 returned complete responses. A total of 25 were excluded because they had a history of hip arthroplasty and 130 because hip radiographs were unavailable, which left a study population of 136 participants. The age range was 12 to 55 years.
55 years (mean age at PROM completion 24 years) and 79% (n = 107) were male.

Using the five-group Stulberg classification, most participants had Stulberg group III hips (n = 37) (Table II). Using the three-group classification, most participants had aspherical hips (n = 53), followed by spherical (n = 46), then ovoid (n = 37) (Table II).

Boxplots were plotted to visually represent the distribution of PROM scores for hip shape and age category, and Stulberg class and age category (Figures 1 to 4).

The relationship between hip shape, age, and sex. Advancing age was significantly associated with worse outcomes according to PROMIS (p = 0.040) but not NAHS (p = 0.099), EQ-5D-5L (p = 0.134), or NRS (p = 0.984). Male sex was associated with
better outcomes across all four PROMs: PROMIS (median 56.9 (interquartile range (IQR) 47.4 to 56.9) vs 47.4 (37.9 to 56.9); p = 0.005), NAHS (92.5 (76.3 to 97.5) vs 75.0 (50.0 to 90.0); p = 0.003), EQ-5D (0.837 (0.642 to 1.000) vs 0.678 (0.479 to 0.837); p = 0.019), and NRS (2 (0 to 5) vs 5 (1 to 7); p = 0.012). In this study, hip shape was not significantly associated with either age (p = 0.608) or sex (p = 0.063).

The effect of hip shape on function. Aspherical hips had significantly worse function compared to spherical hips, as shown by both PROMIS Mobility and NAHS. For both scores, self-reported hip function was worse in participants aged ≥ 26 years compared to those aged 12 to 16 years. Within a generalized linear model, each worsening Stulberg class was associated with significant reductions in the PROMIS Mobility T-score (predicted mean difference -1.77 (95% CI -2.70 to -0.83)) and NAHS (-5.68 (95% CI -8.45 to -2.90)).

Participants aged ≥ 26 years appeared to show the largest differences in function scores between spherical and aspherical hips, compared to other age categories. The NAHS (Figure 2) appeared to be better at distinguishing the range of hip function captured from participants than the PROMIS Mobility score, as it showed more of a difference than PROMIS between hip shapes and Stulberg classes (Figure 1).

The effect of hip shape and age on general quality of life. Within a generalized linear model, worsened Stulberg class was not associated with changes to the EQ-5D-5L (predicted mean difference -0.03 (95% CI 0.72 to 0.11)). Similarly, with regard to function, quality of life was worse in participants aged ≥ 26 years compared to those aged 12 to 16 years. The effect of hip shape and age on pain. Within a generalized linear model, worsening Stulberg class was associated with higher pain scores as captured by the NRS (predicted mean difference 0.61 (95% CI 0.14 to 1.08)). Pain in participants aged older than 26 years was reported to be worse than in participants aged 12 to 16 years.

Discussion
In our study individuals with a spherical femoral head exhibited better function, quality of life, and lower levels of pain than those with aspherical hips. We found that increasing age resulted in a decrease in function and quality of life and increase in pain for all hip shapes with Perthes’ disease. Worsening hip shapes was also associated with poorer patient-reported outcomes. Furthermore, we observed that the Stulberg three-group classification appeared to offer similar information to the five-group classification.

Stulberg et al21 in 1981 were the first to explore the relationship between secondary osteoarthritis and radiological assessment. While the Stulberg classification has largely remained a standard in the assessment of Perthes’ disease, few have compared the classification with functional, quality of life, and pain outcomes in a large group of patients. Wiig et al14 have shown that the Stulberg three-group classification has increased inter-rater reliability, and that the discriminatory ability appeared similar between the three- and the five-group classification in terms of predicting pain, function, and quality of life. The three-group classification is known to be an effective long-term predictor of radiological outcome in Perthes’ disease,22 and our study suggests that it is effective in predicting patient-reported outcomes. Given the proven usefulness of the three-group classification, with improved interobserver reliability and little detail lost compared to the five-group classification, this appears a more useful tool to describe the outcomes of Perthes’ disease. In addition to the three-group classification, other authors have also identified the possibility of additional prognostic information through MRI.23

Joint-specific PROMs such as the HHS and the NAHS have exhibited moderate correlations with radiological measures of hip sphericity deviation.4 Our findings support this as we show a decrease in function in aspherical hips, with the NAHS appearing the most sensitive measure to change in hip shape among the outcomes used within our study. The association between EQ-5D-5L and hip shape was less clear than other functional outcomes, though it seems likely that a quality of life tool was less sensitive to change than other outcomes, which are more specific to pain and limited function.

One potential source of bias is the sampling of our study population, which was affected by non-responders to the invitation and the unavailability of hip radiographs. These biases are likely to have resulted in over-sampling of individuals with poorer-functioning hips compared to the broader Perthes’ population. However, we achieved a similar number of patients among all disease severities, which enables a strong analysis and takes into account the correlation between hip shape and outcomes.

Hips in this study were all classified in the healed stage of Perthes’ disease, though this is often many years after the onset of disease. Huhnstock et al22 employed the three-group classification using radiographs five years after disease onset and found that this was reliable at predicting long-term outcomes at this stage. It would be useful to see if long-term outcomes could be reliably determined even earlier, as the minimum time possible to accurately predict long-term outcomes would be useful to drive the minimum duration of follow-up for randomized controlled trials.

In conclusion, this study has demonstrated that hip shape was associated with long-term patient-reported outcomes. Patient-reported function deteriorated with age and the magnitude of this decline was related to the degree of hip deformity. Furthermore, we found that the three-group Stulberg classification appeared to give similar information to the five-group classification.

Take home message
- Adolescents and young adults with hip shape with greater deviation from sphericity exhibited poorer function and increased pain at the healed stage of Perthes’ disease.
- The three-group Stulberg classification appeared to give similar information to the five-group Stulberg classification.

Twitter
Follow D. Metcalfe @TraumaDataDoc
Follow D. C. Perry @MrDanPerry

References


12. Rose M, Bjornsen JB, Gendek B, Bruce B, Fries JF, Ware JE. The PROMIS Physical function item bank was calibrated to a standardized metric and shown to improve measurement efficiency. *J Clin Epidemiol.* 2014;67(5):516–526.


