An independent assessment of reliability of the Melbourne Cerebral Palsy Hip Classification System

M. W. Shrader1
A. L. Koenig2
M. Falk3
M. Belthur3
C. Boan3

Abstract
Purpose Neuromuscular (NM) hip dysplasia is common in patients with cerebral palsy (CP). Traditionally, migration percentage (MP) has been used to measure the severity of NM hip dysplasia; however, the MP has some limitations. The purpose of this study is to determine the intra- and inter-reliability of the Melbourne Cerebral Palsy Hip Classification System in the typical paediatric population of patients with CP.

Methods A total of 65 anteroposterior pelvis radiographs in patients (age range 12 years to 21 years) with CP spanning all grades (I to VI) of the classification system were identified and collected for analysis in this institutional review board approved study. Four paediatric orthopaedic surgeons and one orthopaedic surgical resident classified each radiograph according to the Melbourne system. Then, at least four weeks later, the raters repeated the process with a re-randomised order of radiographs. Statistical analysis was performed using the intraclass correlation coefficient (ICC) where < 0 denotes poor agreement and > 0.8 indicates almost perfect agreement.

Results The interobserver reliability was found to be excellent with the ICC of 0.853 (0.813 to 0.887) and 0.839 (0.795 to 0.877). The intraobserver reliability was also found to be excellent with the ICC in the range of 0.838 to 0.933 among the raters. Subgroup analysis indicated no differences in the reliability of observers based on clinical experience.

Conclusion This study independently demonstrates that the Melbourne Cerebral Palsy Hip Classification System for NM hip dysplasia in patients with CP can be reliably used for communication among various healthcare providers and research and epidemiological purposes.

Cite this article: Shrader MW, Koenig AL, Falk M, Belthur M, Boan C. An independent assessment of reliability of the Melbourne cerebral palsy hip classification system. J Child Orthop 2017;11:334-338. DOI: 10.1302/1863-2548.11.170077

Keywords: Melbourne Cerebral Palsy Hip Classification System; neuromuscular hip dysplasia; cerebral palsy; migration percentage; reliability

Introduction
Children with cerebral palsy (CP) commonly have significant hip dysplasia that can lead to a negative impact on quality of life. The consequences of hip subluxation and dislocation include significant pain, discomfort and imbalance, decreased ability to maintain hygiene specifically in the perineal area, and the loss of ability to stand and walk.1,2 The severity of hip disease in CP is directly proportional to the patient’s classification using the Gross Motor Functional Classification System.3-5 The ability to accurately and reliably classify the degree of hip pathology in patients with CP has the potential to improve care and outcomes in this patient population in addition to serving as a research tool for determining effectiveness of interventions.

Radiographic measurements used in the evaluation of hips with neuromuscular hip dysplasia include the migration percentage (MP) of Reimers and the acetabular index of Hilgenreiner.6 The MP is a continuous variable measuring the percentage of the femoral head lateral to the acetabular margin. The reliability of using the MP has been investigated and found to have acceptable reliability and repeatability, within an error of 6% to 13%.7-8

However, although it is a commonly used measurement method, the MP has several disadvantages. First, accurate measurements are extremely difficult to obtain in cases of severe dysplasia. Secondly, anterior and posterior dislocations (versus the more common posterolateral dislocation) do not always show an increased MP.9 Thirdly, the MP fails to include additional morphological features of the hip, such as acetabular dysplasia.9 Finally, an ordinal classification system has benefits over a continuous system in
communication among various healthcare providers and for research and epidemiological purposes.10

Recently, a new classification system was developed as a tool to grade the final outcomes of hip pathology in CP. The Melbourne Cerebral Palsy Hip Classification System is based on the following gross morphologic features: integrity of Shenton’s arch; shape of the femoral head; shape of the acetabulum; and pelvic obliquity.8 The classification system ranges from Grade I to VI. Grades I and II represent a normal hip and near normal hip, respectively. Grade III represents a dysplastic hip. Grade IV represents a subluxated hip. Grade V represents a complete hip dislocation and Grade VI represents a dislocated hip that requires salvage surgery.

Reliability of the Melbourne Cerebral Palsy Hip Classification System has been demonstrated by the authors of the original study in the location of the original study, with inter- and intraobserver reliability of excellent to almost perfect.11 An independent evaluation of this system’s reliability is important to further validate its use. The purpose of this study is to test the intra- and inter-reliability of this classification system as an independent evaluation.

Materials and methods

Institutional review board approval was obtained for this radiographic measurement analysis. Hip radiographs (anteroposterior (AP) pelvis) in patients aged 12 to 21 years with a diagnosis of CP were identified at a tertiary care paediatric referral hospital. The range of hip pathology spanning the entire range of the classification system was screened from the medical records. The radiographs were blinded and presented to five fellowship-trained paediatric orthopaedic surgeons and one orthopaedic surgical resident on two separate occasions eight weeks apart. The level of experience varied among the surgeons: two have greater than ten years of experience; two have greater than five years of experience; one recently completed their fellowship; and one is currently a fourth-year orthopaedic general resident on two separate occasions eight weeks apart. The level of agreement varied among the surgeons: two have greater than ten years of experience; two have greater than five years of experience; one recently completed their fellowship; and one is currently a fourth-year orthopaedic resident. The surgeons were provided instruction in this classification system prior to performing this study, and each radiograph was independently classified into this system by each orthopaedic surgeon.

The hip classification is described as follows and is illustrated in Figure 1. The classification system is based on a combination of femoral head and acetabulum morphology, pelvic obliquity and the MP. The original authors felt that this combination of both a description of morphology and the MP made for the best classification system.

Grade I is a normal hip with a MP of < 10°, Shenton’s arch is intact, the femoral head is round, there is normal acetabular development and a pelvic obliquity of < 10°. Grade II is a near normal hip with a MP of 10% to 15%, an intact Shenton’s arch, a round or almost round femoral head, normal or near normal acetabular development and a pelvic obliquity of < 10°. Grade III is a dysplastic hip with a migration percentage of 15% to 30%, Shenton’s arch broken by < 5 mm, a round or mildly flattened femoral head, a normal or mildly dysplastic acetabulum and pelvic obliquity of < 10°. Grade IV is a subluxated hip with a migration percentage of > 30% but < 100%, Shenton’s arch is broken by > 5 mm, there is femoral head and acetabulum variable deformity and variable pelvic obliquity. Grade V is a dislocated hip with a migration percentage of at least 100% with a completely disrupted Shenton’s arch, variable femoral head and acetabulum deformity and variable pelvic obliquity. Grade VI is a hip that has undergone salvage surgery.

In total, 65 AP hip radiographs were selected in a non-randomised fashion to ensure each classification group was represented. Each radiograph was classified into one of six grades by each surgeon (MWS, MB). Each radiograph was de-identified. Eight weeks after the initial evaluation, the same radiographs were then re-presented to the group in random order for the second evaluation. In general, optimal radiograph positioning is standard in the clinical settings studied here. However, the radiographs were not screened for ‘optimal’ positioning, specifically to allow for study of ‘real world’ radiographs.

Statistical analysis

This was performed using the intraclass correlation coefficient (ICC), for both intra- and interobserver reliability. The ICC is equivalent to the weighted kappa value in which less weight is assigned to agreement since categories are further apart and was chosen since the original authors used this system for reliability analysis. The ICC was interpreted using established conventions for kappa where < 0 is poor agreement, 0 to 0.2 is slight agreement, 0.2 to 0.4 is fair agreement, 0.4 to 0.6 is moderate agreement, 0.6 to 0.8 is substantial agreement and > 0.8 is excellent agreement.12 Statistical analysis was performed using SPSS Statistics version 22 (IBM Corp., Armonk, New York).

Results

Interobserver reliability

The ICC for interobserver reliability was 0.853 (0.813 to 0.887) for the first reading and 0.839 (0.795 to 0.877) for the second reading, demonstrating excellent interobserver reliability (Table 1). The data were further analysed to determine if there were any differences of greater than two classifications of an individual subject. In the first analysis, of the 130 hips analysed, four radiographs had a difference of greater than two classifications among the
**Fig. 1** Radiographs and diagrams showing The Melbourne Cerebral Palsy Hip Classification System (reproduced with permission from Robin J, Graham HK, Baker R, et al. A classification system for hip disease in cerebral palsy. *Dev Med Child Neurol* 2009;51:183–192).
AN INDEPENDENT ASSESSMENT OF RELIABILITY OF THE MELBOURNE CEREBRAL PALSY HIP CLASSIFICATION SYSTEM

Table 1. Details of interobserver reliability.

| Rater                  | ICC     | 95% CI     |
|------------------------|---------|------------|
| All raters, reading A  | 0.853   | 0.813 to 0.887 |
| All raters, reading B  | 0.839   | 0.795 to 0.877 |

ICC, intraclass correlation coefficient; CI, confidence interval.

five participants. In the second analysis, five radiographs had a difference of greater than two classifications among the five participants.

Intraobserver reliability

The ICC for intraobserver reliability shown in Table 2 was in the range of 0.838 to 0.933, indicating excellent intraobserver reliability. The data were assessed with a symmetry test to ensure that there was no bias between the first and second readings of the participants.

Discussion

MP is the most widely used radiographic measurement of hip disease in CP. There are advantages to this method including but not limited to its lack of subjective input, the relative ease of measurement, and the extensive use of the MP in the literature in relation to clinical outcomes of surgical intervention. \(^{13,14}\) Currently, in many paediatric orthopaedic centres, management decisions are commonly based on the MP. However, due to the 2D nature of radiographs, anterior and posterior dislocations do not always show a substantial increase in the migration percentage and is thus a limitation of the use of the MP. \(^{11}\) The use of CT imaging with 3D reconstruction and MRI has been proposed to circumvent this limitation. \(^{15,15}\) However, this may not be practical, cost-effective or safe due to high levels of exposure during CT imaging as a means of hip surveillance or clinical study in children. MRI does not provide as much bony detail and almost always requires sedation.

The continuous scale of the MP also has several limitations. Studies have shown that an experienced reader could replicate a MP measurement within 6% to 8%. \(^{3,5,7,8,14}\) In order to determine a ‘true’ change in MP, based on a 95% confidence interval, the change of MP would have to be > 13% from one radiograph to another, which is quite a large amount of change clinically. Furthermore, some providers feel the continuous scale of the MP makes communication of the overall sense of pathology more difficult, and potentially makes research more difficult, rather than using a discrete, ordinal system. \(^{10}\)

The Melbourne Cerebral Palsy Hip Classification System was developed as an ordinal classification system using both quantitative (MP) and qualitative input. \(^{16}\) The system was designed with the understanding that the development of severe displacement in combination with pelvic obliquity and/or deformity of the acetabulum and femoral head increases the difficulty of performing the measurement of MP. \(^{7}\) Finally, from a research perspective, this classification system provides ordinal data resulting in an optimal system for studying clinical outcomes.

The original study of intra- and inter-reliability of this classification system was designed to evaluate radiographs in adolescents with CP with an age range of 14 to 19 years. \(^{11}\) Murnaghan et al. \(^{11}\) performed the first study of the intra- and inter-reliability of the Melbourne Cerebral Palsy Hip Classification System, as part of the group from Melbourne that developed the system. Our study takes place in a different region (South Western United States) and is the first independent analysis of its reliability. Furthermore, there was no attempt to ensure perfect radiographs, but instead, this study investigated how this system would perform with ‘real world’ radiographs. We also chose to expand slightly the age range in this analysis compared with the original study from Melbourne, simply because our Grade V and VI hips often occurred in those slightly older patients still seen in our paediatric CP clinic.

An independent evaluation of this system’s reliability is important to further validate its use. Our study compared very favourably with that from Murnaghan et al. \(^{11}\) with very similar intra- and interobserver reliability. Overall, our data suggest excellent intra- and interobserver reliability in this patient population with the Melbourne Cerebral Palsy Hip Classification System.

Our data indicate that the most variance in rater’s classification occurs between classifications I and III. This also compares very favourably with that from Murnaghan et al. \(^{11}\) The biggest variations occurred in five patients in our study and in six patients in the series from Melbourne. We feel that these relatively few outliers do not negatively impact the usefulness of the classification system.

One of the criticisms of the system is that there are very few differences between a I and a II rating. In fact, along with a ‘near normal acetabulum’ and a ‘nearly round femoral head’, one of the main differences between a I and a II is a difference in MP of only 5%, which is within the error of accepted measurement of MP. We agree that, in practical terms, the differentiation between a I and a II can be difficult. The current numbers in this analysis, and therefore the power of the study, were not sufficient to report...
out separate statistics, as the number of patients in each subset would have limited proper statistical inference.

Another valid weakness of this analysis is the use of only paediatric orthopaedic surgeons as the raters, rather than a mix of healthcare providers, including physiotherapists. The original study by Murnaghan et al.\(^\text{11}\) did use physiotherapists as a subset of raters for their analysis. Our selection of only paediatric orthopaedic surgeons unfortunately reflects the difference in practice in the United States where our institution is located, where currently physiotherapists are not typically involved with independent evaluation of hip radiographs of patients with CP. However, the authors certainly admit that the quality of care in countries that have standardised CP hip surveillance programmes, which includes the practice of physiotherapists evaluating hip radiographs, is superior to that in our location. We believe that well-trained, qualified physiotherapists would likely give similar highly reliable ratings with this classification system.

Our study confirms that this classification system is valid and reliable in patients with CP and neuromuscular hip dysplasia. This classification system can be used as an effective tool for communication among providers and in further studies of prognosis and quality of life for patients with CP.

Received 07 February 2017; accepted after revision 02 August 2017.

**COMPLIANCE WITH ETHICAL STANDARDS**

**FUNDING STATEMENT**

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.

**ACKNOWLEDGEMENTS**

The authors wish to thank G. White, J. Karlen, W. Wood and E. Andrisevic for their assistance in this project.

**OA LICENCE TEXT**

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) licence (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed.

**ETHICAL STATEMENT**

**Ethical approval:** This work was approved by the Institutional Review Board (IRB) at Phoenix Children’s Hospital (PCH). All procedures performed in this study were in accordance with the ethical standards of PCH.

**Informed consent:** The IRB waived the need for informed consent due to the retrospective nature of this work.

**REFERENCES**

1. **Cornell MS.** The hip in cerebral palsy. *Dev Med Child Neurol* 1995;37:3-18.
2. **Root L.** Surgical treatment for hip pain in the adult cerebral palsy patient. *Dev Med Child Neurol* 2009;51:84-91.
3. **Soo B, Howard JJ, Boyd RN, et al.** Hip displacement in cerebral palsy. *J Bone Joint Surg [Am]* 2006;88-A:121-129.
4. **Hägglund G, Laue-Pedersen H, Wagner P.** Characteristics of children with hip displacement in cerebral palsy. *BMC Musculoskelet Disord* 2007;8:101.
5. **Connelly A, Flett P, Graham HK, Oates J.** Hip surveillance in Tasmanian children with cerebral palsy. *J Paediatr Child Health* 2009;45:437-443.
6. **Reimers J.** The stability of the hip in children: a radiological study of the results of muscle surgery in cerebral palsy. *Acta Orthop Scand Suppl* 1980;184:1-100.
7. **Parrott J, Boyd RN, Dobson F, Lancaster A, et al.** Hip displacement in spastic cerebral palsy: repeatability of radiologic measurement. *J Pediatr Orthop* 2002;22:660-667.
8. **Hägglund G, Laue-Pedersen H, Persson M.** Radiographic threshold values for hip screening in cerebral palsy. *J Child Orthop* 2007;1:43-47.
9. **Robin J, Graham HK, Baker R, et al.** A classification system for hip disease in cerebral palsy. *Dev Med Child Neurol* 2009;51:183-192.
10. **Chambers HG.** A classification for hip disease in cerebral palsy. *Dev Med Child Neurol* 2009;51:168-169.
11. **Murnaghan ML, Simpson P, Robin JG, et al.** The cerebral palsy hip classification is reliable. *J Bone Joint Surg [Br]* 2010;92-B:436-441.
12. **Nelson JC, Pepe MS.** Statistical description of interrater variability in ordinal ratings. *Stat Methods Medical Res* 2000;3:475-496.
13. **Taft LT.** Cerebral palsy. *Pediatr Rev* 1995;16:411-418.
14. **Dobson F, Boyd RN, Parrott J, Nattrass GR, Graham HK.** Hip surveillance in children with cerebral palsy. *J Bone Joint Surg [Br]* 2002;84-B:720-726.
15. **Kim HT, Wenger DR.** Location of acetabular deficiency and associated hip dislocation in neuromuscular hip dysplasia: three-dimensional computed tomographic analysis. *Journal Pediatr Orthop* 1997;17:143-151.
16. **Graham HK.** Classifying cerebral palsy. *J Pediatr Orthop* 2005;25:127-128.