OPEN

Interobserver and Intraobserver Reliability in the Salter Classification of Avascular Necrosis of the Femoral Head in Developmental Dysplasia of the Hip

Emily K. Schaeffer, PhD,† Ethan Ponton, BSc,*‡ Wudbhav N. Sankar, MD,§ Harry K.W. Kim, MD,‖¶ Simon P. Kelley, MBChB, FRCS(Tr&Ortho),# Peter J. Cundy, MBBS, FRACS,**†† Charles T. Price, MD, FAAP,‡‡ Nicholas M.P. Clarke, CHM, DM, FRCS,§§‖‖ John H. Wedge, OC, MD, FRCSC,# for International Hip Dysplasia (IHD) Study Group, and Kishore Mulpuri, MBBS, MS(Ortho), MHSc(Epi)*†*

Background: Avascular necrosis (AVN) of the femoral head is a concerning complication that can result from treatments for developmental dysplasia of the hip (DDH). AVN can lead to degenerative osteoarthritis, persistent acetabular dysplasia, reduced function, and continuing hip pain. The incidence of AVN reported in the DDH literature is widely varied (0% to 73%). This variability may arise from lack of consensus on what constitutes true AVN in this patient population, and lack of clear criteria provided in studies reporting incidence rates.

Methods: A multicentre, prospective database of infants diagnosed with DDH between 2010 and 2014 from 0 to 18 months of age was analyzed for patients treated by closed reduction (CR). Twelve pediatric orthopaedic surgeons completed 2 rounds of AVN assessments. Deidentified anteroposterior radiographs at most recent follow-up were provided to surgeons along with patient age at radiographic assessment, length of follow-up, ands affected hip. Ten of 12 surgeons completed a third round of assessments where they were provided with 1 to 2 additional radiographs within the follow-up period. Radiographic criteria for total AVN described by Salter and colleagues were used. Surgeons rated the presence of AVN as "yes" or "no" and kappa values were calculated within and between rounds.

Results: A total of 69 hips in 60 patients were assessed for AVN a median of 22 months (range: 12 to 36) post-CR. Interobserver kappa values for rounds 1, 2, and 3 were 0.52 (range: 0.11 to 0.90), 0.61 (range: 0.21 to 0.90), and 0.53 (range: 0.10 to 0.79), respectively. Intraobserver agreement for AVN diagnosis was an average of 0.72 (range: 0.31 to 0.96).

Conclusions: Despite using the most commonly referenced diagnostic criteria, radiographic diagnosis of AVN following CR in DDH patients demonstrated only moderate agreement across surgeons. The addition of sequential radiographs did not improve cross-observer reliability, and while substantial agreement was seen within observers, the range of intraobserver kappa values was large.

Level of Evidence: Level I—diagnostic study.

Key Words: developmental dysplasia of the hip, DDH, avascular necrosis of the femoral head, AVN, pediatrics, inter-rater reliability, intrarater reliability

(*J Pediatr Orthop* 2022;42:e59–e64)

From the *Department of Orthopaedics, University of British Columbia; †Department of Orthopaedic Surgery, BC Children's Hospital; ‡Office of Pediatric Surgical Evaluation and Innovation, BC Children's Hospital, University of British Columbia, Vancouver, BC; #The Hospital for Sick Children, Toronto, ON, Canada; §Division of Orthopaedics, The Children's Hospital of Philadelphia, Philadelphia, PA; ‖Center for Excellence in Hip Disorders, Texas Scottish Rite Hospital for Children, Dallas, TX; ¶Department of Orthopaedic Surgery, University of Texas Southwestern Medical Center, Dallas, TX; ‡‡University of Central Florida College of Medicine, Orlando, FL; **Centre for Orthopaedic and Trauma Research, The University of Adelaide; ††Department of Orthopaedic Surgery, Women's and Children's Hospital, Adelaide, SA, Australia; §§Department of Pediatric Orthopaedic Surgery, Southampton Children's Hospital; and ‖‖University of Southampton, Southampton, UK.

IHD Study Group: Kishore Mulpuri, Emily K. Schaeffer, Charles T. Price, Nicholas M.P. Clarke, Jose Herrera Soto, Travis Matheney, Pablo Castaneda, Simon P. Kelley, Nicole Williams, Young-Jo Kim, John H. Wedge, Alex Aarvold, Peter Cundy, Bruce Foster, Vidyadhar Upasani, Scott Mubarak, Unni Narayanan, Wudbhav N. Sankar.

Support for this study was provided by the International Hip Dysplasia Institute (IHDI), the Peterson Fund for Global Hip Health, and the I'm a HIPpy Foundation.

The authors declare no conflicts of interest.

Reprints: Kishore Mulpuri, MBBS, MS(Ortho), MHSc(Epi), Department of Orthopaedic Surgery, BC Children's Hospital, 1D.69-4480 Oak Street, Vancouver, BC, Canada V6H 3N1. E-mail: kmulpuri@cw.bc.ca.

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website, www.pedorthopaedics.com.

DOI: 10.1097/BPO.0000000000001979

The term developmental dysplasia of the hip (DDH) includes a spectrum of hip abnormalities which has an incidence of 1 to 34 per 1000 live births. It is the leading cause

of hip osteoarthritis in young adults.[1,2] Treatment for DDH depends both on the age of the patient as well as severity of dysplasia. Typically, abduction brace is first line treatment before 6 months of age, with closed reduction (CR) and spica casting or open reduction as more frequently used options in older or more complex patients.[1,3] Avascular necrosis (AVN) of the femoral head is a concerning complication that can result from treatment for DDH treatment.

Though the exact etiology of AVN is currently unknown, direct vascular compression and/or excessive pressure on the femoral head are thought to be the root causes,[4] which lead to lack of blood flow and cellular death. This can lead to proximal femoral growth disturbance and deformity, degenerative osteoarthritis, persistent acetabular dysplasia, reduced function, continuing hip pain, and potential need for an early hip replacement. At this point, it has not yet been elucidated why certain patients develop AVN and others do not. Many variables have been identified as risk or protective factors, such as age, prereduction traction, and CR versus open reduction; however, some of these have been found to increase risk in some studies and decrease risk in others.[5–12] Furthermore, some studies suggest that none of these factors are related to the development of AVN secondary to DDH treatment.[4,11,13–17] There is consensus, however, that patients who develop AVN are more likely to require subsequent surgeries and report poorer clinical outcomes.[11]

The incidence of AVN in the DDH literature is widely varied (0% to 73%).[4–6,11,13–16,18–20] While some of this variability is inherent to the different treatment methods used, much of it may also arise from lack of consensus on what constitutes true AVN in this patient population, and the lack of clear criteria provided in studies reporting incidence rates. Different AVN classification systems used include the Gage and Winter, Kalamchi and McEwen, Salter and colleagues, and Bucholz and Ogden.[7,9,21,22] The lack of clarity in the DDH literature has made cross-study comparison of this important treatment complication difficult and prevented the definitive identification of risk factors and prognostic factors for the development of AVN following treatment for DDH. There has yet to be a classification system for AVN validated in pediatric patients with DDH. The purpose of this study was to determine the interobserver and intraobserver reliability of the radiographic diagnosis (Salter criteria) of AVN following CR for DDH. The Salter classification was chosen as this method aims for early identification of ischemia that can lead to prolonged growth disturbance and deformity long term.

METHODS

This was a multicentre, international prospective study of patients diagnosed with DDH from 0 to 18 months of age and treated by CR between 2010 and 2014 from 7 centers in the United States, Canada, Australia, and the United Kingdom. Patients with a minimum 1-year follow-up post-CR were included for assessment of AVN at the most recent follow-up radiograph. This study was reviewed and approved by the Research Ethics Board at the University of British Columbia (H11-02191).

Pediatric orthopaedic surgeons with a practice focus on the pediatric hip and contributing to the multicentre prospective study were invited to participate in three rounds of AVN assessment. For round 1, they were all provided with deidentified anteroposterior radiographs from the most recent follow-up along with patient age at radiographic assessment, length of follow-up, and the affected hip. Two weeks after the first round, the radiographs were shuffled and sent back to the surgeons for the second round. Twelve pediatric orthopaedic surgeons—practicing at 2 Canadian centers, 5 US centers, 1 Australian center, and 1 UK center—completed the first 2 rounds of AVN assessments. Ten of the 12 surgeons completed a third round of assessment where they were provided with 1 to 2 additional radiographs within the follow-up period. All surgeons were sent the Salter and colleagues paper, outlining the AVN criteria (Table 1), with the file containing the hip radiographs. An example of the sequential radiograph series is provided in Figures 1A–C.

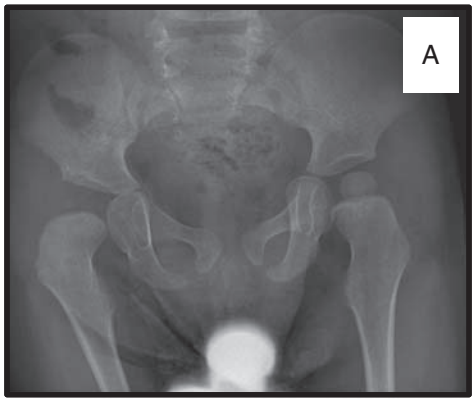
In their assessments, surgeons were instructed to use the radiographic criteria for AVN described by Salter. They rated the presence of AVN as "yes" or "no" and Cohen's Kappa values were calculated within and between rounds to determine interobserver and intraobserver reliability, respectively (Fig. 2). The kappa values were interpreted using the Landis and Koch guidelines, indicating that agreement with $\kappa < 0.00$ is poor, 0.00 to 0.20 is slight, 0.21 to 0.40 is fair, 0.41 to 0.60 is moderate, 0.61 to 0.80 is substantial, and 0.81 to 1.00 is almost perfect.[23] Kappa values are reported along with 95% confidence interval (CI).

RESULTS

A total of 69 hips in 60 patients were assessed for AVN, a median 22 months (range: 12 to 36) post-CR. Interobserver kappa values for round 1, 2, and 3 were 0.52 (95% CI: 0.48-0.57; range: 0.11 to 0.90), 0.61 (95% CI: 0.57-0.65; range: 0.21 to 0.90), and 0.53 (95% CI: 0.48-0.59; range: 0.10 to 0.79), respectively, where values of 0.41 to 0.60 are considered moderate agreement, and 0.61 to 0.80 are considered substantial agreement. Interobserver agreement did not remain constant between pairs over the 3 rounds. For example, surgeon 2 and 7 had the lowest agreement of all pairs in round 1 (kappa = 0.11; 61% agreement) and the highest in round 3 (kappa = 0.79;

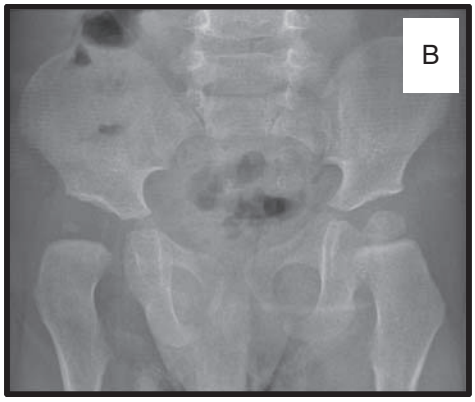
TABLE 1. Criteria for the Diagnosis of AVN[21]

Salter Criteria
1. Failure of appearance of the ossific nucleus of the femoral head during 1 y or longer after reduction
2. Failure of growth of an existing ossific nucleus during 1 year or longer after reduction
3. Broadening of the femoral neck within 1 year after reduction
4. Increased radiographic density of the femoral head followed by the radiographic appearance of fragmentation
5. Residual deformity of the femoral head and neck when reossification is complete. These deformities include coxa magna, coxa plana, coxa vara, and a short, broad femoral neck

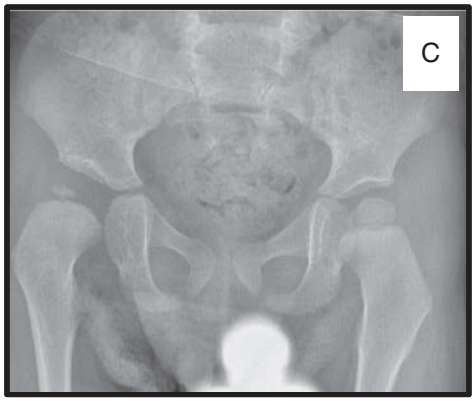


Affected Hip: RIGHT

Patient Age: 26 MONTHS
Post- CR: 12 MONTHS



Patient Age: 30 MONTHS
Post- CR: 17 MONTHS



Patient Age: 38 MONTHS
Post- CR: 24 MONTHS

AVN: Y or N?

FIGURE 1. Example cases—radiograph series provided for rating. In the file that was provided, the surgeons were given the sequential radiographs, the affected hip, the patient age in months, and the number of months post-closed reduction surgery for each radiograph in the series. They were asked to provide a yes or no answer with respect to determining the presence of AVN. AVN indicates avascular necrosis; CR, closed reduction.

93% agreement). All pairwise surgeon inter-rater kappa values are provided in Supplemental Table 1 (Supplemental Digital Content 1, http://links.lww.com/BPO/A415). Intraobserver agreement for AVN diagnosis was an average of 0.72 (95% CI: 0.61-0.83; range: 0.31 to 0.96). All surgeon intrarater reliability values are provided in Supplemental Table 2 (Supplemental Digital Content 1, http://links.lww.com/BPO/A415).

Examples of almost perfect agreement of no AVN, almost perfect agreement of yes AVN, and minimal agreement (50% split) are provided in Figures 3A–C, respectively.

DISCUSSION

The results of this study and the variability seen in the data demonstrate the need for the development of a comprehensive, standardized classification system for AVN specific to the DDH patient population. As of yet, no classification system has been validated for this population. We showed that, despite using the most commonly referenced diagnostic criteria, only moderate agreement was seen across specialized pediatric orthopaedic surgeons for DDH, and the addition of sequential radiographs during follow-up did not improve interob-

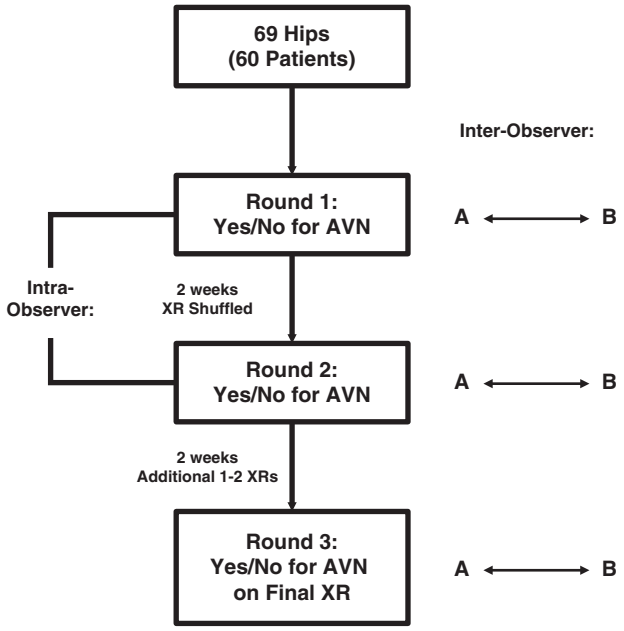


FIGURE 2. Flowchart of methods. A file with 69 radiographs from 60 patients, and their pertaining information, outlined in Figure 1, were sent to surgeons. Two weeks after the first round, they were sent the same file with the radiographs shuffled and asked to repeat the task. Finally, 10 of the 12 surgeons completed the third round 2 weeks later, where they were given 1 to 2 additional radiographs. In each round, interobserver kappa values were calculated, and intraobserver kappa values were calculated between rounds 1 and 2. AVN indicates avascular necrosis.

server reliability. Substantial intraobserver agreement was seen; however, the range of kappa values across surgeons was large, suggesting that surgeon experience and/or practice location may be a contributing factor. Discrepant agreement levels across cases demonstrate that diagnostic variability largely arises from borderline AVN cases.

Few studies have been conducted in the DDH population to determine which classification system is best for diagnosing AVN; however, 1 study looked at the reliability of the Bucholz and Ogden criteria, and found an inter-rater reliability of 0.34 for all raters, which indicates poorer agreement when compared with the Salter criteria, used in this study.[24] Important methodological differences do exist between studies to consider: the former study included surgeons, a musculoskeletal radiologist, and orthopaedic trainees, and we included exclusively pediatric orthopaedic surgeons as participants. In addition, the Bucholz and Ogden criteria are graded, and most of the disagreement was between grades I and II, whereas the Salter criteria is a yes or no diagnosis. Because of these differences, it is difficult to make direct comparisons, but the Salter criteria appears either to be more reliable in the DDH population, or simply provides a more general diagnosis. However, the agreement in the Salter criteria is still only moderate, leaving much room for improvement. Some of the existing variability may be reduced in the future through the development and integration of artificial intelligence technology that can automate and standardize the process. However, artificial intelligence still relies on human input and requires a "perfect case" comparison; consequently, there is an evident need for the development of a better objective classification system for AVN.

In addition, there is ongoing debate about potentially important differences between proximal femoral growth disturbance that may be transient, and true AVN. For example, O'Brien et al[25] suggest that AVN is not an appropriate term to describe the growth disturbances resulting from treatment for DDH. Instead, the authors propose that characterization of growth disturbance lines and the contour of the physis may be a more accurate means of identifying and describing this phenomenon. Keret and MacEwen[26] also suggest a modified approach to the Salter criteria that involves classifying growth disturbances into 3 groups: (1) mild vascular insufficiency; (2) partial necrosis or moderate vascular insufficiency; or (3) total necrosis of the femoral head or severe vascular insufficiency. These proposed modified or alternate classifications will be important to consider in the development of a standardized system for diagnosis.

There are a few limitations of this study. First, we did not control for respondents' years of experience or practice location, and this likely explains some of the observed variation. However, we included only specialized pediatric orthopaedic surgeons, all of whom had been practicing for >5 years with a specific practice focus on pediatric hip. Second, for the x-rays, we included only anteroposterior radiographs in this study because this view was available for all patients, but other views such as frog leg lateral were not consistently available. Third, we do not know the ultimate impact of a mild growth disturbance early after surgery. Some of these early, mild cases may not be clinically impactful to the patient. Finally, hips that are in the early stages of developing AVN may appear radiographically normal, despite the presence of hip pain, and therefore diagnosis may be delayed until discernable on plain radiograph, or through functional exploration of the bone, which involves a bone scan and investigation into medullary circulation hemodynamics.[27] Functional exploration is not routinely carried out, however, and follow-up radiographs are more clinically relevant for the DDH population. Ultimately, to confirm the accuracy of the Salter criteria in correctly diagnosing AVN, large long-term follow-up studies need to be conducted to determine both the accuracy of diagnosis and the clinical and functional impact. In addition, alternate diagnostic methods such as those proposed in O'Brien and colleagues and Karat and colleagues will be valuable to consider.

In conclusion, we found that the inter-rater reliability for the radiographic diagnosis of AVN secondary to DDH treatment "moderate," and while intrarater reliability was "substantial," there was a large range of kappa values. It is imperative that an accurate and reliable classification system for this patient population be developed so that we can

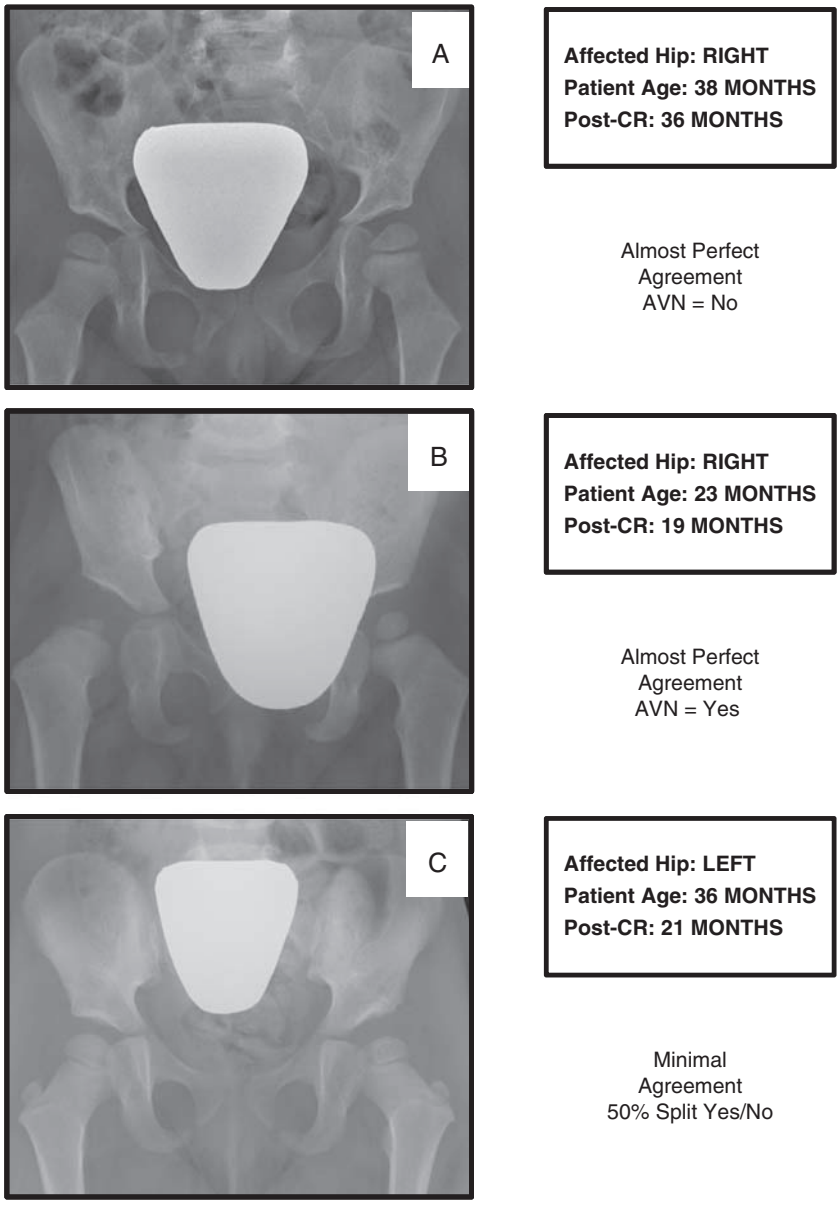


FIGURE 3. Example cases—levels of agreement between surgeons. In these examples, there was near perfect agreement between surgeons to the presence (A) or absence (B) of AVN in these children. In some radiographs, there was minimal agreement between surgeons, with 50% of surgeons answering yes and 50% answering no (C). AVN indicates avascular necrosis; CR, closed reduction.

begin to quantify the incidence of AVN as a treatment complication and identify risk factors and prognostic factors. Further studies are needed to differentiate between a temporary growth disturbance and true AVN. These will help direct diagnosis and further research to resolve the current controversies regarding protective and risk factors, ultimately improving clinical outcomes.

ACKNOWLEDGMENTS

The authors would like to thank the I'm a HIPpy Foundation, the Peterson Fund for Global Hip Health, the International Hip Dysplasia Institute, and the additional surgeons who participated in this study to complete the ratings: Dr Anthony Cooper, Dr Bruce Foster, Dr Jose Herrera Soto, Dr Travis Matheney, Dr Scott Mubarak, Dr Salil Upasani, and Dr Nicole Williams.

REFERENCES

1. Al-Essa RS, Aljahdali FH, Alkhilaiwi RM, et al. Diagnosis and treatment of developmental dysplasia of the hip: a current practice of paediatric orthopaedic surgeons. *J Orthop Surg (Hong Kong)*. 2017;25:1–7.
2. Noordin S, Umer M, Hafeez K, et al. Developmental dysplasia of the hip. *Orthop Rev (Pavia)*. 2010;2:e19–e24.
3. Kotlarsky P, Haber R, Bialik V, et al. Developmental dysplasia of the hip: what has changed in the last 20 years? *World J Orthop*. 2015;6:886–901.
4. Park K-B, Vaidya VN, Shin H, et al. Prereduction traction for the prevention of avascular necrosis before closed reduction for

developmental dysplasia of the hip: a meta-analysis. *Ther Clin Risk Manag*. 2018;14:1253–1260.
5. Al Faleh AF, Jawadi AH, Sayegh SA, et al. Avascular necrosis of the femoral head: Assessment following developmental dysplasia of the hip management. *Int J Health Sci (Qassim)*. 2020;14:20–23.
6. Domzalski M, Synder M. Avascular necrosis after surgical treatment for developmental dysplasia of the hip. *Int Orthop*. 2004;28:65–68.
7. Gage JR, Winter RB. Avascular necrosis of the capital femoral epiphysis as a complication of closed reduction of congenital dislocation of the hip: a critical review of twenty years' experience at Gillette children's. *J Bone Joint Surg*. 1972;54:373–388.
8. Gregosiewicz A, Wośko I. Risk factors of avascular necrosis in the treatment of congenital dislocation of the hip. *J Pediatr Orthop*. 1988; 8:17–19.
9. Kalamchi A, MacEwen GD. Avascular necrosis following treatment of congenital dislocation of the hip. *J Bone Joint Surg Am*. 1980;62:876–888.
10. Langenskiöld A, Paavilainen T. The effect of prereduction traction on the results of closed reduction of developmental dislocation of the hip. *J Pediatr Orthop*. 2000;20:471–474.
11. Schur MD, Lee C, Arkader A, et al. Risk factors for avascular necrosis after closed reduction for developmental dysplasia of the hip. *J Child Orthop*. 2016;10:185–192.
12. Wang Y-J, Yang F, Wu Q-J, et al. Association between open or closed reduction and avascular necrosis in developmental dysplasia of the hip: a PRISMA-compliant meta-analysis of observational studies. *Medicine*. 2016;95:e4276–e4284.
13. Luhmann SJ, Schoenecker PL, Anderson AM, et al. The prognostic importance of the ossific nucleus in the treatment of congenital dysplasia of the hip. *J Bone Joint Surg Am*. 1998;80:1719–1727.
14. Roposch A, Stöhr KK, Dobson M. The effect of the femoral head ossific nucleus in the treatment of developmental dysplasia of the hip. A meta-analysis. *J Bone Joint Surg Am*. 2009;91:911–918.
15. Segal LS, Boal DK, Borthwick L, et al. Avascular necrosis after treatment of DDH: the protective influence of the ossific nucleus. *J Pediatr Orthop*. 1999;19:177–184.
16. Novais EN, Hill MK, Carry PM, et al. Is age or surgical approach associated with osteonecrosis in patients with developmental dysplasia of the hip? A meta-analysis. *Clin Orthop Relat Res*. 2016; 474:1166–1177.
17. Chen C, Doyle S, Green D, et al. Presence of the ossific nucleus and risk of osteonecrosis in the treatment of developmental dysplasia of the hip: a meta-analysis of cohort and case-control studies. *J Bone Joint Surg*. 2017;99:760–767.
18. Bradley CS, Perry DC, Wedge JH, et al. Avascular necrosis following closed reduction for treatment of developmental dysplasia of the hip: a systematic review. *J Child Orthop*. 2016;10:627–632.
19. Pospischill R, Weninger J, Ganger R, et al. Does open reduction of the developmental dislocated hip increase the risk of osteonecrosis? *Clin Orthop Relat Res*. 2012;470:250–260.
20. Kruczynski J. Avascular necrosis of the proximal femur in developmental dislocation of the hip: Incidence, risk factors, sequelae and MR imaging for diagnosis and prognosis. *Acta Orthopaedica Scandinavica*. 1996;67:3–48.
21. Salter RB, Kostuik J, Dallas S. Avascular necrosis of the femoral head as a complication of treatment for congenital dislocation of the hip in young children: a clinical and experimental investigation. *Can J Surg*. 1969;12:44–61.
22. Bucholz RW, Ogden JA, eds. Patterns of ischemic necrosis of the proximal femur in nonoperatively treated congenital hip disease. *The Hip: Proceedings of the Sixth Open Scientific Meeting of the Hip Society*. St Louis, MO: Mosby; 1978:43–63.
23. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.
24. Roposch A, Wedge JH, Riedl G. Reliability of Bucholz and Ogden classification for osteonecrosis secondary to developmental dysplasia of the hip. *Clin Orthop Relat Res*. 2012;470:3499–3505.
25. O'Brien T, Millis MB, Griffin PP. The early identification and classification of growth disturbances of the proximal end of the femur. *J Bone Joint Surg*. 1986;68-A:970–980.
26. Keret D, MacEwen GD. Growth disturbance of the proximal part of the femur after treatment for congenital dislocation of the hip. *J Bone Joint Surg Am*. 1991;73-A:410–423.
27. Ficat RP. Idiopathic bone necrosis of the femoral head. Early diagnosis and treatment. *J Bone Joint Surg Br*. 1985;67:3–9.