Original Clinical Article

JOURNAL OF CHILDREN'S ORTHOPAEDICS

Reliability of the sourcil method of acetabular index measurement in developmental dysplasia of the hip

C. L. Maddock[1]
S. Noor[1]
A. Kothari[1]
C. S. Bradley[1,2]
S. P. Kelley[1,3]

Abstract

Purpose The ability to monitor and study developmental dysplasia of the hip (DDH) requires validated radiographic outcome measures. The sourcil method of acetabular index measurement (AI-S) has not yet been shown to be a reliable measure of acetabular dysplasia in a DDH population, despite its widespread use. The aims of this study were to test the reliability of the AI-S method in a DDH population, and to compare the reliability of the AI-S method with that of the classic lateral edge method (AI-L).

Methods From an institutional database, standardized anteroposterior hip radiographs were obtained from a cohort of 35 female patients (70 hips) at two and five years of age who had been treated nonoperatively for DDH. Three observers independently measured the acetabular index using the AI-L and AI-S methods on all 70 hips at two time points, four weeks apart.

Results The inter-rater reliability intraclass correlation coefficient (ICC) for the AI-L and AI-S methods was between good and excellent at 0.94 (confidence interval (CI) 0.89 to 0.96) and 0.91 (CI 0.87 to 0.94), respectively. The ICCs for intra-rater reliability for the AI-L method were excellent at 0.93 (CI 0.90 to 0.95), 0.95 (CI 0.93 to 0.97) and 0.95 (CI 0.94 to 0.97) for raters 1, 2 and 3, respectively. The ICCs for intra-rater reliability for the AI-S method were between good and excellent at 0.91 (CI 0.87 to 0.93), 0.93 (CI 0.90 to 0.95) and 0.90 (CI 0.86 to 0.93) for raters 1, 2 and 3 respectively.

Conclusion Both AI-S and AI-L methods are equally reliable radiographic measures of DDH.

[1]The Hospital for Sick Children, Toronto, Ontario, Canada
[2]Department of Rehabilitation Sciences, University of Toronto, Toronto, Ontario, Canada
[3]Department of Surgery, University of Toronto, Toronto, Ontario, Canada

Correspondence should be sent to Simon P. Kelley, 555 University Ave., Toronto, Ontario, Canada M5G 1X8.
E-mail: simon.kelley@sickkids.ca

Level of Evidence: Level III (diagnostic)

Cite this article: Maddock CL, Noor S, Kothari A, Bradley CS, Kelley SP. Reliability of the sourcil method of acetabular index measurement in developmental dysplasia of the hip. *J Child Orthop* 2019;13:167-171. DOI: 10.1302/1863-2548.13.190015

Keywords developmental dysplasia of the hip; reliability; acetabular index; sourcil; measurement

Introduction

Developmental dysplasia of the hip (DDH) is the most common paediatric musculoskeletal disorder with an incidence of approximately four in 1000 live births.[1,2]

If detected in the neonatal period, nonoperative treatment of DDH is considered to be the benchmark of management and is generally successful. Following completion of treatment, it is important to review children both clinically and radiographically to assess ongoing hip development for prognosis and clinical decision-making. During routine follow-up for DDH there are often no abnormal clinical signs to find on physical examination, therefore, the ability to monitor hip development and to study the outcomes of treatment for DDH depends upon the availability of valid and reliable radiographic outcome measures.

The acetabular index (AI) has long been utilized as one of the key radiographic parameters in the evaluation of acetabular morphology in DDH. It is considered useful both in research as an outcome measure and also clinically to monitor and guide treatment, such as to determine the need for surgical intervention.[3-6]

The method of AI measurement on hip radiographs and how it represents 3D acetabular morphology has been questioned by two differing approaches. The classic lateral edge method described by Hilgenreiner,[7] and clarified by Tönnis,[8] utilizes a measure of acetabular slope based on the most lateral aspect of the acetabulum (Fig. 1), and when increased above normal values quantifies anterolateral acetabular deficiency as demonstrated on advanced axial imaging.[9] An alternative method of calculating the acetabular index has been described by Kim et al,[9] which measures the acetabular slope based on the lateral edge of the sourcil (Fig. 1) and when increased above normal values quantifies deficiency of the mid-superior portion of the acetabulum as demonstrated on advanced axial imaging.[9,10]

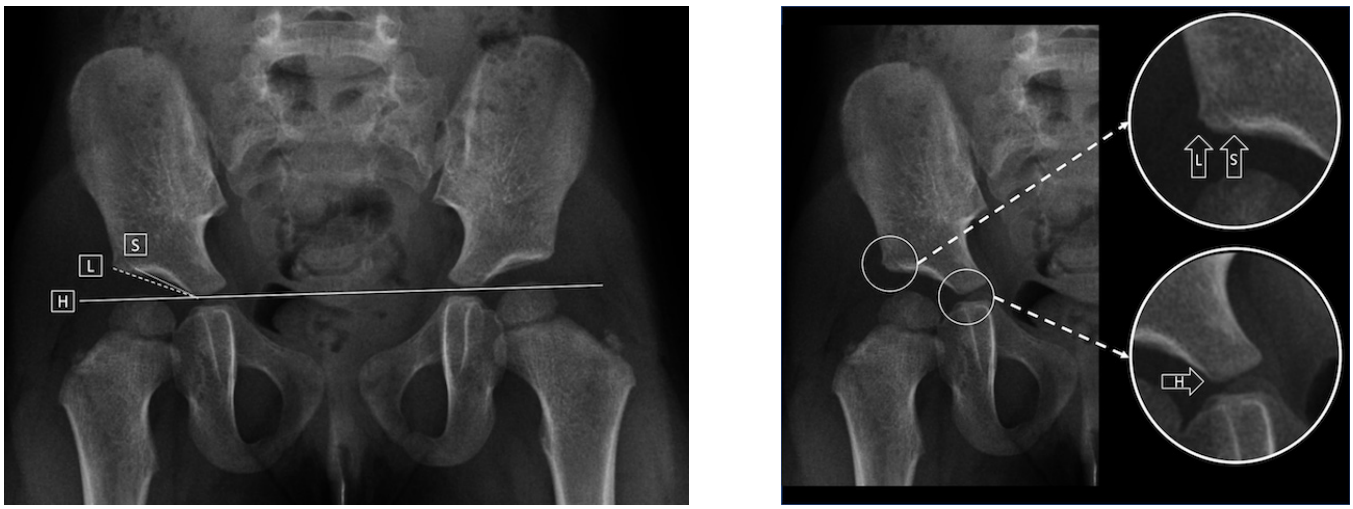

Fig. 1 Method to measure acetabular indices: (**a**) H denotes Hilgenreiner's line. S denotes line from which AI-S is measured. L denotes line from which AI-L is measured; (**b**) exploded view of important anatomical landmarks for defining lines used to measure acetabular indices. H denotes most inferior point of bony ilium at triradiate cartilage from which Hilgenreiner's line is drawn. S is the lateral edge of the sourcil from which AI-S is measured. L is the lateral edge of the acetabulum from which AI-L is measured.

On normal hip radiographs the lateral edge of the acetabulum and lateral edge of the sourcil are closely related if not contiguous, but yet in the dysplastic hip the lateral edge of the sourcil is often noted to be an irregular, short, interrupted shape with an upward slope, and ends short, medial and superior to the bony lateral edge of the acetabulum.[9,10] In the dysplastic hip the measurement of AI using the sourcil method or lateral edge method may, therefore, result in significantly different values depending on the method used.

It has been hypothesized that the sourcil method may be a better measure of acetabular dysplasia than the classic lateral edge method. The reason for this is that the sourcil represents the radiographic representation of the compressive stress in the hip and better correlates to the extent of the important weight-bearing zone of the roof of the acetabulum.[9-11] Alternatively, it is also possible that these two distinct methods of AI measurement are equally important as they contribute to a more detailed map of acetabular deficiency in DDH which may offer prognostic information or more specific guidance on the need for surgical intervention and the type of reconstructive osteotomy that may be required.

Regardless of the method used to measure AI, it must be reliable so that it can be trusted to be reproducible when repeated at random to allow consistent decision-making in treatment algorithms and to facilitate further research.

Unlike the Hilgenreiner lateral edge method of AI measurement, the sourcil method has not yet been shown to be a reliable measure of acetabular dysplasia in a DDH population, despite its use in clinical practice and reporting in research publications,[9-12] thus its utility as an outcome measure is still in question.

The research aims of this study were: 1) to test the reliability of the sourcil method of AI measurement in a DDH population, both by an experienced orthopaedic surgeon and surgical trainees; and 2) to compare the reliability of the sourcil method of AI measurement with that of the classic Hilgenreiner lateral edge method.

Materials and methods

Sample size for both inter-rater and intra-rater intraclass correlation coefficient (ICC) estimations was determined based on the accepted levels of $\alpha = 0.05$, $\beta = 0.2$, $\rho o = 0.7$ and $\rho 1 = 0.85$, where a minimum ICC of 0.7 is recommended for implementation in health research.[13-15] This calculation identified that a total of 70 pelvic radiographs were required for the study.

After obtaining institutional review board approval, a departmental infant hip clinic database was queried between the years 2012 to 2018 to produce a list of female patients who had undergone standardized Pavlik harness treatment for either unilateral or bilateral DDH and for whom follow-up anteroposterior pelvic radiographs at age two years (+/- 4 months) and age five years (+/- 4 months) were available on the hospital's picture archiving and communication system (PACS). Ages two and five years were selected in part because it is standard practice at our institution to image all previously treated children at these ages, but also provides initial measurements after walking age and at a time when acetabular development

JOURNAL OF CHILDREN'S ORTHOPAEDICS

decelerates. The diagnosis of DDH necessitating harness treatment had been made in each case using a combination of clinical and ultrasound examination of the hips.

Patients were excluded if there was subsequent evidence of a musculoskeletal diagnosis other than DDH, incomplete radiographic follow-up, radiographic follow-up outside the predetermined age range or abnormal radiographic orientation. Pelvic rotation and tilt were determined as described by Tönnis[8] and any radiograph showing excessive rotation was subsequently excluded. The resulting set of standardized anteroposterior radiographs consisted of 35 female patients from a DDH population with 70 hips for review at the two standardized time points.

Prior to radiographic measurement, three observers (one experienced orthopaedic surgeon (SPK) and two fellowship-trained orthopaedic surgeons (SN, AK)) jointly reviewed the published descriptions of acetabular index measurement for both the lateral edge and sourcil methods to confirm the exact radiographic landmarks used in this study. For purposes of clarity, consistency and comparison we have repeated the methods popularized by Tönnis,[8] Tönnis and Brunken,[16] Ogata et al[10] and Kim et al[9] and most recently evaluated in a population of normal children by Novais et al.[12]

In both methods, Hilgenreiner's line (Line H), is drawn connecting the inferior margins of the ilia at the level of the triradiate cartilages bilaterally. The lateral edge method is depicted in Figure 1 where Line L is drawn from the lateral most aspect of the acetabular roof to intersect with Line H, where the inferior margin of the ilium abuts the triradiate cartilage. The 'AI-L' is the angle that is formed between Lines H and L. Similarly, the sourcil method is also depicted in Figure 1, by drawing Line S from the lateral most aspect of the sourcil to Line H, where the inferior margin of the ilium abuts the triradiate cartilage. The 'AI-S' is the angle that is formed between Lines H and S.

Each observer independently measured the acetabular index using the AI-L and AI-S methods for 70 hips on 35 radiographs at two time points, four weeks apart (140 hips, 280 measurements). All measurements were conducted using line-drawing tools included in the PACS software. At the time of measuring, each observer was blinded to the measurements of the other observers and to the initial diagnosis with respect to the laterality and severity of the patients' DDH. This same procedure was repeated by each observer at each time point.

Statistical analysis

The statistical analyses were performed using SPSS software version 23 (SPSS Inc, Chicago, Illinois). ICC estimates and their 95% confidence intervals (CI) for inter-rater and intra-rater reliability for AI-L and AI-S measures were calculated based on a two-way random effects, absolute agreement, multiple raters/measurement model.[17,18] This model allows for subsequent generalizability to other situations and other raters. Inter-rater reliability was measured between raters using the first round of measurements. Intra-rater reliability was assessed for the same rater at both time points. ICC estimates were also calculated based on left or right hip between raters using the first set of measurements. Interpretation of ICC estimates was as follows: < 0.50, poor; between 0.50 and 0.75, fair; between 0.75 and 0.90 good; above 0.90, excellent.[19]

Results

Inter-rater reliability

The inter-rater reliability ICC for the AI-L method was between good and excellent at 0.94 (CI 0.89 to 0.96).

The inter-rater reliability ICC for the AI-S method was also between good and excellent at 0.91 (CI 0.87 to 0.94).

Intra-rater reliability

The ICCs for intra-rater reliability for the AI-L method was excellent at 0.93 (CI 0.90 to 0.95); 0.95 (0.93 to 0.97) and 0.95 (0.94 to 0.97) for raters 1, 2 and 3, respectively.

The ICCs for intra-rater reliability for the AI-S method was between good to excellent at 0.91 (CI 0.87 to 0.93); 0.93 (CI 0.90 to 0.95) and 0.90 (CI 0.86 to 0.93) for raters 1, 2 and 3, respectively.

Laterality

ICC estimates for either AI-L or AI-S methods for right and left hips, respectively, were between good to excellent with the lowest lower limit of the 95% CI at 0.84 and the highest upper limit of the 95% CI at 0.97 across all measures.

The inter- and intra-rater reliability ICC values are presented in Tables 1 and 2, respectively.

Discussion

The acetabular index is a very common, and probably the most widely used, radiographic measure of hip dysplasia in routine clinical use. It can give a snapshot of the severity of dysplasia as well as being a useful marker of hip development over time. The classic lateral-edge method (AI-L) is easy to measure and has been shown to be reliable. More recently an alternative 'sourcil method' of acetabular index measurement (AI-S) has been popularized, but to date has not been shown to be reliable in a DDH population, for which it was designed.

The purpose of this study was, therefore, to determine the reliability of the sourcil method of AI measurement in a population of children with DDH, and also to compare its reliability with the classic lateral edge method. For completeness, we included patients with radiographs at

Table 1 Inter-rater reliability values

Laterality	Method	ICC value	95% confidence interval
Both hips	AI-L	0.94	0.89 to 0.96
	AI-S	0.91	0.87 to 0.94
Left hip	AI-L	0.94	0.90 to 0.97
	AI-S	0.92	0.87 to 0.95
Right hip	AI-L	0.93	0.87 to 0.96
	AI-S	0.90	0.84 to 0.94

ICC, intraclass correlation coefficient; AI-L, lateral edge; AI-S, sourcil

Table 2 Intra-rater reliability values

	Method	ICC value	95% confidence interval
Rater 1	AI-L	0.93	0.90 to 0.95
	AI-S	0.91	0.87 to 0.93
Rater 2	AI-L	0.95	0.93 to 0.97
	AI-S	0.93	0.90 to 0.95
Rater 3	AI-L	0.95	0.94 to 0.97
	AI-S	0.90	0.86 to 0.93

ICC, intraclass correlation coefficient; AI-L, lateral edge; AI-S, sourcil

two and five years of age which we believe to be representative of a range of ages where acetabular index is most useful in clinical practice for decision-making. The sourcil has long been known as an excellent indicator of hip development, reflecting a reaction to the local stress within the hip, and is frequently abnormal in children with hip dysplasia. The edge of the sourcil has been determined to correlate to the mid-superior portion of the acetabulum, thus being an important metric of femoral head coverage and extent of acetabular dysplasia. Increasing numbers of publications have suggested that measurement of the AI, and the centre-edge angle (CEA), both well-known radiographic outcome measures of hip dysplasia, are better made to the edge of the sourcil compared with the traditional method using the lateral-most edge of the acetabular bone.[9-11]

We found that the AI-S method of acetabular index measurement in children with DDH showed good to excellent inter-rater reliability and intra-rater reliability, with equally good ICC values whether performed by an experienced orthopaedic surgeon or surgical trainees. Furthermore, the measure of AI-S is equally reliable for both right and left hips. In addition, we have shown that both AI-S and AI-L are equally reliable measures of acetabular dysplasia.

Previous measures of reliability of acetabular index (AI-L) have been somewhat variable from fair to excellent.[9,12,20] Our study has demonstrated good to excellent reliability between the measures. We believe that our reliability results are higher than some previous studies due to two factors. Firstly, prior to measurements being taken, the three observers met to discuss in detail the exact anatomical landmarks which would be used to make the measurements of both AI-L and AI-S. Secondly, the published descriptions of how to best measure these radiographic indices are becoming clearer and more detailed, which one would assume leads to more accurate, precise and reliable measurements.

The sourcil was shown to be an important feature on which to base radiographic measures of hip dysplasia by Ogata et al.[10] who demonstrated that some children with DDH, and marked acetabular deficiency (as identified on CT scans), could achieve normal radiographic CEA values based on the traditional method of measurement to the lateral edge of the acetabulum. Ogata et al[10] showed that by measuring the CEA to the lateral edge of the sourcil instead resulted in more accurate characterization of hip dysplasia that better correlated to acetabular coverage on CT scans.

Kim et al[9] subsequently showed similar findings when evaluating the AI and used axial imaging on CT and MRI to demonstrate that the lateral edge of the acetabulum on an anteroposterior radiograph corresponded to anterolateral coverage of the hip, while the lateral edge of the sourcil corresponded to the lateral margin of the mid-superior portion of the acetabulum. Kim et al[9] suggested that measuring the AI to the edge of the sourcil would better characterize acetabular deficiency and one should be clear as to which measure of AI was being used to prevent errors in interpretation of the nature of DDH in each particular patient.

Despite multiple publications recommending the use of the AI-S method and its increasing use in clinical practice, no normal values for this measure were available until recently. Using normal reference values of AI-L from Tönnis and Brunken's[16] classic work with clinical measures of AI-S would have likely led to misclassification, especially in a DDH population, where the edge of the sourcil and true lateral bony edge of the acetabulum are markedly separated on an anteroposterior hip radiograph. Novais et al[12] have recently published a complete range of normal values and reference curves for the measurement of AI-S in a population of normal children, with good reliability of the measure. They also demonstrated that these normal values differed from normal values published by Tönnis and Brunken,[16] and thus reiterated the point that one must be consistent in the use of AI measurements, specifying the method used (sourcil or lateral edge) and relating the findings to their newly published reference data.

Limitations of our study include the fact that we have not addressed some further important characteristics of the AI-L and/or AI-S methods. For example, it is not yet known whether AI-L and AI-S measurements are correlated, show agreement and may be used interchangeably to describe hip dysplasia in general or alternatively complement each other by measuring different aspects of dysplasia. Therefore, it is not clear whether the AI-L or AI-S have equivalent clinical utility in monitoring progress of hip development in DDH, are equivalent predictors for further surgery or most importantly which one best correlates with symptomatic dysplasia at skeletal maturity. All of these factors will need to be addressed in future studies.

Until further investigation is complete, we continue to recommend clarification in the medical chart which method of AI measurement is being used, that comparison with appropriate normal reference data can be made and consistency in repeated measurements is assured for purposes of monitoring clinical progress in DDH.

In conclusion, we have found that the sourcil method of acetabular index measurement (AI-S) in children with DDH showed good to excellent reliability, whether performed by an experienced orthopaedic surgeon or surgical trainees. Furthermore, the measure of AI-S is equally reliable for both right and left hips. In addition, we have shown that both sourcil (AI-S) and lateral edge (AI-L) methods are equally reliable measures of acetabular dysplasia. Clinicians must be aware, however, that clinical equivalence has yet to be proven and until then clear documentation of which method is used and comparison with their specific normative data must be assured.

Received 31 January 2019; accepted after revision 28 February 2019.

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.

ETHICAL STATEMENT

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent: Informed consent was obtained from all individual participants included in the study.

ICMJE CONFLICT OF INTEREST STATEMENT

All authors declare that they have no conflict of interest.

AUTHOR CONTRIBUTIONS

CLM: study design; data acquisition; data analysis and interpretation; manuscript drafting; manuscript revision.

SN: data acquisition; data analysis and interpretation; manuscript drafting; manuscript revision.

AP: data acquisition; data analysis and interpretation; manuscript drafting; manuscript revision.

CSB: study design; data analysis and interpretation; manuscript drafting; manuscript revision.

SPK: study design; data acquisition; data analysis and interpretation; manuscript drafting; manuscript revision.

REFERENCES

1. **Kocher MS.** Ultrasonographic screening for developmental dysplasia of the hip: an epidemiologic analysis (Part I). *Am J Orthop (Belle Mead NJ)* 2000;29:929-933.

2. **Woodacre T, Ball T, Cox P.** Epidemiology of developmental dysplasia of the hip within the UK: refining the risk factors. *J Child Orthop* 2016;10:633-642.

3. **Omeroğlu H, Ağuş H, Biçimoğlu A, Tümer Y.** Evaluation of experienced surgeons' decisions regarding the need for secondary surgery in developmental dysplasia of the hip. *J Pediatr Orthop* 2012;32:58-63.

4. **Grissom L, Harcke HT, Thacker M.** Imaging in the surgical management of developmental dislocation of the hip. *Clin Orthop Relat Res* 2008;466:791-801.

5. **Klein C, Fontanarosa A, Khouri N, et al.** Anterior and lateral overcoverage after triple pelvic osteotomy in childhood for developmental dislocation of the hip with acetabular dysplasia: Frequency, features, and medium-term clinical impact. *Orthop Traumatol Surg Res* 2018;104:383-387.

6. **Ezirmik N, Yildiz K.** Advantages of single-stage surgical treatment with salter innominate osteotomy and Pemberton pericapsular osteotomy for developmental dysplasia of both hips. *J Int Med Res* 2012;40:748-755.

7. **Hilgenreiner H.** Translation: Hilgenreiner on congenital hip dislocation. *J Pediatr Orthop* 1986;6:202-214.

8. **Tönnis D.** Normal values of the hip joint for the evaluation of X-rays in children and adults. *Clin Orthop Relat Res* 1976;119:39-47.

9. **Kim HT, Kim JI, Yoo CI.** Diagnosing childhood acetabular dysplasia using the lateral margin of the sourcil. *J Pediatr Orthop* 2000;20:709-717.

10. **Ogata S, Moriya H, Tsuchiya K, et al.** Acetabular cover in congenital dislocation of the hip. *J Bone Joint Surg [Br]* 1990;72-B:190-196.

11. **Omeroğlu H, Ağuş H, Biçimoğlu A, Tümer Y.** Analysis of a radiographic assessment method of acetabular cover in developmental dysplasia of the hip. *Arch Orthop Trauma Surg* 2002;122:334-337.

12. **Novais EN, Pan Z, Autruong PT, Meyers ML, Chang FM.** Normal percentile reference curves and correlation of acetabular index and acetabular depth ratio in children. *J Pediatr Orthop* 2018;38:163-169.

13. **Portney L, Watkins M.** *Foundations of clinical research: application to practice*. 2nd ed. New Jersey: Prentice Hall Health, 2000.

14. **Cohen J.** *Statistical power analysis for the behavioral sciences*. 2nd ed. New Jersey: Lawrence Erlbaum Associates, 1977.

15. **Walter SD, Eliasziw M, Donner A.** Sample size and optimal designs for reliability studies. *Stat Med* 1998;17:101-110.

16. **Tönnis D, Brunken D.** Differentiation of normal and pathological acetabular roof angle in the diagnosis of hip dysplasia. Evaluation of 2294 acetabular roof angles of hip joints in children. *Arch Orthop Unfallchir* 1968;64:197-228.

17. **McGraw K, Wong S.** Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1996;1:30-46.

18. **Shrout PE, Fleiss JL.** Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86:420-428.

19. **Koo TK, Li MY.** A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;15:155-163.

20. **Upasani VV, Bomar JD, Parikh G, Hosalkar H.** Reliability of plain radiographic parameters for developmental dysplasia of the hip in children. *J Child Orthop* 2012;6:173-176.