European Spine Journal
https://doi.org/10.1007/s00586-018-5821-8

ORIGINAL ARTICLE

The "Risser+" grade: a new grading system to classify skeletal maturity in idiopathic scoliosis

Michael J. Troy[1] · Patricia E. Miller[1] · Nigel Price[2] · Vish Talwalkar[3] · Fabio Zaina[4] · Sabrina Donzelli[4] · Stefano Negrini[4] · M. Timothy Hresko[1]

Received: 19 March 2018 / Accepted: 4 November 2018

Abstract

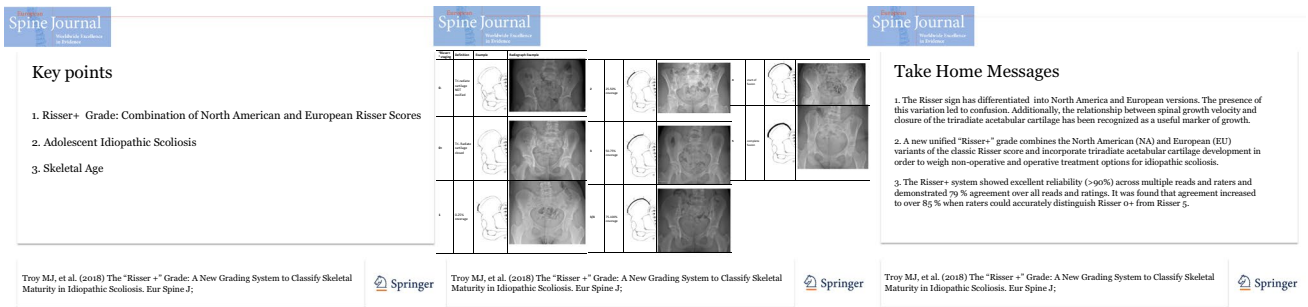
Purpose This study aims to propose and validate a new unified "Risser+" grade that combines the North American (NA) and European (EU) variants of the classic Risser score. The "Risser+" grade can effectively combine the North American and European Risser Classifications for skeletal maturity with adequate intra-rater/inter-rater reliability and agreement.

Methods Agreement and reliability were evaluated for 6 raters (3-NA, 3-EU) who assessed 120 pelvic radiographs from the BrAIST trial, all female, average age 13.4 (range 10.1–16.5 years). Blinded raters reviewed x-rays at two time-points. Intra- and inter-rater agreement (RA) were established with Krippendorff's alpha (*k*-alpha), while intra- and inter-rater reliability (RR) were established with intraclass correlation coefficients (ICC). Acceptable agreement and reliability were set a priori at 0.80.

Results Inter-RA for the second reading met study requirements (*k*-alpha = 0.86 [0.81–0.90]) compared to the first reading (0.72 [0.63–0.79]) while combined readings was close to target agreement (0.79 [0.74–0.84]). Removal of 20 readings demonstrating outlier tendencies increased agreement for the first, second, and combined reads (*k*-alpha = 0.85, 0.89, 0.87, respectively). Intra-RA was sufficient for 4 out of 6 raters (*k*-alpha > 0.80) and one rater from EU and NA presented subpar intra-RA (*k*-alpha = 0.64 and 0.74, respectively). Inter-RR met study requirements overall reads (ICC = 0.96 [0.95–0.97]) including the first (0.94 [0.92–0.95]) and second (0.97 [0.97–0.98]) reads, independently.

Conclusions The Risser+ system showed excellent reliability across multiple reads and raters and demonstrated 79% agreement overall reads and ratings. Agreement increased to over 85% when raters could distinguish Risser 0 + from Risser 5.

Graphical abstract These slides can be retrieved from electronic supplementary material.



Keywords Risser · Scoliosis · Idiopathic scoliosis · Skeletal age

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00586-018-5821-8) contains supplementary material, which is available to authorized users.

Extended author information available on the last page of the article

Published online: 16 November 2018

Introduction

Growth and development are imperative in assessment and treatment of children with idiopathic scoliosis. The measures of skeletal maturity have been developed based on physical development as well as the radiographic measurements of bone maturation in the hand or pelvis. The Risser sign has long been used as a surrogate for spine growth. Risser described the pattern of ossification of the iliac apophysis and subsequent fusion of the apophysis to the ilium [1]. He described 5 phases of apophyseal development which was subsequently validated [2]. The Risser sign has also been proven to correlate with skeletal age assessment [2]. However, the accuracy of the system has been contested depending on the use of anteroposterior or posterior–anterior, oblique, or lateral radiographs. Alternative skeletal systems using the elbow, calcaneus, and hip growth centers have been described [3–6]. Sanders et al. also developed a prognostic skeletal maturity scoring system demonstrating stronger correlation with curve progression, yet requires a modest learning curve on behalf of the clinician [7]. Independent of criticism, the Risser sign remains a clinical gold standard for growth evaluation, and it is used in the main indications for clinical practice [8–10]. The Risser sign has differentiated into North America (NA) and European (EU) versions [11]. The presence of this variation led to confusion in the medical literature as the NA Risser divides the ossification of the iliac wing into quarters (Risser stages 1, 2, 3, 4) and complete apophysis fusion to the iliac wing as Risser 5. Yet in the EU version, the ossification is broken into thirds, as Risser stage 4 denotes the apophysis beginning to fuse to the iliac wing, and stage 5 represents the conclusion of the fusion to the iliac wing [12, 13]. Additionally, the relationship between spinal growth velocity and closure of the triradiate acetabular cartilage has been recognized as a useful marker of growth, particularly at the worst clinical period corresponding to Risser 0. Nault et al. [14] added an assessment of the triradiate cartilage maturation to the Risser system in an attempt to add an estimate peak height velocity into the equation. The concept has been well accepted; however, a formally revised Risser classification has yet to be introduced.

In 2014, the SOSORT and the SRS nonoperative committee published a consensus statement on research in AIS [10]. The statement included the need for radiographic documentation of curve magnitude of research subjects and growth status based on a newly defined "Risser+ system" (Table 1) which combines the Nault proposal [14], the NA, and EU Risser classification. Of note, the Risser+ system divides the ossification of the iliac crest into quarters, with a new grade of "3/4" to denote 75–100% ossification. The SOSORT and SRS nonoperative committee recommends the clinical use of the Risser+ staging system, but acknowledge the need for repeatability studies for validation.

The primary purpose of this study aimed to validate and establish the reliability and agreement of the Risser+ system as it is recognized by experienced scoliosis clinicians from NA and EU. We hypothesized that the Risser+ system would exhibit acceptable reliability indicating that raters could consistently distinguish between Risser stages and that the Risser+ grading system would exhibit acceptable agreement indicating its reproducibility across multiple international raters.

Methods

The Risser+ system is an 8 point system which combines the NA and EU variations into a common system while also assessing the triradiate cartilage (TRC) maturity. The Risser+ system consists of grade 0-(open TRC), 0 + (Closed TRC), 1, 2, 3, 3/4, 4 and 5.

Posteroanterior (PA) spinal radiographs including the pelvis and femoral head were obtained from the BrAIST study group. The BrAIST trial collected images throughout the course of treatment during various stages of development, with a radiographic panel assigning Risser scores. Two hundred images with an even distribution of Risser stages 0–4 were screened for appropriate framing to include the femoral head and overall quality. Of these, 125 radiographs were de-identified, randomized, and entered into a REDcap (Research Electronic Data Capture) database. Three NA and three EU physicians experienced in treatment of patients with scoliosis were asked to grade each of the 125 images using the Risser+ model (Table 1). After an average time span of 2 weeks, each rater repeated the procedure. After both rounds of rating, 5 radiographs were deemed unreadable by a majority of the raters due to image quality and were subsequently excluded from analysis. The final cohort included 120 patient radiographs, was entirely female with an average age of 13.4 years (range 10.1–16.5 years) (Table 2), and included a distribution of Risser 0, 29%; Risser 1, 15%; Risser 2, 13%; Risser 3, 13%; Risser 4, 30%. Seventeen subjects (14%) were included with open triradiate cartilage based on the BRAIST study imaging assessment.

Statistical analysis

Primary analysis was to assess the reproducibility and efficiency of the Risser+ system as determined by inter-rater and intra-rater reliability and agreement. Reliability represents the ability to distinguish different stages or subjects from one another despite measurement errors, while agreement

Table 1 Visual representation of the Risser+ system

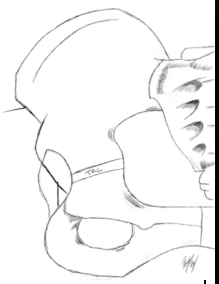
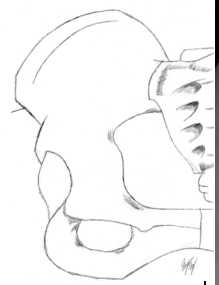
"Risser+" staging	Definition	Example	Radiograph example
0–	Tri-radiate cartilage NOT ossified		
0+	Tri-Radiate cartilage closed		
1	0–25% coverage		
2	25–50% coverage		

Table 1 (continued)

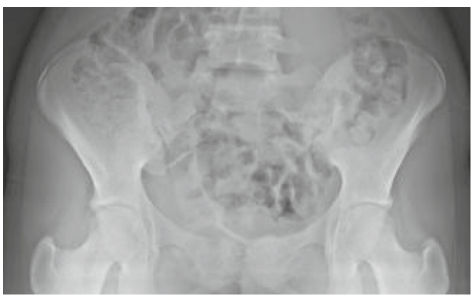
3	50–75% coverage		
3/4	75–100% coverage		
4	Start of fusion		
5	Complete fusion		

Table 2 Cohort characteristics ($N = 120$)

Characteristic	Freq.	(%)
Age (years; mean ± SD)	13.4	± 1.48
BMI percentile	55.5	± 25.65
Open triradiate	17	(14%)
Menarche	87	(73%)
Expected 5 point Risser grade		
0	35	(29%)
1	18	(15%)
2	16	(13%)
3	15	(13%)
4	36	(30%)

assesses the ability of raters to correctly identify the same stages multiple times.

Inter- and intra-rater reliability were quantified using intraclass correlation coefficients (ICC) along with 95% confidence. For inter-rater reliability, an ICC (2, 6) model was used and for intra-rater reliability an ICC (2, 1) model was used. Due to 9 missing ratings, 5 radiographs were excluded from all ICC calculations. Inter- and Intra-rater agreement were quantified using Krippendorff's alpha (*k*-alpha) along with boot-strapped 95% confidence limits. Boot-strapping was based on 1000 replications.

Reliability and agreement statistics were estimated for the first and second readings separately, and for all readings collectively. To conclude that the Risser+ grading scale is a satisfactory tool for medical diagnostics we required that the agreement and reliability coefficients reach a minimum threshold of 0.80.

Secondary analysis included comparing reliability and agreement estimates between EU and NA raters. In addition, we compared the agreement of ratings Risser ≤ 2 versus Risser > 2 in order to assess the clinical implications when considering whether a patient is at high risk of progression (Risser ≤ 2). Similarly, we assessed whether clinicians agreed on Risser grades for determining skeletal maturity and discontinuation of bracing: Risser ≥ 4 versus Risser < 4. Lastly, we estimated agreement between 0− and 0+ categories to assess clinicians' ability to distinguish between open and closed triradiate cartilage.

Sample size considerations

Power analysis determined that in order to detect a minimum ICC of 0.80 using a one-sided test compared to a null ICC of 0.70, we would require a minimum of 58 subjects with 6 raters to achieve 80% power. For the secondary aims of the study, we would require a minimum of 80 subjects to estimate a similar ICC with 3 raters to achieve 80% power. Because Krippendorff's alpha is estimated through boot-strapping, its accuracy is not dependent on sample size [15]. Our sample of 120 radiographs, that spanned a full distribution of Risser grades, was adequate to achieve the primary and secondary goals of this study.

Table 3 Inter- and intra-rater reliability for all raters and by continent for the 8-point Risser grading scale

	Number of ratings	*ICC*	95% CI
Inter-rater reliability			
All raters			
First reading	115	0.94	(0.92–0.95)
Second reading	115	0.97	(0.97–0.98)
Overall readings	230	0.96	(0.95–0.97)
North America			
First reading	115	0.85	(0.80–0.89)
Second reading	115	0.95	(0.94–0.97)
Overall readings	230	0.91	(0.88–0.93)
Europe			
First reading	115	0.91	(0.88–0.94)
Second reading	115	0.95	(0.94–0.97)
Overall readings	230	0.93	(0.91–0.95)
Intra-rater reliability			
North America			
Rater #1	115	0.86	(0.80–0.90)
Rater #2	115	0.90	(0.86–0.93)
Rater #3	115	0.66	(0.54–0.75)
Europe			
Rater #4	115	0.91	(0.87–0.94)
Rater #5	115	0.94	(0.92–0.96)
Rater #6	115	0.72	(0.62–0.80)

Results

Reliability

Inter-rater reliability of the 8-point scale sufficiently met the requirements of the study overall reads (ICC = 0.96; 95%CI = 0.95–0.97) as well as for the first (ICC = 0.94; 0.92–0.95) and second (ICC = 0.97; 0.97–0.98) reads independently (Table 3). EU raters exhibited slightly higher inter-rater reliability for the first reading (EU: 0.91 vs NA: 0.85) but comparable reliability for the second reading (ICC = 0.95) (Table 3). Intra-rater reliability was sufficient for 4 out of the 6 raters in the study (all ICC > 0.8) (Table 3). However, one rater from each of EU and NA presented subpar intra-rater reliability (ICC = 0.66 and 0.72, respectively).

Agreement

Inter-rater agreement of the 8-point scale overall reads and raters was substantial with *k*-alpha = 0.79 (95%CI = 0.74–0.84) but did not meet the requirements of the study set at 0.80. The second reading provided significantly higher agreement (*k*-alpha = 0.86; 0.81–0.90) compared to the first (*k*-alpha = 0.72; 0.63–0.79) (Table 4). EU raters exhibited slightly better agreement than NA raters for both the first (EU: 0.78 vs NA: 0.66) and second readings (0.88 vs 0.87) (Table 4). Intra-rater agreement was sufficient for 4 out of the 6 raters in the study (all *k*-alpha > 0.80) (Table 4). However, one rater from each of EU and NA presented subpar intra-rater agreement (*k*-alpha = 0.64 and 0.74, respectively).

Agreement for indication of high risk of progression (Risser ≤ 2 versus Risser > 2) was substantial for the first (*k*-alpha = 0.64; 95% CI = 0.54–0.71), second (*k*-alpha = 0.74; 95%CI = 0.66–0.82, and across all reads (*k*-alpha = 0.69; 95%CI = 0.63–0.75). Conversely, agreement with respect to discontinuing bracing Risser ≥ 4 versus Risser < 4 was insufficient for the first (*l*-alpha = 0.50; 95%CI = 0.39–0.59), the second (*k*-alpha = 0.55; 95%CI = 0.45–0.64), and across all readings (*k*-alpha = 0.52; 95% CI = 0.45–0.59). Rater agreement between Risser 0 − versus Risser 0 + was poor for the first (*k*-alpha = 0.56; CI = 0.28–0.76), second (*k*-alpha = 0.53; CI = 0.29–0.71), and across all reads (*k*-alpha = 0.55; CI = 0.37–0.69).

Table 4 Inter- and intra-rater agreement for all raters and by continent for the 8-point Risser grading scale

	Number of ratings	*K*-alpha	95% CI
Inter-rater agreement			
All raters			
First reading	120	0.72	(0.63–0.79)
Second reading	120	0.86	(0.81–0.90)
Overall readings	240	0.79	(0.74–0.84)
North America			
First reading	120	0.66	(0.52–0.77)
Second reading	120	0.87	(0.80–0.91)
Overall readings	240	0.77	(0.68–0.83)
Europe			
First reading	120	0.78	(0.69–0.85)
Second reading	120	0.88	(0.81–0.93)
Overall readings	240	0.83	(0.76–0.87)
Intra-rater agreement			
North America			
Rater #1	120	0.87	(0.76–0.94)
Rater #2	120	0.90	(0.82–0.96)
Rater #3	120	0.64	(0.44–0.78)
Europe			
Rater #4	120	0.91	(0.86–0.95)
Rater #5	120	0.94	(0.91–0.96)
Rater #6	120	0.74	(0.59–0.85)

Discussion

The Risser scale is internationally recognized as a clinically significant measure of skeletal maturity. The Scoliosis Research Society mandates that acceptable radiographs show C7 to the femoral heads in order to sufficiently frame the triradiate cartilage [16], thus making the Risser+ sign consistently clinically available. Although the Sanders score can be considered more indicative of curve progression, it requires a greater learning curve and a left-hand radiograph when compared to the Risser score available on any adequate anteroposterior radiograph. However, there has been controversy around the accuracy of the Risser sign [17] and the disparity between the NA and EU versions. Bitan et al. [12] and Nault et al. [14] demonstrated that the existence of two versions of the Risser scale might cause confusion for clinicians or international collaboratives. The aim of this study was to assess the reproducibility of the Risser+ system based on inter- and intra-rater reliability and agreement. Reliability measures the ability of raters to distinguish different stages from one another despite measurement errors, while agreement measures the ability of raters to correctly identify the same measures multiple times. The Risser sign bears limitations as noted by Shuren et al. [18], Zaoussis and James [19] and even Risser [1] due to atypical iliac apophysis (abnormal ossification, shortened excursion, fragmented ossification). Izumi [17] demonstrated that radiographs between posteroanterior and anteroposterior Risser staging varied in agreement. Yang et al. [20] used both the NA and EU Risser system to reveal that the Risser staging may be reliable, but not accurate when compared to 3D-CT images.

The Risser+ system exhibited excellent reliability (ICC = 0.96) across all reads and raters but inconsistent agreement. The second reading across all raters met the study requirements for acceptable agreement; however, the first did not. Further review of the data found that there were 20 outlying individual readings where at least 1 rater recorded a rating that was more than 4 levels different than the other 5 raters for the same radiograph (i.e., one rater marked a Risser 0 vs another marked a Risser 5). It was determined that removal of these 20 readings increased agreement of the Risser 8-point grading scale for the first, second, and overall reads (*k*-alpha = 0.85, 0.89, 0.87, respectively). This error may root from the examination of a single radiograph in the absence of relevant clinical data such as age, Tanner stage, or puberty history.

Previous studies by Goldberg et al. [21] and Dhar et al. [22] noted high inter-observer and intra-observer agreement when assessing Risser stages. Yet, Shuren et al. [18] denoted low inter-observer agreement with a kappa between 0.31 and 0.53. Reem [23] demonstrated low inter-observer agreement, but high intra-observer agreement between multiple rounds of rating. Hammond [24] found that there was low agreement between a radiologist and orthopedic surgeons with 50% agreement. Risser+ raters demonstrated better agreement during the second rating in all aspects (intra-rater and inter-rater) compared to the first rating; perhaps highlighting an increased level of comfort with the new Risser+ system after initial rating.

Our secondary analyses indicated that substantial agreement was present in the indication of high-risk progression (Risser ≤ 2 versus Risser > 2; $k > 0.6$); however, agreement did not meet the thresholds of the study ($k \geq 0.80$). Similarly, agreement with respect to discontinuing bracing (Risser ≥ 4 versus Risser < 4) and comparing 0− and 0 + across all reads was insufficient. The varying quality of radiographs available when compared to recent advances in low-dose, 3-D imaging may have contributed to the unsatisfactory agreement. The BrAIST study and Karol et al. provided strong evidence that there is greater risk of progression for patients of Risser 0 are greater than Risser 1 and 2 patients [25]. These results demonstrate the clinical impact of the Risser+ score when monitoring the progression of skeletal maturity and contemplating deformity treatment by nonoperative means but highlight the need for more for precision in decision making (such as the Scoliosis Digital Scoring system).

Prospective trials should address clinical relevance and optimization of the Risser+ score in comparison with the Scoliosis Digital Scoring system. The addition of sequential radiographs and more clinical data such as age, Tanner stage, or puberty history may increase overall agreement of the Risser+ grading scale. Moreover, the clinical impact of the ratings should be established to correlate Risser+ grade and bracing. Although previous iterations of the Risser score may vary, the Risser score is still very prominent within the clinical setting. The Risser+ score may allow for a common language between orthopedists, physical therapists, orthotists, etc., from various nations and regions. Future research may exhibit the potential impact and usefulness of the combined Risser+ scale in the clinical setting as the universality could provide consistency in reporting treatment results. The Risser+ could potentially lead to treatment consensus in terms of the commencement of bracing, scoliosis specific exercises, or other methods of growth modulation by operative techniques.

Conclusion

The Risser+ system showed excellent reliability (> 90%) across multiple reads and raters; however, demonstrated only 79% agreement overall reads and ratings. It was found that agreement increased to over 85% when raters could accurately distinguish Risser 0 + from Risser 5. This error may be due to the examination of a single radiograph in the absence of relevant clinical context. While the Risser+ system proved to have high reliability, the less than adequate agreement indicates that it is not a valid classification system in the absence of relevant clinical data. Additional research needs to be conducted to prove that proper agreement thresholds can be met when clinical data is known for participants in scoliosis treatment research studies. Yet the Risser+ score allows for a universal language between orthopedists and allied health professionals regardless of location and minimal learning curve.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Risser JC (2010) The classic: the iliac apophysis: an invaluable sign in the management of scoliosis. Clin Orthop Relat Res 468(3):643–653
2. Biondi J et al (1985) Correlation of Risser sign and bone age determination in adolescent idiopathic scoliosis. J Pediatr Orthop 5(6):697–701
3. Dimeglio A (2001) Growth in pediatric orthopaedics. Journal of Pediatric Orthopaedics 21(4):549–555
4. Little DG et al (2000) Relationship of peak height velocity to other maturity indicators in idiopathic scoliosis in girls*. J Bone Joint Surg 82(5):685–689
5. Sanders JO (2007) Maturity indicators in spinal deformity. J Bone Joint Surg Am 89(suppl 1):14–20
6. Sanders JO et al (2008) Predicting scoliosis progression from skeletal maturity: a simplified classification during adolescence. J Bone Joint Surg Am 90(3):540–553
7. Sanders JO et al (2008) Predicting scoliosis progression from skeletal maturity: a simplified classification during adolescence. JBJS 90(3):540–553
8. Richards BS et al (2005) Standardization of criteria for adolescent idiopathic scoliosis brace studies: SRS Committee on Bracing and Nonoperative Management. Spine 30(18):2068–2075
9. Negrini S et al (2012) 2011 SOSORT guidelines: orthopaedic and rehabilitation treatment of idiopathic scoliosis during growth. Scoliosis 7(1):1
10. Negrini S et al (2015) Recommendations for research studies on treatment of idiopathic scoliosis: consensus 2014 between SOSORT and SRS non–operative management committee. Scoliosis 10(1):1

11. Stagnara P et al (1965) Structural scoliosis. Terminal results of orthopedic treatments during the growth period. Rev Chir Orthop Reparatrice Appar Mot 51:33–52
12. Bitan FD, Veliskakis KP, Campbell BC (2005) Differences in the Risser grading systems in the United States and France. Clin Orthop Relat Res 436:190–195
13. Hacquebord JH, Leopold SS (2012) In brief: the Risser classification: a classic tool for the clinician treating adolescent idiopathic scoliosis. Springer, Berlin
14. Nault ML et al (2010) A modified Risser grading system predicts the curve acceleration phase of female adolescent idiopathic scoliosis. J Bone Joint Surg Am 92(5):1073–1081
15. Krippendorff K (2004) Reliability in content analysis. Hum Commun Res 30(3):411–433
16. O'Brien M, et al (2008) Radiographic measurement manual. Spinal Deformity Study Group (SDSG) Medtronic Sofamor Danek USA, Inc
17. Izumi Y (1995) The accuracy of Risser staging. Spine (Phila Pa 1976) 20(17):1868–1871
18. Shuren N et al (1992) Reevaluation of the use of the Risser sign in idiopathic scoliosis. Spine (Phila Pa 1976) 17(3):359–361
19. Zaoussis AL, James JI (1958) The iliac apophysis and the evolution of curves in scoliosis. J Bone Joint Surg Br 40-B(3):442–453
20. Yang JH et al (2014) Evaluation of accuracy of plain radiography in determining the Risser stage and identification of common sources of errors. J Orthop Surg Res 9:101
21. Goldberg et al (1988) Observer variation in assessing spinal curvature and skeletal development in adolescent idiopathic scoliosis. Spine (Phila Pa 1976) 13(12):1371–1377
22. Dhar S et al (1993) Correlation between bone age and Risser's sign in adolescent idiopathic scoliosis. Spine (Phila Pa 1976) 18(1):14–19
23. Reem J et al (2009) Risser sign inter-rater and intra-rater agreement: is the Risser sign reliable? Skeletal Radiol 38(4):371–375
24. Hammond KE et al (2011) Inter-observer and intra-observer reliability of the Risser sign in a metropolitan scoliosis screening program. J Pediatr Orthop 31(8):e80–e84
25. Karol LA et al (2016) The effect of the Risser stage on bracing outcome in adolescent idiopathic scoliosis. J Bone Joint Surg Am 98(15):1253–1259

Affiliations

Michael J. Troy[1] · Patricia E. Miller[1] · Nigel Price[2] · Vish Talwalkar[3] · Fabio Zaina[4] · Sabrina Donzelli[4] · Stefano Negrini[4] · M. Timothy Hresko[1]

✉ M. Timothy Hresko
Timothy.Hresko@childrens.harvard.edu

[1] Boston Children Hospital, Harvard Medical School, Boston, USA

[2] Children's Mercy Hospital, Kansas City, USA

[3] Shriners Hospital, Lexington, USA

[4] ISICO - L'Istituto Scientifico Italiano Colonna Vertebrale, Milan, Italy