

Identifying Sources of Variability in Scoliosis Classification Using a Rule-Based Automated Algorithm

Ian A. F. Stokes, PhD, and David D. Aronsson, MD

Study Design. Use of a rule-based automated algorithm to determine sources of variability in radiographic classification.

Objectives. To determine whether unambiguous rules encoded in a computer program would ensure reliable classification.

Summary of Background Data. Reliability problems have been identified in classifications used in surgical planning for patients with thoracic idiopathic scoliosis, but the sources of unreliability are not understood.

Methods. Objective classification methodology was tested on the King *et al* (1983) scheme. There were two novel components: 1) positions of the corners of vertebrae in radiographs were digitized relative to a defined axis system and used in automated evaluation of spinal shape parameters required for classification; and 2) the assignment of a classification was done with a rule-based algorithm. The algorithm was implemented after some ambiguities and absence of precise definitions in the King *et al* classification scheme had been resolved. The algorithm was tested with radiographs of patients having adolescent idiopathic scoliosis.

Results. The automated procedure could encounter reliability problems in cases in which a lumbar curve was very close to crossing the midline, thoracic and lumbar curves were of approximately equal value, when the apex level in the thoracolumbar region was ambiguous, when a Cobb angle was close to 10°, or when the flexibility index was close to unity.

Conclusion. Objective measurements and rule-based algorithms can eliminate some sources of interobserver and intraobserver errors in classification of spinal deformity. When classification parameters fall close to the boundaries for classification, reliability problems will persist. [Key words: classification, scoliosis, radiography, algorithm] **Spine 2002;27:2801–2805**

Classification systems used to guide the management of orthopedic problems should be reliable (the classification must be consistent), and each classification should relate to a different prognosis or management strategy.¹ In surgical planning for arthrodesis in the surgical manage-

ment of adolescent idiopathic scoliosis, classification by radiographic measures is used to help determine the extent of the arthrodesis. The King *et al* classification³ is probably still the most widely used in planning of a spinal arthrodesis, although it was originally developed to aid the planning of a spinal arthrodesis using Harrington instrumentation. It defines five thoracic scoliosis curve types and an additional group called “miscellaneous.” The King *et al* classification³ relies on subjective identification of the radiographic features that provide the measures used for classification. It may also require individual interpretation and memory of the classification criteria. As a result, there are numerous opportunities for variable implementation that produce interobserver and intraobserver variability. Two recent empirical studies^{2,4} of repeat classification by the King *et al* method³ have demonstrated poor reliability.

Empirical studies can evaluate the overall reliability of a classification of a given sample of patients by statistics such as the kappa statistic. However, this approach has not been successful in identifying sources of variability between observers or between repeated evaluations by the same observer. In examining the procedure used to classify scoliosis by the King *et al*³ method, we identified several possible sources of imprecision. These include variation in evaluating measurements used in classification as well as observer errors in interpreting or memorizing the criteria for assigning curve types. The measurements used in classification include Cobb angles, curve endpoints, range of lateral bending, and parameters relating to spinal alignment and vertebral tilt. The present study aimed to provide insights into how to reduce or eliminate the sources of errors.

The purpose of this report was to determine whether radiographs of patients with idiopathic scoliosis could be reliably classified by the King *et al* method,³ using unambiguous rules for classification encoded in a computer program. The method required two steps: the first step was to record the coordinates of defined landmarks on the radiographs of the patient to assist objective determination of the parameters used to distinguish between groups, and the second step was to apply a rule-based algorithm to assign the King *et al* classification³ based on these measurements.

Materials and Methods

Based on the description given by King *et al*,³ the following four sets of parameters may be required to classify a given patient: 1) from standing posteroanterior or anteroposterior radiographs,

From the Department of Orthopaedics and Rehabilitation, University of Vermont, Burlington, Vermont.

Acknowledgment date: June 20, 2001. First revision date: January 2, 2002. Second revision date: March 28, 2002. Acceptance date: May 13, 2002.

The manuscript submitted does not contain information about medical device(s)/drug(s).

No funds were received in support of this work. No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this manuscript.

Address correspondence to Ian A. F. Stokes, PhD, Department of Orthopaedics and Rehabilitation, 434 Stafford Hall, University of Vermont, Burlington, VT 05405, USA; E-mail: Ian.Stokes@uvm.edu

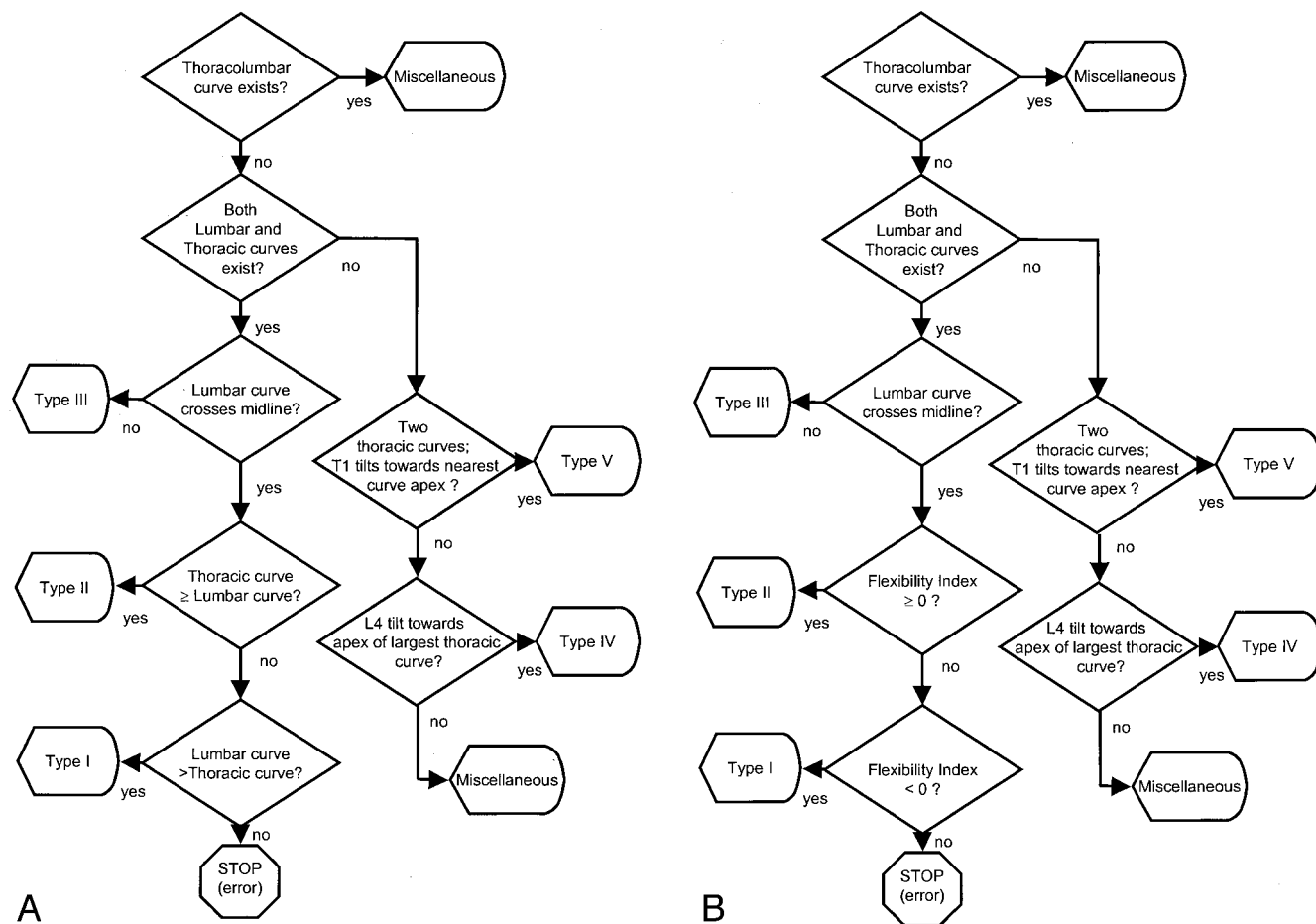


Figure 1. **A**, Flowchart of the algorithm used to classify each patient using measurements from a posteroanterior radiograph. **B**, Flowchart of the algorithm to classify each patient using measurements from a posteroanterior radiograph and supine lateral bending radiographs.

the number of scoliosis curves, each identified by apex level, end-vertebrae, and Cobb angle; 2) positions of certain vertebrae relative to the plumbline, and whether the lumbar curve and thoracic curves crossed the midline; 3) the tilt of L4 and T1 and the elevation of the first ribs; and 4) the flexibility index (FI), defined as the percent correction of the thoracic curve on supine side bending, subtracted from the percent correction of the lumbar curve.

According to Table 1 in King *et al*³ and the associated text of that report, patients with Type I or Type II scoliosis have an S-shaped curve pattern (in which both a lumbar and thoracic curve cross the midline).

- Patients with Type I scoliosis have a lumbar curve Cobb angle 3 or more degrees greater than the thoracic curve on a standing radiograph or have a negative flexibility index.
- Patients with Type II scoliosis have a thoracic curve Cobb angle equal to or greater than the lumbar curve on a standing radiograph or have a zero or positive flexibility index.
- Patients with Type III scoliosis have a thoracic curve, but the lumbar curve does not cross the midline, and the plumbline is centered over the sacrum.
- Patients with Type IV scoliosis have a thoracic curve, the L5 vertebra is centered over the sacrum, and the L4 vertebra tilts into the long thoracic curve.
- Patients with Type V scoliosis have a double thoracic curve pattern with T1 tilted into the convexity of the upper curve and have an elevated first rib.

To implement this classification in objective algorithmic form, it was realized that the descriptions of curve types included ambiguities and were incompletely defined. First, Type I and Type II were not necessarily mutually exclusive (*e.g.*, if the flexibility index were negative and the Cobb angle of the thoracic curve were greater than that of the lumbar curve, the patient could be classified as either Type I or Type II). Second, the definitions would exclude any patient having a lumbar curve whose Cobb angle is between 0° and 3° larger than that of the thoracic curve. To resolve the first problem, our algorithm used either the Cobb angle magnitude criterion (Figure 1A) or the flexibility criterion (Figure 1B). To resolve the second problem, our criterion for Type I curve required that the lumbar Cobb angle only be greater than the thoracic Cobb angle, rather than a 3° difference (as implied by Table 1 in King *et al*³).

We noted that the definition of the plumbline in general requires a repeat radiograph to level the iliac crests, and this is probably seldom done in modern practice. Therefore, we defined the central sacral line as the vertical line passing through a sacral midpoint, irrespective of whether the pelvis was level. The concept of “crossing the midline” is not fully defined; therefore, we adopted a definition that required that all four marked corners of the apical vertebra lie on the same side of the central sacral line (see below).

In a Type III curve pattern, the plumbline is directly centered over the sacrum, and in Type IV curves, the L5 vertebra is centered over the sacrum. It was noted that these concepts are

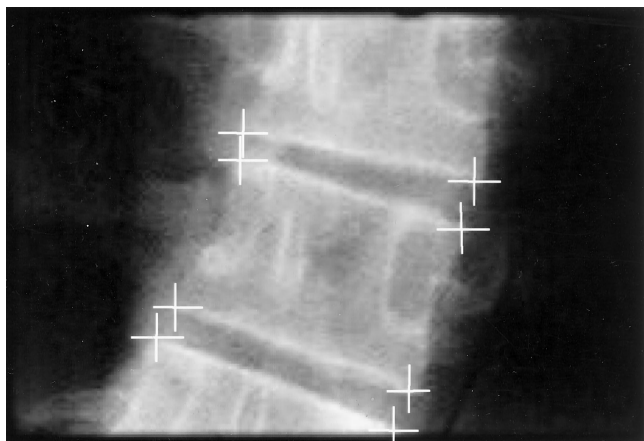


Figure 2. Detail of an image obtained by the digital camera, showing the digitized points at the corners of the vertebrae.

not defined quantitatively and also add additional requirements for patients to be included in those groups, potentially leaving some patients unclassified. Therefore, we did not include centering of the plumbline or centering of L5 in our algorithm. Similarly, elevation of the first rib is not defined for Type V, so we included only the T1-tilt criterion.

To obtain the quantitative measurements used in classification unambiguously and objectively, the coordinates of bony landmarks identified visually in the patients' posteroanterior radiographs were digitized. Two alternative but similar digitization procedures using custom software were developed. In one, the full-size radiograph was placed on a backlit translucent digitizing tablet (GTCO Corp., Rockville, MD), and image points were selected using the tablet's cursor and transmitted to a personal computer. In the alternate procedure, a digital photograph of the radiograph was taken using a Coolpix 950 camera (Nikon, Tokyo, Japan), and then the image points were selected by a "mouse" pointing device on the photographic image displayed on a personal computer screen. The digitized landmarks were the four corners of each vertebra from T1 through L5 (Figure 2) and two symmetric landmarks on the proximal sacrum (*e.g.*, right and left margins of the S1 endplate), for a total of 70 points. We adopted the following procedures relating to the posteroanterior radiograph: the midpoint between the symmetric digitized points on the sacrum defined the origin of the patient's x-y coordinate system. The central sacral line was defined as the vertical line passing through this origin. The lateral edge of the radiograph was used to define the vertical direction (as for the center sacral vertical line [CSV] defined by Lenke *et al*⁵). The total image digitizing time by the human operator was found to be 2.5–3 minutes, depending on image quality and landmark visibility.

To have a computer program identify curves and measure their geometric properties automatically (without subjective intervention), the position and tilt of each endplate and vertebra were calculated from the digitized coordinates. The endplate tilt angle was calculated as the arc tangent of the difference in y-coordinates divided by the difference in the x-coordinates of the corresponding two points. The tilt of a vertebra was calculated as the mean tilt of its two endplates. The position of each vertebra and of each disc in the patient's coordinate system was defined by the average x- and y-coordinates of its four corners. (This point is equivalent geo-

metrically to the intersection of lines joining diagonally opposite corners of the vertebra.)

According to accepted definitions, the curve apex is the vertebra or disc having the greatest lateral deviation from the vertical axis in the apical region of the spinal curve, and the end-vertebrae of each curve are those with the greatest tilt of their superior endplate (proximal end vertebra) or inferior endplate (distal end vertebra). The Cobb angle is the difference in tilt angle of endplates of the end vertebrae. Automated identification of curves and curve apexes according to the Cobb definitions required the following procedures:

1. To automate the identification of the apex, an apical region was first located that included the two vertebrae above and the two vertebrae below a disc separating two vertebrae tilting in opposite directions to the horizontal. The vertebra or disc in this apical region that had the greatest horizontal distance from the y-axis was then identified as the apex.
2. To identify the end-vertebrae automatically, an inflection (change of spinal curvature) above and below each apex was first located. Then the two vertebral levels above and below each inflection point were examined to find the endplate with the greatest tilt. Finally, to comply with the normal definition of what constitutes a curve, any curve with a Cobb angle of $<10^\circ$ was excluded from further analysis.

The classification algorithm then identified curves patterns either by posteroanterior radiographic measures alone (Figure 1A) or by using flexibility index instead of Cobb angle criteria to distinguish Type I and Type II curves. For the Cobb angle magnitude criterion, the automated classification algorithm (Figure 1A) used the following logic:

If at least one thoracic curve and one lumbar curve exist, then the curve pattern is Type I if both curves cross the midline and the lumbar Cobb angle is greater than the thoracic Cobb angle, Type II if both curves cross the midline and the thoracic Cobb angle is greater than or equal to the lumbar Cobb angle, and Type III if there is a lumbar curve that does not cross the midline. Otherwise, the curve pattern is Type IV if there is no lumbar curve and L4 tilts toward the apex of the largest thoracic curve, and Type V if two or more thoracic curves exist, and T1 tilts toward the apex of the most cephalad thoracic curve.

The algorithm using curve magnitude (Figure 1A) was implemented by coding it in a computer program written in Matlab® language (The MathWorks, Natick, MA). This program first located the curve apexes, end-vertebrae, and the Cobb angle (if $>10^\circ$). It then determined whether apical vertebrae crossed the midline, according to the above definitions. Finally, the King *et al*³ classification of the curve pattern was assigned. In the algorithm given in Figure 1B, the flexibility criterion is used to distinguish between Type I and Type II. An example of the digitized points and features identified by the computer program is shown in Figure 3.

The Figure 1A algorithm was verified on the five curve type examples (tracings of radiographs) in the report by King *et al*³ and evaluated using the six radiographic examples in the report by Lenke *et al*⁴ and the one radiographic example in the report by Cummings *et al*.² The published figures from each article were enlarged on a photocopying machine, and the 70 required points were marked by each of two observers (the authors of this article) on fresh copies. The algorithm was also tested with

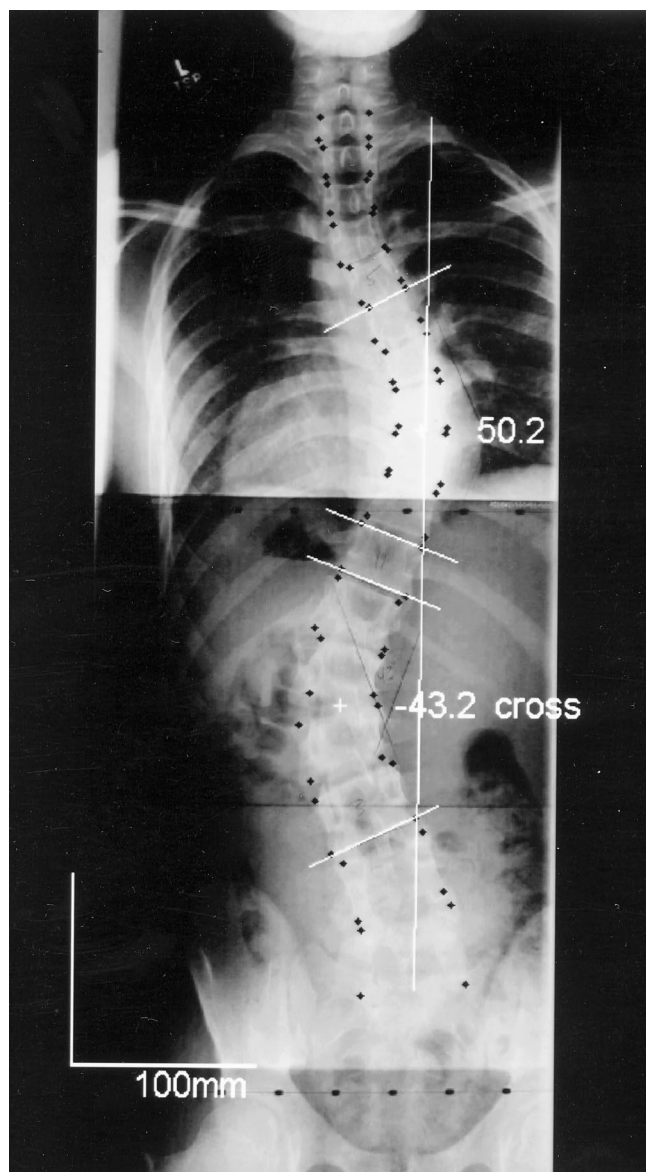


Figure 3. Sample output from the classification program showing the four digitized points at the corners of each thoracic and lumbar vertebra, derived lines and identified curves overlaid on the radiograph, and noting the curve pattern (King *et al*³ Type II) deduced by the computer program.

anteroposterior radiographs of 17 patients with adolescent idiopathic scoliosis. Patients were selected if they had a previously measured curve magnitude $>30^\circ$ Cobb (mean 41° , SD 11° , range $30\text{--}63^\circ$) and the radiograph was of adequate quality to mark thoracic and lumbar vertebral corners. Radiographs were marked and digitized twice (after a delay of several days), and four of these films were each marked twice by two observers. Among these patients, there was 1 case with a Type I curve pattern, 3 with type II, 10 with Type III, 2 Type IV, and 1 case that was variously classified at Type V or Type III.

■ Results

The examples of each of the five curve types in the report by King *et al*³ were correctly classified, except the Type I example that was classified as thoracolumbar because

the apical level was determined to be at L1. The second and third examples from the report by Lenke *et al*⁴ were inconsistently classified as either Type II or Type III. These two examples had a lumbar apical vertebra that was questionable as to whether or not it crossed the midline, and were reported by Lenke *et al*⁴ as having been unreliably classified by the observers in that study. In another case, there was inconsistency as to whether the apex of the lower curve was determined to be at a lumbar or thoracolumbar level. The five other published examples were consistently classified, and the classification was in agreement with the most common classification by the observers as reported by the original authors.

The 17 original radiographs that were selected for this study were consistently classified, with two exceptions. In one case, there was a lumbar curve with a Cobb angle close to 10° that resulted in this curve pattern being variously classified as Type III or Type V depending on whether the magnitude of the lumbar curve was determined to be greater or less than the threshold value. In the other exception, the patient was variously classified as Type II or Type III because of uncertainty as to whether the lumbar curve crossed the midline.

■ Discussion

We set out to determine whether radiographs of patients with idiopathic scoliosis could be reliably classified by the King *et al* method,³ using unambiguous rules for classification encoded in a computer program. This required that the classification scheme be defined unambiguously, with all classification criteria uniquely defined. Consequently, the only source of variability in classification of any given radiograph was variable landmark identification by the observer.

The classifications that were determined automatically by using the computerized algorithm were found to be accurate and reliable in most cases. The exceptions occurred when a radiographic curve pattern included a feature close to a cutoff value of a variable used in the classification. This was found to occur in three situations: 1) when it was borderline, whether or not the apical vertebra of a lumbar curve crossed the midline; 2) when a Cobb angle was close to 10° ; or 3) when the apical vertebra of a lumbar curve level was close to thoracolumbar levels. Another possible source of variability in this automated classification would arise if the thoracic and lumbar curves were of very similar magnitude (curve magnitude criterion) or if the flexibility index were close to 0. Then the curve pattern could be identified as either Type I or Type II, depending on which curve was determined to have the greater Cobb angle or percent correction on side bending.

To obtain a unique classification for any patient, some additional definitions were required to eliminate ambiguity in the King *et al* report,³ and also the plumbline and rib-tilt criteria were omitted. The unique distinction between Type I and Type II patients was only found to be possible by using either flexibility index or curve magni-

tude criteria, but not both. We chose to test the algorithm using curve magnitude (as would normally be used before immediate preoperative planning when lateral bending films are made). For classification using FI, the exact procedure for making lateral bending films ought to be standardized, and it must be specified that the same end-vertebrae as those located in the standing film are used to measure curve correction.

These difficulties in automating a specific scoliosis classification point to inherent problems that may exist in any classification. If patients are frequently observed to be close to the border between two groupings, reliability problems are likely to occur. These will be exacerbated if the procedures used in classification are incompletely defined, or the measurements that are required are liable to technical errors. Ideally, a classification distinguishes between groups of patients using parameters that rarely have values close to the cutoff criteria. This occurs when affected patients have values of the classification parameters that form a bimodal distribution, with the classification criterion that separates two groups lying between the two modes. In the classification of patients with adolescent idiopathic scoliosis, the distribution of relative Cobb angle magnitudes, curve apex levels, and positions of lumbar curve apical vertebrae relative to the central sacral line are continuous and apparently do not have the bimodal distribution that would be amenable to distinct classification. Probably there are many patients being managed clinically who have measurements close to the values used in classification and, therefore, are liable to be classified unreliably. An additional source of variability would be exposed if repeat radiographs were done, *e.g.*, at different times of the day.

For these reasons, the empirical evaluation of the reliability of a classification scheme by empirical studies of interobserver and intraobserver variability quantified by the kappa statistic or other measures of overall repeatability might produce quite different findings depending on which specific patients or radiographs are selected to evaluate the reliability. In particular, the level of reliability that is found in such a study might depend on whether the test radiographs include cases that have features close to the cutoff criteria. Among the patients whose radiographs were tested in the present study, there were several for whom the mea-

asures of Cobb angle, position of a lumbar apex relative to the central sacral line, and location of a curve apex happened to occur at close to the cutoff values. The quality of the radiographs and postural control of patients during radiography are additional potential sources of error in identifying vertebral positions and endplate tilt and, hence, in curve identification.

The algorithmic approach to automated classification that was used here could potentially be applied to other orthopedic classification schemes, especially those that use measurement of radiographic features. A prerequisite for automated classification is that the method to measure the classification variables, as well as cutoff values between classification and the classification algorithm, are all precisely defined. For reliable classification, it is also important that the parameters that distinguish between groups rarely occur with values close to the cutoff criteria.

■ Key Points

- A rule-based algorithm should eliminate sources of errors and variability in radiographic classification of spinal deformity, other than problems from variable radiographic landmark identification.
- It was found that the process of writing a rule-based scoliosis classification algorithm identified and resolved ambiguities in definitions of curve types.
- When measured parameters used in classification are close to the thresholds separating curve types, reliability problems persist.

References

1. Burstein AH. Fracture classification systems: do they work and are they useful? *J Bone Joint Surg Am* 1993;75:1743–4.
2. Cummings RJ, Loveless EA, Campbell J, et al. Interobserver reliability and intraobserver reproducibility of the system of King *et al.* for the classification of adolescent idiopathic scoliosis. *J Bone Joint Surg Am* 1998;80:1107–11.
3. King HA, Moe JH, Bradford DS, et al. The selection of fusion levels in thoracic idiopathic scoliosis. *J Bone Joint Surg Am* 1983;65:1302–13.
4. Lenke LG, Betz RR, Bridwell KH, et al. Intraobserver and interobserver reliability of the classification of thoracic adolescent idiopathic scoliosis. *J Bone Joint Surg Am* 1998;80:1097–106.
5. Lenke LG, Betz RR, Harms J, et al. Adolescent idiopathic scoliosis: a new classification to determine extent of spinal arthrodesis. *J Bone Joint Surg Am* 2001;83:1169–81.