SPINE Volume 31, Number 6, pp 665–670

Computer-Assisted Algorithms Improve Reliability of King Classification and Cobb Angle Measurement of Scoliosis

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

Study Design. Interobserver and intraobserver reliability study of improved method to evaluate radiographs of patients with scoliosis.

Objective. To determine the reliability of a computer-assisted measurement protocol for evaluating Cobb angle and King *et al* classification.

Summary of Background Data. Evaluation of scoliosis radiographs is inherently unreliable because of technical and human judgmental errors. Objective, computer-assisted evaluation tools may improve reliability.

Methods. Posteroanterior preoperative radiographic images of 27 patients with adolescent idiopathic scoliosis were each displayed on a computer screen. They were marked 3 times in random sequence by each of 5 evaluators (observers) who marked 70 standardized points on the vertebrae and sacrum in each radiograph. A computer program (*Spine* 2002;27:2801–5) that identified curves, calculated Cobb angles, and generated the King *et al* classification automatically analyzed coordinates of these points. The interobserver and intraobserver variability of the Cobb angle and King *et al* classification evaluations were quantified and compared with values obtained by unassisted observers.

Results. Average Cobb angle intraobserver standard deviation was 2.0° for both the thoracic and lumbar curves (range 0.1 to 8.3° for different curves). Interobserver reliability was 2.5° for thoracic curves and 2.6° for lumbar curves. Among the 5 observers, there was an inverse relationship between repeatability and time spent marking images, and no correlation with image quality or curve magnitude. Kappa values for the variability of the King *et al* classification averaged 0.85 (intraobserver).

Conclusions. Variability of Cobb measurements compares favorably with previously published series. The classification was more reliable than achieved by unassisted observers evaluating the same radiographs. The same principles may be applicable to other radiographic measurement and evaluation procedures.

Key words: scoliosis, radiographs, computer-assisted, classification, Cobb angle. **Spine 2006;31:665–670**

The treatment of patients with idiopathic scoliosis relies heavily on radiographic measurements to identify coronal and sagittal curves, detect progression of deformity, and assist in the planning of conservative and surgical management. The Cobb angle has become the basis for quantifying scoliosis curve magnitude. Studies of interobserver and intraobserver variability in measurement of this angle[1–6] have revealed that the errors in radiographic measurements are typically ±5°, and this is comparable with thresholds of change that can influence treatment decisions.[3] The sources of the errors may include incorrect selection of the most tilted endplates, random errors in drawing lines across the endplates, and systematic errors caused by inaccurately manufactured protractors.[3]

Spinal curve pattern classifications that rely on radiographic measures are used in surgical planning for patients with adolescent idiopathic scoliosis, to select fusion levels.[7] The classification by King *et al*[8] is still the most widely used in surgical planning, although it was originally developed specifically for procedures using Harrington instrumentation. It defines 5 thoracic scoliosis curve types and an additional group called "miscellaneous" based on measurements on standing radiographs, and can include measurements from lateral bending films. The King *et al* classification relies on subjective identification and measurement of the radiographic features, including the apical and end vertebrae of curves, vertebral endplate tilt angles, and the origin and alignment of the central sacral line. It also requires individual interpretation and memory of the classification criteria. Errors in identifying these radiographic landmarks and using the resulting measurements in identifying the pattern of deformity provide numerous opportunities for both technical and judgmental errors, producing interobserver and intraobserver variability.[9] Empirical studies[10–13] of repeat classification by the King *et al* method have shown problems with reliability.

A computer-assisted algorithm[9] intended to minimize human involvement in Cobb angle measurement and, in King *et al* classification, identified potential sources of classification errors. Specifically, it was reported that the classification was often unreliable when a radiographic measurement was close to a threshold value used to distinguish between 2 curve types. For example, when the apical vertebra of a lumbar curve was close to the central sacral line, repeated evaluations differed as to whether the curve crossed the midline, and, consequently, the classification was inconsistent. Digital radiography can facilitate software-assisted evaluation of radiographs to replace traditional "pencil and ruler" measurement

From the Department of Orthopaedics and Rehabilitation, University of Vermont, Burlington, VT.
Acknowledgment date: September 23, 2004. First revision date: February 7, 2005. Acceptance date: April 4, 2005.
The manuscript submitted does not contain information about medical device(s)/drug(s).
Federal 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 and reprint requests to Ian A. Stokes, PhD, Department of Orthopaedics and Rehabilitation, University of Vermont, Stafford Hall 434 Burlington, VT 05405-0084; E-mail: Ian.Stokes@uvm.edu

methods, and evaluations that rely on subjective assessment and memory of classification criteria.

The purpose of this report was to determine the reliability of the previously published[9] computer-assisted protocol for evaluating Cobb angle and King *et al* classification of radiographs of patients with idiopathic scoliosis who were candidates for surgery. Some possible influences on reliability, including the experience of the individual using the computer-assisted tool, magnitude of the scoliosis, and image quality, were investigated.

■ Methods

Posteroanterior radiographs taken before surgery of 27 patients with adolescent idiopathic scoliosis who had been selected previously for studies of classification reliability[10,14] were used in the present study, thus permitting direct comparisons. The average Cobb angle in this series was 64° (range 45° to 105°). For the present study, the radiographic images had been digitized, and were displayed in a randomized sequence on a computer screen and marked 3 times by each of 5 observers. The radiographic images were provided as a "PowerPoint" (Microsoft, Redwood, WA) file, from which gray-scale image files were extracted. The image sizes were 925 pixels high, by typically 475 pixels wide (*i.e.*, each pixel was about 1 mm on the original spinal radiographs that were about 900-mm high). This format simulated digital radiography, although the pixel resolution was somewhat less. A total of 70 standardized radiographic landmarks were marked on each image using custom software and a computer "mouse" to "click" on select points whose coordinates were then stored. The landmarks were the corners of the vertebral bodies (extremes of the endplate images) from T-1 to L-5 and 2 symmetrical landmarks on the sacrum, used to obtain the "central sacral line" (Figure 1). The images were supplied with the anatomic level T-1 identified.

The 5 observers included a pediatric orthopedic surgeon member of the Scoliosis Research Society, a nonclinician researcher having many years experience of marking spinal radiographs in a research context, an orthopedic resident, a musculoskeletal radiologist, and a premedical student having no experience with spinal radiographs. The latter 3 evaluators had not used this computer-assisted tool before the present study. They learned how to use it by having it shown to them, emphasizing that they should identify and mark the corners of the vertebral body images and symmetrical points on the pelvis. This supervised instruction took about 30 minutes, including the time required to learn how to start the program and to identify locations of image files, *etc.* Subsequently, the evaluators used the tool without supervision and without discussion of the findings. The computer recorded the time of completion of processing of each film, from which the time taken to process individual images was derived.

The stored coordinates were input to a published computer algorithm[9] that used derived vertebral positions and endplate tilt angles to identify scoliosis curves, their apexes, end vertebrae, and their Cobb angles, using a strict rule-based approach. The classification algorithm (Figure 2) implemented the King *et al* classification using the published rules,[8] incorporating the Cobb measurements and other criteria (positions of vertebrae and endplate tilt angles). For type 1 and 2 curves, the King *et al* classification distinguishes between these classes based either on relative Cobb angle magnitudes or the "flexibility index"

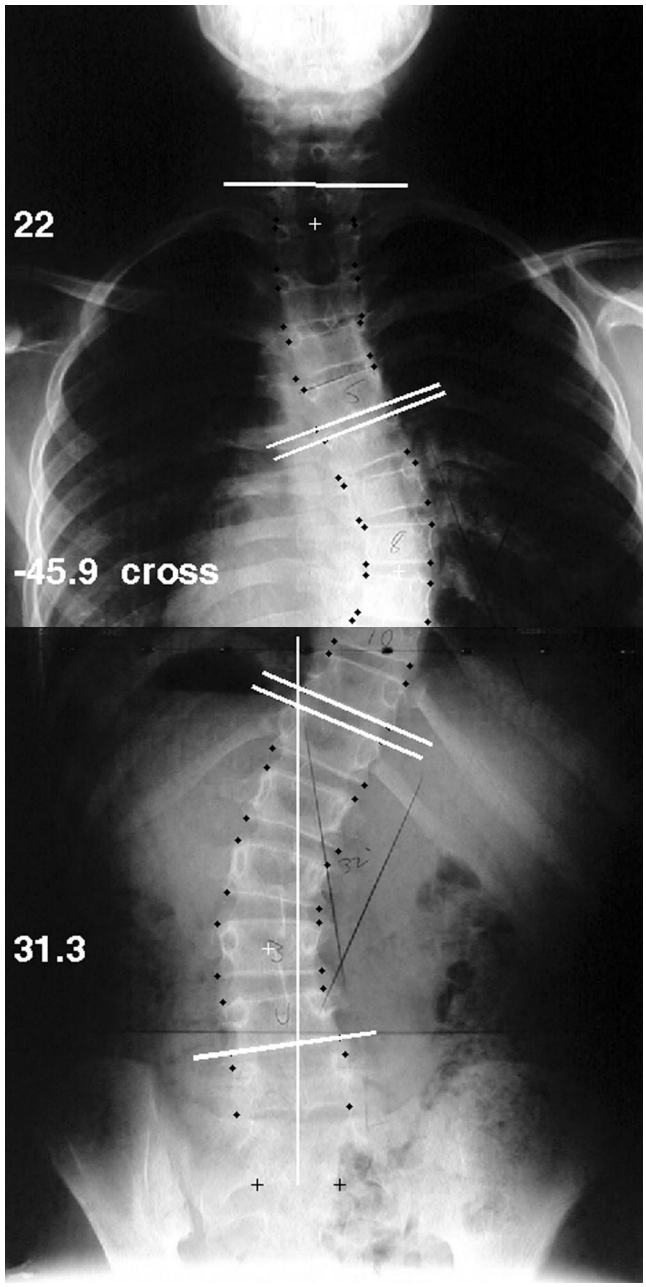

Figure 1. Radiographic image showing marked points (4 on each vertebral body, 2 on the sacrum), and the curve end vertebrae and Cobb angle values, as determined from the landmarks by the computerized algorithm. The curve pattern was automatically classified as type 3.

obtained from lateral bending films. The criteria to distinguish these curve types as stated in King *et al*[8] are ambiguous, requiring alternate algorithms in automated classification.[9] Here, the Cobb angle criterion was used, not the flexibility index (Figure 2). The radiographs were assumed to be aligned with the vertical, so the central sacral line was considered parallel to the film edge, and passing through the midpoint of the 2 sacral landmarks.

The quality of each radiograph was evaluated subjectively by two of the observers. They rated each image using a scale from 0 to 10, where a score of 10 would indicate that all landmark points were easily identifiable, and 0 would indicate that none were visible.

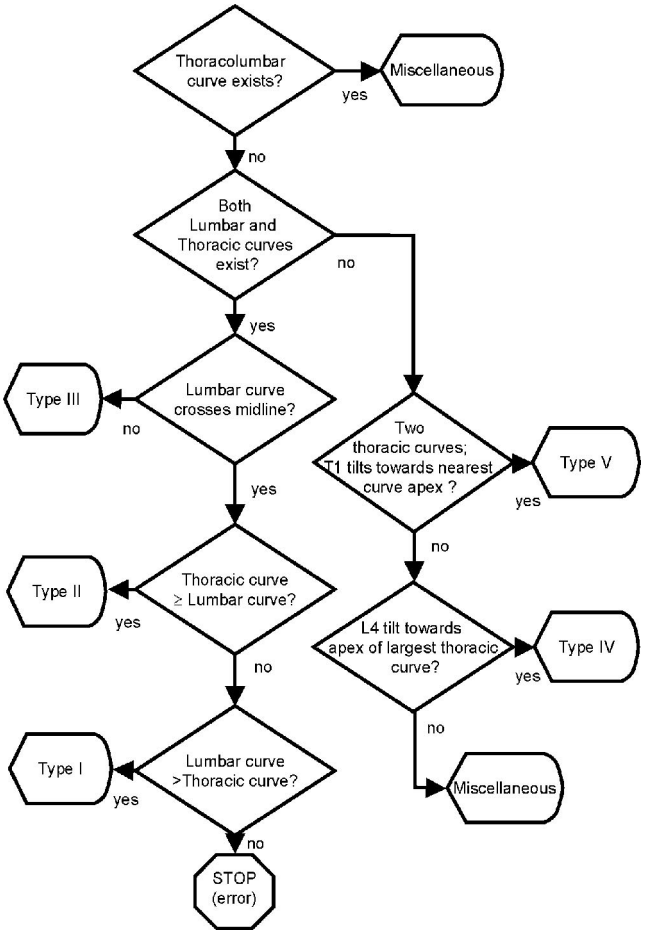

Figure 2. Flowchart of the algorithm used in the computer-assisted classification. Adapted with permission from *Spine* 2002; 27:2801–5.[9]

Statistical Methods. The intraobserver repeatability of the Cobb angle measurements for each curve/observer combination was calculated as the sample standard deviation of each observer's 3 measurements of each curve. Similarly, the interobserver reliability for each curve/trial combination was calculated as the sample standard deviation of the 5 observers' measurements in each trial. These values were then averaged across the 27 patients. For the purpose of statistical analyses of the Cobb angle data, each patient was assumed to have provisionally 2 curves, which were analyzed separately. If ≥2 curves were present, the main thoracic curve was designated as the upper curve, and the thoracolumbar or lumbar curve was designated as the lower curve. If an upper thoracic curve was present, it was omitted from the Cobb angle reliability analyses.

The kappa statistic[15] that measures the proportion of consistent classifications in 2 sets of observations, corrected for the observed frequency of each class, was used to assess the variability in the King *et al* classification. This statistic was calculated for paired sets of classifications by each observer (intraobserver repeatability) or between observers (interobserver reliability), using all combinations of paired observations. The resulting values were averaged over combinations of pairs (between or within observers) to provide an overall measure of interobserver and intraobserver variability.

■ Results

Intraobserver Repeatability in Cobb Angle Evaluation

Average sample standard deviations of 2.0° for both the upper and lower curves (Table 1) characterized Cobb angle variability between measurements of each individual patient. The highest individual measurement error (*i.e.*, difference from the overall mean for that curve) was 8.3°. There was no significant correlation between Cobb angle repeatability for each patient and the image quality score ($R^2 = 0.17$ for upper curves; $R^2 = 0.02$ for lower curves). The image quality evaluation was reliable, based on a correlation with $R^2 = 0.62$ between the numerical scores assigned by the 2 independent observers.

Interobserver Reliability in Cobb Angle Evaluation

The standard deviations of the samples of repeated observations averaged 2.5° for upper curves, and 2.6° for lower curves. There was no trend of the reliability increasing or decreasing over trials (Table 1) and no evidence of systematic differences between observers.

Intraobserver Repeatability in King et al Classifications

Kappa values for the intraobserver repeatability of the King *et al* classification averaged 0.85. The range was from 0.81 to 0.88 for the 5 observers (Table 2). Of the 5 observers, the Cobb angle and King classification repeatability did not correlate with experience in treating patients with scoliosis, but there was an inverse relationship between the rate of marking the radiographic images and repeatability (Figure 3). The Spearman rank correlation coefficient between the number of films marked per hour and Cobb angle repeatability was 1.0 ($P < 0.001$), and it was 0.82 ($P < 0.05$) for the correlation between films marked per hour and kappa statistic for classification repeatability.

Interobserver Reliability in King et al Classifications

The overall interobserver kappa values increased from 0.72 to 0.91 over the 3 series of measurement. The average interobserver kappa was 0.82 (Table 2). All 5

Table 1. Standard Deviations of Repeated Measures of Cobb Angles (degrees)

Intraobserver Repeatability		
Observer	Upper Curve	Lower Curve
1	2.1	2.1
2	1.8	1.8
3	1.9	1.7
4	2.3	2.4
5	2.0	2.2
Average	2.0	2.0
Interobserver Reliability		
Trial	Upper Curve	Lower Curve
1	2.6	2.6
2	2.3	2.4
3	2.7	2.6
Average	2.5	2.6

Table 2. Kappa Values for the King *et al* Classifications

Intraobserver Repeatability		
Observer	Classification Consistency (%)	Kappa Value
1	88	0.84
2	90	0.87
3	90	0.88
4	85	0.81
5	90	0.87
Average	89	0.85
Interobserver Reliability		
Trial	Classification Consistency (%)	Kappa Value
1	79	0.72
2	87	0.84
3	93	0.91
Average	86	0.82

observers in all 3 trials consistently classified 13 patients. Of the remaining 14 patients who were classified inconsistently, 4 different causes of the inconsistency were identified:

1. Inconsistent detection of an upper thoracic curve (King-type assignment as variably 5 or 4, 3 or 4, or 3 or 5), which occurred in a total of 4 patients in this study.
2. Inconsistent detection of the lumbar curve crossing the midline (King-type either 2 or 3), which occurred in 2 patients in this study.
3. Inconsistent identification of the thoracolumbar curve apex level (King-type assigned variably as miscellaneous or type 1, or type 5), which occurred in 4 patients in this study.
4. Inconsistent calculation of the relative Cobb angle magnitudes of the upper and lower curve (assigned either type 1 or type 2), which occurred in 4 patients in this study.

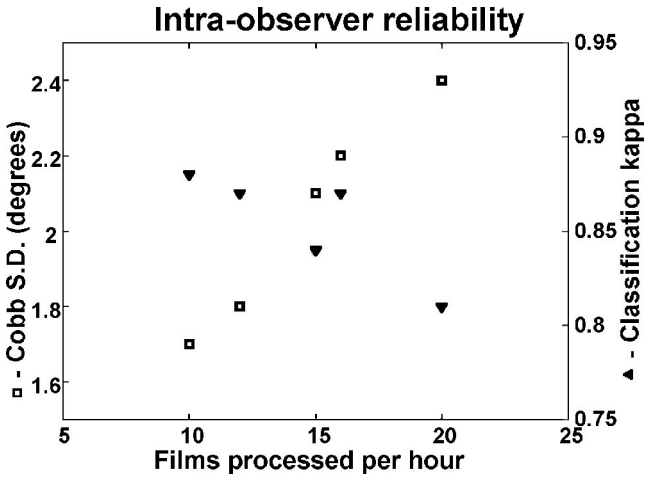

Figure 3. Intraobserver variability as a function of time spent by each observer marking the radiographs. Mean standard deviation (SD) for Cobb angle measurements (unfilled squares) and mean kappa statistic for King *et al* classifications (filled triangles) are shown.

■ Discussion

Human observers can make technical and judgmental errors in evaluating spinal radiographs, thereby reducing accuracy and reliability. The findings of this study indicate that the task of identifying scoliosis curves in radiographs and subsequently classifying the curve type is more reliable with the assistance of a computerized tool. With the use of this tool, standardized measurement procedures can reduce the technical errors, and formal objective algorithms can reduce the judgmental errors. Clinical experience was not a factor in determining an individual observer's reliability. Instead, the time spent in selecting the radiographic landmarks was a significant factor. Relative to manual marking and classification, any additional time spent, potentially by a lesser trained person, could have worthwhile benefits in more accurate monitoring of curve progression and treatment planning.

The reliability of the King classification obtained in this study was superior to previously published series (Table 3), despite the fact that the classification was performed without premarking the radiographs, as in some previous studies.[10,12] Premarking has been identified as a significant factor in facilitating classification.[11] In the present study, the observers did not need to be coached or trained in the classification groupings, or to memorize them, because they only had to identify and mark the vertebral body and sacral landmarks. These comparisons between series, including the direct comparison with the study of Lenke *et al*[10] using the same radiographs used in this study but premarked with Cobb angles and the central sacral line, suggests that judgmental, rather than technical measurement errors, predominate in the King *et al* classification. Further evidence for this comes from the finding that the Cobb angle variability inherent in the computer-assisted method is comparable with that reported previously.

It has been noted[9] that there is ambiguity as to the use of lateral bending measurements in distinguishing between type 1 and 2 curves in the King classification, and the classification of these curve types is frequently performed based on the relative curve magnitudes present in a standing radiograph, as in this study. The validity of the computer-assisted algorithm was reported[9] by testing it against the examples given in King *et al*.[8]

Table 3. Mean Kappa Values Obtained for King *et al* Classification Reliability in Published Studies and in the Present Study

	Interobserver	Intraobserver
Lenke *et al*[10]	0.49	0.62
Cummings *et al*[13]	0.44	0.64
Richards *et al*[11]	0.61	0.81
Behensky *et al*[12]	0.46	0.79
Present study	0.82	0.85

The reproducibility of the Cobb angle measures obtained here appears equal to or better than previously reported.[1–6,16] However, direct comparisons cannot be made with the previous studies because different radiographs were evaluated, and differing statistical methods have been used in those studies to evaluate Cobb angle reproducibility. Some published reports premarked the end vertebrae, and some preselected good quality films or those having smaller curve magnitudes than in the presurgical group studied here. Oda *et al*[1] reported that 5 surgeons, measuring 50 radiographs, had an average error of 9° (calculated as twice the standard deviation) and that the main error source was in identifying end vertebrae. Morrissy *et al*[3] reported repeated measurements by 4 surgeons of 48 "good quality" radiographs of patients having Cobb angle in the range 20° to 40°. When the end vertebrae were not preselected, the standard deviation of paired differences was 2.4°. Carman *et al*[4] reported an average difference of 3.8° (95% of differences less than 8.0°) in repeated measurements by 5 readers of 8 radiographs. They inferred from analysis of variance components that the overall standard deviation was 2.97°. These findings indicated that a change in a Cobb angle measurement of less than 10° cannot be interpreted with confidence as a real change.

Goldberg *et al*[5] showed interobserver variability of 2.5° and intraobserver reliability of 1.9° in a study by 4 evaluators of the primary curve identified in 30 radiographs. They also reported that the interclass correlation coefficient for the Cobb angle was 0.98. The interobserver standard deviation was 2.8° and the intraobserver standard deviation was 1.8° in a study by Ylikoski and Tallroth[6] of Cobb angle measurements of 30 consecutive untreated patients having a mean Cobb angle of 24.4° by 2 readers using a specially designed angle-measuring instrument ("Plurimeter"). In the present study of patients with larger (preoperative) scoliosis, the average sample standard deviations of the Cobb angle were (intraobserver) 2.0° for upper and lower curves, and (interobserver) 2.5° and 2.6° for upper and lower curves, respectively.

There is some disagreement as to whether the precision of Cobb angle measurements is substantially improved when the end vertebrae are preselected[1,3] or not.[4,5] In the present study, the end vertebrae were selected automatically based on the values of endplate inclination calculated from the vertebral body landmarks. The accuracy of marking points on endplates was studied by Cheung *et al*,[17] who reported a coefficient of repeatability 0.8 and 1.3 mm in horizontal and vertical directions, respectively, suggesting an angular error of about 2° for the determination of each endplate inclination.

Here, the variability of the Cobb angle determination was not found to vary significantly with the radiographic quality. However, the image resolution (pixel size approximately 1 mm) was rather low, relative to original full-size films, and relative to that available in digital radiographs. Because a radiograph only records a patient's spinal shape at an instant of time, repeated radiographs would introduce additional variability because of differing radiographic technique, postural sway, *etc.*[16] For instance, Beauchamp *et al*[18] reported diurnal variation of Cobb angle measurement. In the present and most previous studies, the additional radiographic dose has precluded the use of repeat radiographs, and this additional source of variability is ignored.

In the assignment of the King *et al* classification, several factors have been noted previously as contributing to variability when a specific patient has a scoliosis deformity with features close to classification criteria.[9] These factors influenced the findings in the present study. They include the observed presence or absence of third (upper thoracic) curve, uncertainty as to whether a curve "crosses the midline," and the relative magnitude of thoracic and lumbar curves. Problems occur when a patient has a spinal shape very close to any of these criteria. These kinds of problems could occur in alternate classifications.

For example, the newer classification scheme developed by Lenke *et al*[14] has reliability characterized by kappa values in the range of 0.64–0.89,[11,14,19] with lower values if the radiographs were not premarked (*i.e.*, premeasured). The "lumbar modifier" used in this classification recognizes this possibility of a pedicle lying very close to the "cut-off" point, and reliability in identifying this feature is relatively high (kappa statistic equal to 0.89[19]). Nevertheless, factors such as variability in marking the central sacral line could still affect this judgment. The same kind of computer-assisted algorithmic approach as used in the present study could be applied to other classification systems that have precisely defined classification criteria, and this might improve their reliability. If the classification was taken as the sole factor in deciding the extent of a spinal arthrodesis for each patient, then the variation between observers and observations would alter the surgical plan. For instance, the difference between a type 5 and either a type 2 or 3 classification, if the detection of an upper thoracic curve was inconsistent, would influence whether the upper curve were fused.

As digital imaging and computer-assisted medical decision making become increasingly available, clinicians can increasingly turn to computerized tools to assist in analyzing, classifying, and treating patients with adolescent idiopathic scoliosis. Computerized tools can be helpful in the automated interpretation of data, as well as its storage and display. For evaluation of radiographs of patients with scoliosis, it would be beneficial to replace the traditional pencil, ruler, and protractor methods with interactive marking of landmark points, and having the display software also include formal algorithms for measurement and classification. This process can reduce technical errors, as well as the need for memorization of measurement and classification procedures.

■ Key Points

- Computerized algorithms can assist in the complex task of identifying end vertebrae of curves in radiographs of patients with scoliosis, measuring curve magnitudes, and in applying complex classification rules, potentially overcoming human judgmental errors.
- A tool using a computerized algorithm to facilitate Cobb angle measurements and King classifications had reliability superior to that achieved by unassisted individuals.
- Computer-assisted evaluation was performed equally well by lesser-trained individuals, and the time spent marking the films, not an observer's experience in treating patients with scoliosis, was the major determinant of accuracy and reliability.

Acknowledgments

Dr. L. Lenke kindly provided to us the radiographic images. The authors thank Randall Risinger, MD, Ketan Davae, MD, and Katherine Clark, BS, who marked the radiographs. Richard Single, PhD, advised on the statistical analysis of the data.

References

1. Oda M, Rauh S, Gregory PB, et al. The significance of roentgenographic measurement in scoliosis. *J Pediatr Orthop* 1982;2:378–82.
2. Diab KM, Sevastik JA, Hedlund R, et al. Accuracy and applicability of measurement of the scoliotic angle at the frontal plane by Cobb's method, by Ferguson's method and by a new method. *Eur Spine J* 1995;4:291–5.
3. Morrissy RT, Goldsmith GS, Hall EC, et al. Measurement of the Cobb angle on radiographs of patients who have scoliosis. Evaluation of intrinsic error. *J Bone Joint Surg Am* 1990;72:320–7.
4. Carman DL, Browne RH, Birch JG. Measurement of scoliosis and kyphosis radiographs. Intraobserver and interobserver variation. *J Bone Joint Surg Am* 1990;72:328–33.
5. Goldberg MS, Poitras B, Mayo NE, et al. Observer variation in assessing spinal curvature and skeletal development in adolescent idiopathic scoliosis. *Spine* 1988;13:1371–7.
6. Ylikoski M, Tallroth K. Measurement variations in scoliotic angle, vertebral rotation, vertebral body height, and intervertebral disc space height. *J Spinal Disord* 1990;3:387–91.
7. Lenke LG, Betz RR, Clements D, et al. Curve prevalence of a new classification of operative adolescent idiopathic scoliosis: Does classification correlate with treatment? *Spine* 2002;27:604–11.
8. 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.
9. Stokes IA, Aronsson DD. Identifying sources of variability in scoliosis classification using a rule-based automated algorithm. *Spine* 2002;27:2801–5.
10. 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.
11. Richards BS, Sucato DJ, Konigsberg DE, et al. Comparison of reliability between the Lenke and King classification systems for adolescent idiopathic scoliosis using radiographs that were not premeasured. *Spine* 2003;28:1148–56.
12. Behensky H, Giesinger K, Ogon M, et al. Multisurgeon assessment of coronal pattern classification systems for adolescent idiopathic scoliosis: Reliability and error analysis. *Spine* 2002; 27:762–7.
13. 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.
14. 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-A:1169–81.
15. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;20:37–46.
16. Pruijs JE, Stengs C, Keessen W. Parameter variation in stable scoliosis. *Eur Spine J* 1995;4:176–9.
17. Cheung J, Wever DJ, Veldhuizen AG, et al. The reliability of quantitative analysis on digital images of the scoliotic spine. *Eur Spine J* 2002;11:535–42.
18. Beauchamp M, Labelle H, Grimard G, et al. Diurnal variation of Cobb angle measurement in adolescent idiopathic scoliosis. *Spine* 1993;18:1581–3.
19. Ogon M, Giesinger K, Behensky H, et al. Interobserver and intraobserver reliability of Lenke's new scoliosis classification system. *Spine* 2002;27:858–62.