ORIGINAL ARTICLE

The reliability and validity of three non-radiological measures of thoracic kyphosis and their relations to the standing radiological Cobb angle

G. A. Greendale · N. S. Nili · M.-H. Huang · L. Seeger ·
A. S. Karlamangla

Abstract

Summary Hyperkyphosis is implicated in a mounting list of negative outcomes, including higher mortality. Hyperkyphosis research is hindered due to difficulties inherent in its measurement. By showing that three clinical measures of kyphosis are suitable for use in large scale, longitudinal, hyperkyphosis studies, we will facilitate much needed research in this field.

Introduction The objective of this study is to describe the reliability of three non-radiological kyphosis measures (Debrunner kyphosis angle, flexicurve kyphosis index, and flexicurve kyphosis angle) and their validity compared to the Cobb angle and to approximate a Cobb angle from non-radiological kyphosis measures.

Methods We analyzed data from 113 participants aged ≥60 years with kyphosis angle ≥40°. Cobb angle was measured on a standing lateral thoracolumbar radiograph using bounds at T4 and T12. Non-radiological measures of kyphosis were made three times by a single rater and a 4th time by a blinded second rater.

Results Intra- and inter-rater reliabilities for non-radiological assessments were high (intra-class correlations of 0.96 to 0.98) and did not differ from each other. Pearson correlations, estimating validity, ranged from 0.62 to 0.69 and did not differ. The Debrunner angle was close to the Cobb angle, with scaling factor of 1.067 and an offset of 5°. The Flexicurve kyphosis angle had to be scaled by 1.53 to obtain the equivalent Cobb angle. The scaling factor for the Flexicurve kyphosis index to Cobb angle was 315, with an offset of 5°. Compared to the measured Cobb angle, Cobb angles predicted using the non-radiological measures had similar magnitude errors (standard deviations of the differences ranging between 10.24 and 11.26).

Conclusions Each non-radiological measurement had similar reliability and validity. Low cost, ease of use, and robustness to variations in spine contour argue for the Flexicurve in longitudinal kyphosis assessments. The approximate conversion factors provided will permit translation of non-radiological measures to Cobb angles.

Keywords Cobb angle · Kyphosis · Reliability · Validity

G. A. Greendale (✉) · N. S. Nili · M.-H. Huang ·
A. S. Karlamangla
Division of Geriatrics, David Geffen School of Medicine,
University of California,
10945 Le Conte Avenue, Suite 2339,
Los Angeles, CA 90095-1687, USA
e-mail: ggreenda@mednet.ucla.edu

L. Seeger
Department of Radiology, Division of Musculoskeletal Imaging,
David Geffen School of Medicine at UCLA,
University of California,
Los Angeles, CA, USA

Introduction

Adverse consequences of hyperkyphosis (excessive thoracic kyphosis) include physical functional limitations [1–4], injurious falls [5], back pain [6], respiratory compromise [7], restricted spinal motion [8], fractures [9, 10], and mortality [11–13]. However, a recent randomized, controlled trial found that hyperkyphosis was remediable, encouraging further study of its prevention and treatment [14].

Impediments to large-scale hyperkyphosis research are the difficulties inherent in obtaining the criterion standard measurement, the modified Cobb angle [15–19], including expense, limited portability of X-ray equipment, X-ray exposure, and the time necessary to procure and read the radiographic image.

To facilitate hyperkyphosis research, investigators have developed inexpensive and X-ray-free kyphosis measures, such as the Debrunner kyphometer and the flexicurve ruler. The Debrunner kyphometer consists of a protractor mounted on two arms, the ends of which are positioned on specified bony landmarks; kyphosis angle is read from the protractor [6, 20]. While those with advanced training may readily recognize the landmarks, other research staff may have a difficult time accurately and reproducibly identifying the correct levels. The flexicurve ruler, gently pressed onto the back, adopts the thoracic and lumbar contours of the participant. The researcher then traces the ruler's retained shape onto paper and calculates the kyphosis index (Fig. 1) [21]. One can also calculate an inscribed angle of kyphosis from the tracing, using geometric formulae (Fig. 1) [14].

Although the non-radiological kyphosis measures minimize cost and obviate radiation, they have enjoyed limited adoption. One explanation may be that they are not calibrated to the Cobb angle, which limits their clinical interpretation. A metric that translates a non-radiological kyphosis result into an approximate Cobb angle would allow estimation of clinical severity from non-Cobb measures. Demonstrations of the reliability and validity of the non-radiological measures, especially in older persons, have been minimal, a possible second reason for limited use [13, 20, 22–24].

Therefore, we designed this study to describe: (1) the intra-rater and inter-rater reliability of three non-radiological kyphosis measures, the Debrunner kyphosis angle, the flexicurve kyphosis index, and the flexicurve kyphosis angle; (2) the validity of each non-radiological measure using the modified Cobb angle as the criterion standard; and (3) a translational formula that provides an approximate Cobb angle based on results of the non-radiological measures. We used baseline data from the Yoga for Kyphosis trial, during which we performed standing lateral radiographs to assess modified Cobb angle as well as multiple, same-day, intra-rater and inter-rater measures of the non-radiological assessments.
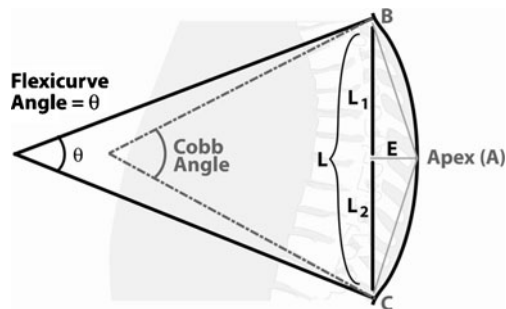
Methods

Participants

The analysis sample came from the Yoga for Kyphosis Trial, a single masked, randomized, controlled trial (RCT) of Yoga intended to improve thoracic hyperkyphosis [14]. The trial enrolled 118 participants aged ≥60 years with Debrunner kyphometer-assessed kyphosis angle ≥40°. Major RCT exclusions were: serious comorbidity; use of an assistive device; or unable to pass a movement-safety screen. Of 118 persons enrolled in the RCT, 113 had a standing radiological Cobb angle and at least one non-radiological assessment of kyphosis at RCT baseline, making them eligible for this analysis.

Kyphosis measurement

All kyphosis measures were made on the same day, within a 4-h window. The *modified Cobb angle*, based on the technique originally described by Cobb to quantify scoliosis, was measured on standing lateral thoracolumbar radiographs [17–19], specifying the limit vertebrae at T4 and T12 [18]. Because some radiographs did not permit use of specified limit vertebrae (e.g., due to overlying structures) Cobb angles from 20 films were based on eight vertebrae (T4–T11 or T5–T12) and Cobb angles from six films were based on seven vertebrae (T5–T11). Non-radiological measures of kyphosis included the Debrunner kyphometer angle, the Flexicurve kyphosis index, and the Flexicurve kyphosis angle. The upper arm of the *Debrunner kyphometer* was placed on C-7 and the lower arm on T-12. The circumscribed kyphosis angle was read from the protractor [6, 20]. Debrunner measurements were flagged as problematic in eight cases, because it was difficult to get the base of the arms flush on the landmarks. The *Flexicurve kyphosis index* was measured using a Flexicurve [21, 25]. The cephalic end of the Flexicurve was placed on C-7, and it was molded to the spine in the caudal direction. The shape was traced onto paper, and the apex kyphosis height was estimated relative to the length of the entire thoracic

Fig. 1 Three methods of quantifying thoracic kyphosis angles are illustrated. The modified T_4–T_{12} Cobb angle (*dotted lines*) measures the angle created by lines drawn parallel to the limit vertebrae visualized on a lateral standing thoracolumbar radiograph. In this case, the limit vertebrae are pre-specified at T4 and T12. The *Flexicurve kyphosis index and angle* are computed using measurements taken from the flexicurve tracing of the thoracic curve, represented here by the *solid dark curve* posterior to the thoracic vertebral bodies. To calculate the *Flexicurve kyphosis index*, the apex kyphosis height (E) is divided by the length of the entire thoracic curve (L). The *Flexicurve kyphosis angle*, Theta (θ), is calculated using *lines drawn perpendicular to the short sides of the triangle* inscribed by the thoracic curve. This *triangle* is demarcated by points **a** (Apex), **b** (at the cranial end of the curve), and **c** (at the caudal end). Theta equals arc tan (E/L_1) + arc tan (E/L_2)

spine; this is the Flexicurve kyphosis index (Fig. 1). Using geometric formulae, the *Flexicurve kyphosis angle* was also calculated from the Flexicurve tracing. By definition, this inscribed angle is systematically less than the circumscribed angle (Fig. 1).

Training and time required for non-radiological kyphosis measures

Research staff had baccalaureate degrees, but none had formal training in anatomy. Staff training consisted of an initial didactic and demonstration (with the aid of volunteer subjects) by Principal Investigator (GAG). It included: review of basic spine anatomy using illustrations; instruction in how to find landmarks by palpation; demonstration of the placement of the kyphometer and how to read the angle from the instrument's protractor; demonstration of how to apply the flexible ruler and how to make measurements from it. Each staff member then practiced identifying landmarks and conducting the measures. In aggregate, the didactics and staff practice took approximately 40 min. During the conduct of the study, each Debrunner measurement took between 1 and 2 min to make and record, depending on the degree of difficulty ascertaining landmarks. Each flexible ruler measure took 30 s to make; subsequent tracing of the shape on paper and taking the measurements to calculate the angle and index took 2.5 min.

Intra-rater and inter-rater reliability

Each clinical kyphosis assessment was made three times for each participant (with repositioning) by the same staff person; the average was the primary value. These three measures also permitted evaluation of intra-rater reliability. For inter-rater reliability, immediately following the first set of measures, one other masked research associate made a 4th assessment, with repositioning, in 54 participants. (Inter-rater sample size ranged from 51 to 54 due to missing values.)

Statistical analyses

We examined the within-rater, intra-class correlation coefficients (ICC = between-person variance divided by total variance) for each of the non-radiological kyphosis measures using the three measurements made on each participant by the primary rater. In the 54 participants in the inter-rater subset, who had paired ratings made by a single first and a single second rater, we compared the average of the three measures from the primary rater with the single measure from the secondary rater, calculating inter-rater ICCs. Both intra-rater and inter-rater ICCs were also examined after stratification by kyphosis severity, defined by Cobb angle median split: moderate if <53°, severe if ≥53°. To compare the non-

radiological kyphosis measures with the Cobb angle criterion standard, we examined Pearson correlations between each non-radiological measure and Cobb angle. These analyses were repeated after first excluding 26 participants whose Cobb angles did not span T4–T12 and then excluding seven individuals whose Debrunner measurements were flagged as problematic. In each of these samples, correlations were also examined after stratification by kyphosis severity. We created mathematical formulae to convert the non-radiological results to equivalent Cobb angles. Formulae were created by simple linear regression of the Cobb angle on each of the non-radiological measures in the sample that excluded participants whose Cobb angles did not span T4–T12 and whose Debrunner measurements were flagged as problematic. To test if Cobb angles measured using alternate landmarks had systematic error, in the 20 participants whose Cobb angle measurements spanned either T5–T12 or T4–T11, we compared the measured Cobb angle with the Cobb angle predicted by the clinical measures, using the paired *t* test. Finally, in the sample in which we derived the Cobb angle prediction equations (Table 5), we conducted Bland–Altman analyses. Bland–Altman analysis consists of the examinations of two graphs. The first graph is an identity plot, a scatter plot of the two measurements along with the line $y=x$. If the measurements agree closely, then the scatter plot points will line up near to the line $y=x$. The identity plot was produced only for measured Cobb angle and the measured Debrunner kyphosis angle, because they measure the same thing (circumscribed kyphosis angle) and use the same metric (degrees). The second graph is a Bland–Altman plot, a scatter plot of the variable's means plotted on the horizontal axis and the variable's differences plotted on the vertical axis; it includes approximate 95% confidence bands (the confidence bands assume normality of differences). The Bland–Altman plot illustrates the amount of disagreement between the measures being compared. Bland–Altman plots were created for the measured Cobb angle and each of the following: measured Debrunner kyphosis angle; Debrunner-predicted Cobb angle; Flexicurve kyphosis index-predicted Cobb angle; and Flexicurve kyphosis angle-predicted Cobb angle. The scientific importance of these differences is judged qualitatively; however, we also computed the standard deviation of the mean difference between the Cobb angle and each comparator to gauge the magnitude of the error [26].

Results

The mean age of the study sample was 75.3 years, average body mass index was 26.5, and 80.5% were women. These and other characteristics of the full sample and the inter-rater reliability sample are summarized in Table 1.

Table 1 Baseline demographic, behavioral and medical characteristics of study participants

Characteristic	Full sample (N=113)	Inter-rater reliability sample [a] (N=54)
Age (years)	75.3±7.5	75.5±7.7
Height (cm)	160.7±8.9	161.1±9.0
Weight (kg)	68.8±15.1	68.3±14.3
Body mass index (kg/m^2)	26.5±4.5	26.1±4.3
Female gender: % (N)	80.5 (91)	81.8 (45)
Usual physical activity	2.3±0.5	2.3±0.6
Chronic conditions (#)	5.6±3.8	5.4±2.9
Vertebral fractures [b,c]		
None % (N)	75.2 (85)	74.6 (41)
Thoracic % (N)	19.5 (22)	20.0 (11)
Lumbar % (N)	7.1 (8)	9.1 (5)

[a] All P values for full vs. inter-rater samples >0.05

[b] Percentage of lumbar and thoracic fractures sum to greater than 100% because some participants had fractures of both spinal regions

[c] Vertebral fractures defined as ≥25% decrement in interior, middle, or posterior vertebral body height

Shown in Table 2, the mean Cobb angle in the full sample was 53.76°. In the 87 cases with T4–T12 Cobb angles, the mean Cobb angle value was 55.43. Average Debrunner kyphosis angle was similar to the average Cobb angle. As expected, the inscribed flexicurve kyphosis angle averaged about 20° less than the circumscribed Cobb and Debrunner angles.

In the full sample, intra- and inter-rater reliabilities (ICCs) were uniformly high for all kyphosis assessments, 0.96 to 0.98 (Table 3). We also computed ICCs in subsamples, using the median value of the sample Cobb angle to define severity. Restriction of range in subsamples compared to the full sample systematically lowers the ICC value, but ICCs of the two subsamples can be compared to each other: reliabilities were similar in those with moderate and severe kyphosis. We also calculated the inter-rater reliability based on only the

first measurement from the rater one and the 4th from rater two; results did not differ (data not shown). Analyses excluding eight cases that were flagged for difficult kyphometer placement did not alter the intra- or inter-rater reliability estimates for that device (data not shown).

Table 3 Intra- and inter-rater reliabilities of three non-radiological kyphosis assessments

	Intra-rater reliability (N=113)	Inter-rater reliability[a] (N=51–54)
Full sample		
Debrunner kyphosis angle	0.98	0.98
Flexicurve kyphosis index	0.96	0.96
Flexicurve kyphosis angle	0.96	0.96
Moderate Kyphosis [b]		
Debrunner kyphosis angle	0.97	0.98
Flexicurve kyphosis index	0.94	0.93
Flexicurve kyphosis angle	0.94	0.94
Severe Kyphosis		
Debrunner kyphosis angle	0.97	0.98
Flexicurve kyphosis index	0.94	0.97
Flexicurve kyphosis angle	0.94	0.95

Values in table are intra-class correlation coefficients, defined as between-person variance divided by total variance

[a] The average of the first three measurements made by the first rater was compared to one measurement performed by the second rater

[b] Moderate kyphosis is defined as a Cobb angle of less than 53°, the sample median. Severe kyphosis is defines as a Cobb angle of greater than or equal to 53°

Table 2 Average values and distributions of standing Cobb angle and non-radiological kyphosis measurements

Kyphosis measurement	Sample size	Mean	Standard deviation	Median
Cobb angle, entire sample[a] (degrees)	113	53.76	14.76	53.10
Cobb angle, subset in which T4–T12 landmarks were used (degrees)	87	55.43	13.62	53.1
Debrunner kyphosis angle (degrees)	113	57.68	9.60	58.00
Flexicurve kyphosis index	113	0.162	0.033	0.161
Flexicurve kyphosis angle [b] (degrees)	113	36.50	6.82	36.48

[a] Cobb angle in the entire study sample includes 26 cases in which the desired T4–T12 landmarks could not be used, requiring alternate landmarks (see Methods for details)

[b] The Flexicurve kyphosis angle is an inscribed angle, which by definition will be smaller than the circumscribed angles estimated using the Cobb or Debrunner methods

Table 4 Validity of three non-radiological measurements of kyphosis compared to the Cobb angle criterion standard

Non-radiological kyphosis measurement and kyphosis severity	Full sample	Cobb-restricted sample[a]	Cobb and Debrunner-restricted samples[b]
Full range of Kyphosis	(N=113; Std error=0.094)	(N=87; Std error=0.107)	(N=80;Std error=0.112)
Debrunner kyphosis angle	0.622	0.715	0.762
Flexicurve kyphosis index	0.686	0.725	0.756
Flexicurve kyphosis angle	0.686	0.721	0.758
Moderate Kyphosis[c]	(N=55; Std error=0.135)	(N=41; Std error=0.156)	(N=37 ;Std error=0.164)
Debrunner kyphosis angle	0.275	0.354	0.405
Flexicurve kyphosis index	0.335	0.426	0.428
Flexicurve kyphosis angle	0.328	0.397	0.406
Severe Kyphosis	(N=58 ;Std error=0.131)	(N=46;Std error=0.149)	(N=43; Std error=0.152)
Debrunner kyphosis angle	0.447	0.602	0.641
Flexicurve kyphosis index	0.517	0.600	0.597
Flexicurve kyphosis angle	0.532	0.626	0.627

Values in table are Pearson correlation coefficients for each non-radiological measure compared to the Cobb angle

[a] Cobb-restricted sample excludes data from subjects whose Cobb angles did not span T4–T12

[b] Cobb and Debrunner-restricted sample excludes data from subjects whose Cobb angles did not span T4–T12 and those whose Debrunner kyphometer measures were flagged as difficult (see Methods for details)

[c] Moderate kyphosis is defined as a Cobb angle of less than 53°, the sample median. Severe kyphosis is defines as a Cobb angle of greater than or equal to 53°

The modified Cobb angle was our criterion measurement; non-radiological measures were compared to it to gauge their validity (Table 4). In the full sample, the Pearson correlations between the non-radiological kyphosis measures and the Cobb angle ranged from 0.62 to 0.69 (95% confidence Interval [CI] for each estimate was ±0.184). Correlations between each non-radiological measure in the 87 persons with T4–T12 Cobb angles were approximately 0.72, somewhat higher than the correlations based on the entire sample. In the sample that was also restricted to those whose Debrunner measures were not flagged as difficult (N=80), the Pearson correlations between the clinical kyphosis measures and the Cobb angle were even higher, and ranged from 0.762 to 0.758. In aggregate, there was a trend towards

Table 5 Calibration of non-radiological kyphosis measurements to theT4–T12 Cobb angle (n=80)

Non-radiological kyphosis measurements	β coefficient	Intercept	R^2
Debrunner kyphosis angle	1.067	−5.40	0.58
Flexicurve kyphosis index	314.61	5.11	0.57
Flexicurve kyphosis angle	1.53	0.30	0.57

Results in table are from simple linear regression, with T4–T12 Cobb angle as outcome and each non-radiological measure as predictor. To convert a non-radiological measure to equivalent T4–T12 Cobb angle, scale by corresponding β and add intercept

Calibration was performed using a sample restricted to persons with a T4–T12 Cobb angle and a Debrunner kyphometer measurement that was not flagged as difficult (see Methods for details)

higher correlations as the samples were progressively restricted. Comparing the severity subsamples, correlations between each non-radiological measure and the Cobb angle were somewhat higher in those with severe compared to those with moderate hyperkyphosis, but overlapping CIs did not support a statistically significant difference between them.

Non-radiological tests were calibrated to the Cobb angle, using linear regression: the T4–T12 Cobb angle was the outcome and each non-radiological kyphosis measure was the predictor (Table 5). The R^2 was 0.57–0.58 for each of the measures. Except for a systematic bias of about 5°, the Debrunner kyphosis angle was very similar to the Cobb angle: the beta coefficient, or scaling factor, to convert Debrunner angle to Cobb angle was 1.067. As expected, the flexicurve angle was systematically smaller than the Cobb angle; it had to be scaled by 1.53 to get the equivalent Cobb angle. The kyphosis index may also be approximated to the Cobb angle by using the conversion factor (about 315) and an offset of about 5°.

In the 20 individuals with Cobb angle measurements that spanned one less vertebral body (i.e., T4–T11 or T5–T12), mean Cobb angle was smaller than the Cobb angle predicted by the clinical kyphosis measures by about 8° in each case (data not shown), indicating that when the Cobb angle measure spans fewer vertebral bodies, the Cobb angle is systematically underestimated.

An identity plot graphically displays the agreement between the measured Cobb angle and the Debrunner angle (Fig. 2a). To graphically portray the disagreement between

the kyphosis measures, Bland–Altman plots, scatter plots of the variable means on the horizontal axis and the variable differences on the vertical axis, were created. These plots include approximate 95% confidence bands. We also computed the standard deviation (SD) of the mean difference between the Cobb angle and each comparator to gauge the magnitude of the error. Figure 2b, c, displays Bland–Altman plots for the measured Cobb angle and each of the following: measured Debrunner kyphometer angle (SD of mean difference, 11.4); Cobb angle-predicted using the Debrunner angle (SD of mean difference, 10.96); Cobb angle-predicted using the Flexicurve kyphosis index (SD of mean difference, 11.26); and Cobb angle-predicted using the Flexicurve kyphosis angle (SD of mean difference, 10.24).

Discussion

The overarching goals of this study were to calculate the reliability and validity of the Debrunner kyphometer angle, flexicurve kyphosis index, and flexicurve kyphosis angle and to calibrate each to the Cobb angle. Intra- and inter-rater reliabilities for the three non-radiological kyphosis assessments were uniformly high (0.96 to 0.98) and did not
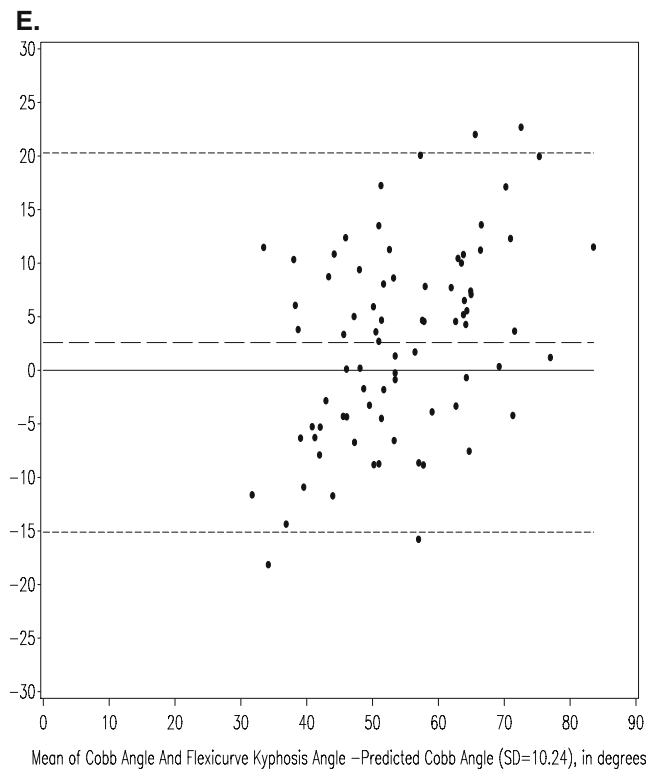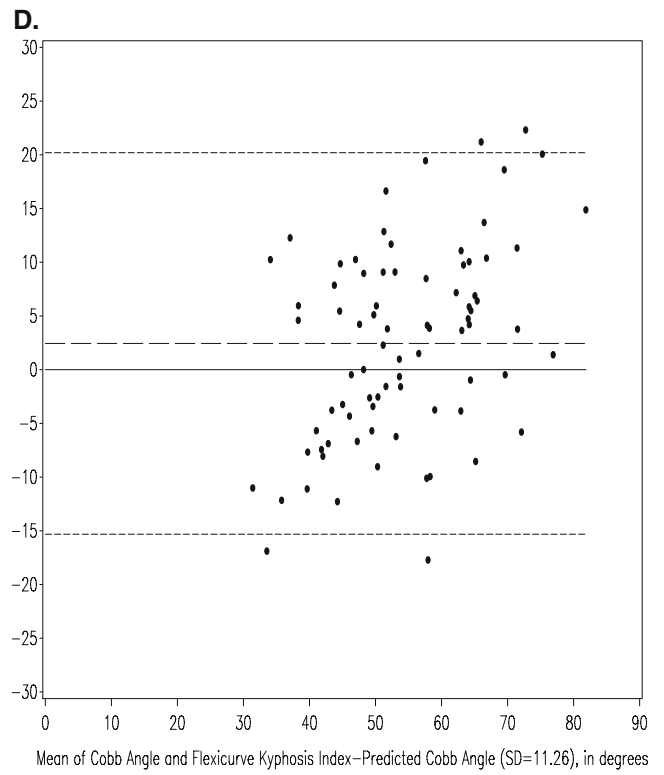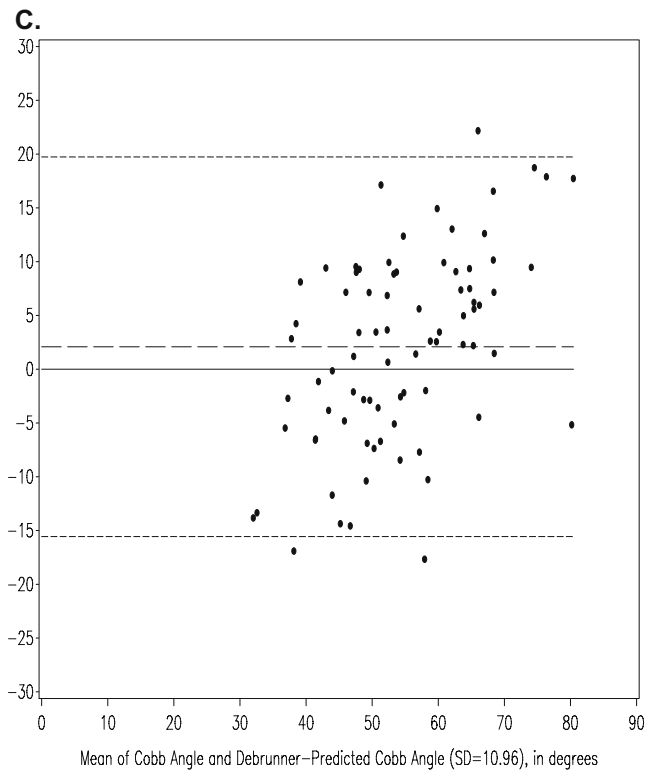
differ statistically from each other. Comparing the non-radiological kyphosis measurements to the Cobb angle also yielded validity estimates that were not distinguishable; all correlations were moderate (0.62 to 0.69). Our derived regression equations that scaled the non-radiological kyphosis estimates to the Cobb angle had robust R^2 values, between 0.57 and 0.58.

This study's high inter-rater and intra-rater reliabilities of Debrunner kyphometer and the Flexicurve kyphosis index, based on ICC values, mirrored reliabilities developed in a sample of 26 postmenopausal women with osteoporosis (but whose age range and degree of kyphosis was not specified); in that sample, inter-rater and intra-rater ICCs between 0.89 and 0.99 were found for each test [20]. The present analysis expands upon prior work by including a greater sample size, older subjects (in whom measurements may be more challenging), and a broad range of kyphosis over which reliabilities were assessed. The two studies agree, however: inter- and intra-rater reliabilities approach perfect and do not differ between the Debrunner kyphometer and the Flexicurve kyphosis index [27]. Although Ohlen examined reliability of the Debrunner kyphometer in 31 young volunteers and Ettinger tested reliability of the Flexicurve kyphosis index in 75 women aged 65–91 years, these two studies used

A.

B.
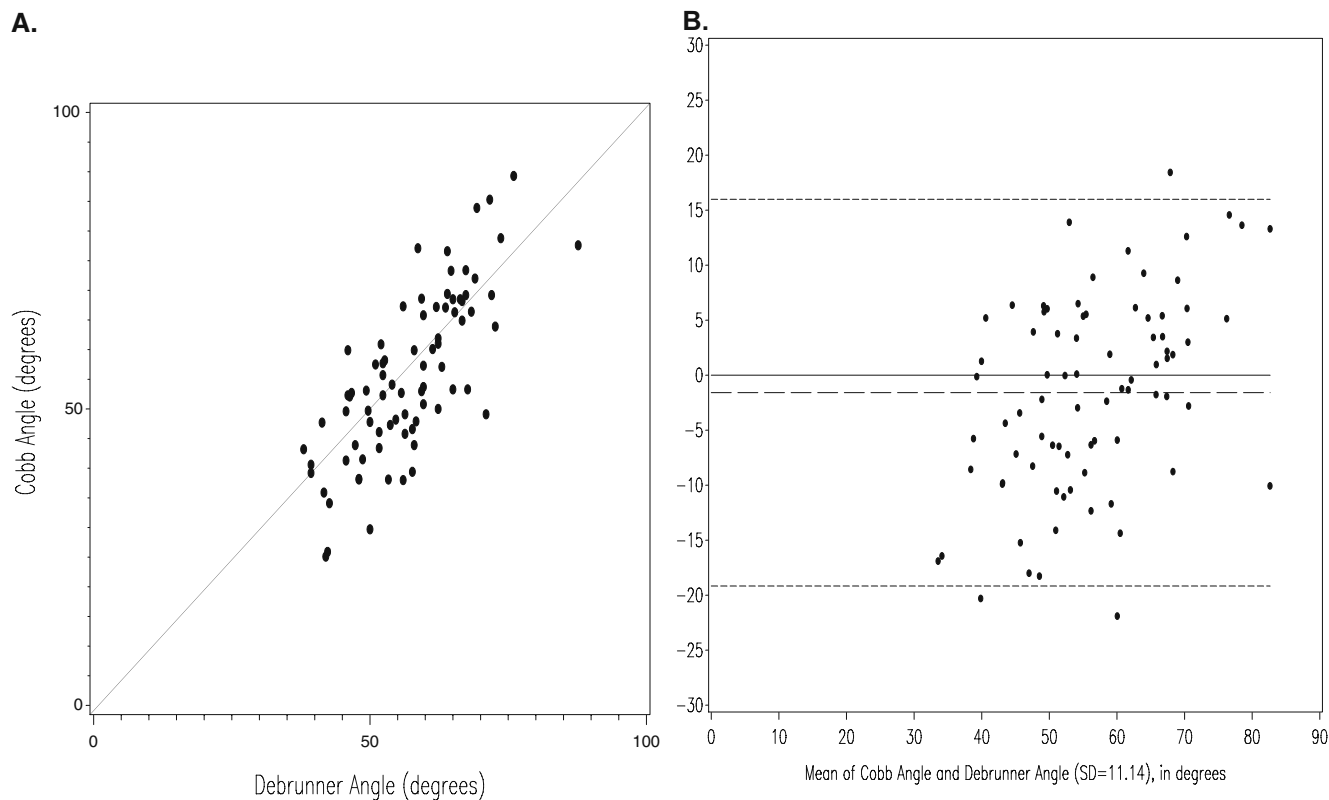
Fig. 2 Identity plot of the measured Cobb angle and the measured Debrunner angle (**a**). Bland–Altman plots of the measured Cobb angle and each of the following: measured Debrunner angle (**b**); Cobb angle predicted using the Debrunner angle (**c**); Cobb angle predicted using the Flexicurve kyphosis Index (**d**); and Cobb angle predicted using the Flexicurve kyphosis angle (**e**). Bland–Altman plots include approximate 95% confidence bands and also provide the SD of the difference between the Cobb angle and each comparator. Please see Methods for details

C.

x-axis: Mean of Cobb Angle and Debrunner−Predicted Cobb Angle (SD=10.96), in degrees

D.

x-axis: Mean of Cobb Angle and Flexicurve Kyphosis Index−Predicted Cobb Angle (SD=11.26), in degrees

E.

x-axis: Mean of Cobb Angle And Flexicurve Kyphosis Angle −Predicted Cobb Angle (SD=10.24), in degrees

Fig. 2 (continued)

different statistical methods to quantify reliability than those used in the present study, precluding direct comparison of their reliability estimates to ours [22, 24].

To our knowledge, published work has not reported the validity of the Debrunner kyphometer or the Flexicurve kyphosis index compared to the standing Cobb angle. Based on a sub-sample of 120 women from the Fracture Intervention Trial, Kado et al. calculated an ICC of 0.68 for the kyphosis index compared to a supine Cobb angle; however, the supine position would be expected to lessen the angle of kyphosis and lower the validity estimate [28].

Creating a mathematical formula that approximates Cobb angle based on a non-radiological kyphosis measure is not a novel idea and its value in avoiding radiation and facilitating longitudinal measurement has been recognized [23]. However, cross-calibration has been done only for the Debrunner instrument in an adolescent sample [23]. The present study offers metrics that allow researchers and clinicians to scale the Debrunner angle, Flexicurve kyphosis index, and the newly developed Flexicurve kyphosis angle to a standing radiological Cobb angle in adults with hyperkyphosis. For example, the Flexicurve kyphosis index–Cobb translations could enhance the interpretation of an important finding from the Study of Osteoporotic Fractures (SOF): that greater Flexicurve kyphosis indices predicted higher mortality independently of vertebral fracture [13]. It is now possible to approximate the Cobb angles that these indices represented: using the current study's metric, the SOF sample's mean predicted Cobb angle would be 43.8° (standard deviation, 10.7). Thus, the relative mortality hazard per kyphosis index standard deviation developed in SOF can be roughly translated to a 15% increase in mortality per each 10.7° increment in Cobb angle.

This study intended to inform deliberations about which of the three non-radiological tests used in the Yoga for Kyphosis project might be best suited to large observational or interventional kyphosis studies, in which sizable numbers of participants would be evaluated at multiple times. Because these types of studies necessitate multiple raters, the first consideration is the inter- and intra-rater reliabilities. On this basis, all three assessments performed nearly perfectly and equally. A second basis for ranking the three tests is validity, but this also did not discriminate among them. Finally, when compared to the criterion standard measured Cobb angle, Cobb angles predicted using each of the non-radiological measures had similar magnitude errors according to the Bland–Altman plots. Therefore, factors such as simplicity of use and sensitivity to anatomical variability may suggest the most favorable approach. The flexicurve may be easier for research staff without medical training, as it does not require identification of caudal landmarks. The flexicurve traces the contour of the entire spine; the inflection points between the cervical lordosis,

thoracic kyphosis, and lumbar lordosis define the spinal curves. In contrast, the Debrunner kyphometer must be placed on palpated landmarks [6]. Despite careful protocols, the inferior landmark can be particularly difficult to discern, especially when lumbar lordosis has reversed [21]. The Cobb and Debrunner angles base their measurements entirely on the two ends of the spinal curve. If there are no problems at these locations (such as endplate tilt of limit vertebrae or difficult Debrunner placement), dependence on the terminal portions of the curve will not be strongly influential [29]. However, when anatomical abnormalities are present, then an instrument such as the Flexicurve, which uses the entire spinal contour, will be more robust because deformities in part of the spine will not introduce large errors. In this regard, the Flexicurve is akin to the centroid angle, which computes kyphosis using the midpoints of all vertebral bodies from T1–T12 [29]. Indicative of the error introduced by difficult landmark determination was the trend toward higher a correlation between the Debrunner and Cobb angles when eight individuals with difficult Debrunner measures were omitted from the validity computation (Table 4).

Use of the T4–T12 constrained Cobb angle had merits and limitations. In favor of the constrained Cobb is that the uppermost thoracic vertebrae are often poorly visualized due to overlying tissue density. Another attribute of the constrained technique is that the identification of the most inclined vertebral body, which marks the transition from the thoracic to the lumbar curves, can be difficult, leading to low intra-rater reliability for determination of limit vertebrae, a problem circumvented by using the constrained Cobb technique [30, 31]. It must be acknowledged that the constrained method will misestimate the true kyphosis angle when the transition vertebra is not at the same level as the specified level. In aggregate, the potential measurement errors in the Cobb angle degrade the accuracy of the criterion standard, conservatively biasing this study's validity estimates.

The reliability and validity estimates of the non-radiological measures of kyphosis calculated from this sample cannot be assumed to apply to all instances in which these measuring devices are used; they are not immutable characteristics of the tests themselves [32]. Deterioration of reliability and validity may occur due to subject characteristics (e.g., obesity hampers landmark location) or to operator characteristics (e.g., staff capability). Because the research associates who performed the measures in the current study had no formal training in anatomy and likely comparable to other entry-level research or clinical staff, we believe that operator characteristics are unlikely to be influential in other settings.

The metrics developed in this study to scale the non-radiological tests to the standing Cobb angle must be viewed as approximations, intended to give investigators and clinicians a "feel" for what the values of the non-radiological tests mean in Cobb angle terms. They are not

intended to translate individual patient's non-radiological measures to Cobb angle values in clinical practice. Rather, these approximate conversion formulae are meant to help researchers get a handle on what the non-radiological tests mean in Cobb angle terms, which will inform the *general* clinical translation of research results.

In summary, in our study sample, we found that the Debrunner kyphometer, the flexicurve kyphosis angle and the flexicurve kyphosis index had strong and similar validity and reliability. Its low cost, ease of use by entry-level research staff, short measurement time, and relative robustness to variations in spine contour and deformity argue for use of the Flexicurve in longitudinal assessments of kyphosis. This study also provides approximate conversion factors that permit translation of results from three non-radiological kyphosis measures to an approximate Cobb angle value, which will assist researchers in interpreting the clinical meaning of the non-radiological tests.

References

1. Chow RK, Harrison JE (1987) Relationship of kyphosis to physical fitness and bone mass on post-menopausal women. Am J Phys Med 66:219–227
2. Ryan SD, Fried LP (1997) The impact of kyphosis on daily functioning. J Am Geriatr Soc 45:1479–1486
3. Kado DM, Huang MH, Barrett-Connor E, Greendale GA (2005) Hyperkyphotic posture and poor physical functional ability in older community-dwelling men and women: the Rancho Bernardo Study. J Gerontol A Biol Sci Med Sci 60:633–637
4. Takahashi T, Ishida K, Hirose D et al (2005) Trunk deformity is associated with a reduction in outdoor activities of daily living and life satisfaction in community-dwelling older people. Osteoporos Int 16:273–279
5. Kado DM, Huang MH, Nguyen CB et al (2007) Hyperkyphotic posture and risk of injurious falls in older persons: the Rancho Bernardo Study. J Gerontol A Biol Sci Med Sci 62(6):652–657
6. Ensrud KE, Black DM, Harris F et al (1997) Correlates of kyphosis in older women. J Am Geriatr Soc 45:682–687
7. Leech JA, Dulberg C, Kellie S et al (1990) Relationship of lung function to severity of osteoporosis in women. Am Rev Respir Dis 141:68–71
8. Miyakoshi N, Itoi E, Kobayashi M, Kodama H (2003) Impact of postural deformities and spinal mobility on quality of life in postmenopausal osteoporosis. Osteoporos Int 14(12):1007–1012
9. McGrother CW, Donaldson MM, Clayton D et al (2002) Evaluation of a hip fracture risk score for assessing elderly women: the Melton Osteoporotic Fracture (MOF) study. Osteoporos Int 13(1):89–96
10. Huang MH, Barrett-Connor E, Greendale GA et al (2006) Hyperkyphotic posture and risk of future osteoporotic fractures: the Rancho Bernardo study. J Bone Miner Res 21(3):419–423
11. Milne JS, Williamson J (1983) A longitudinal study of kyphosis in older people. Age Ageing 12(3):225–233
12. Kado DM, Huang MH, Greendale GA et al (2004) Hyperkyphotic posture predicts mortality in older community-dwelling men and women: a prospective study. JAGS 52:1662–1667
13. Kado DM, Lui LY, Ensrud KE et al (2009) Hyperkyphosis Predicts mortaility independent of vertebral osteoporosis in older women. Ann Intern Med 150:681–687
14. Greendale GA, Huang MH, Karlamangla AS et al (2009) Yoga decreases in senior women and men with adult-onset hyperkyphosis: results of a randomized controlled trial. JAGS 57:1569–1579
15. Fon GT, Pitt MJ, Thies AC (1980) Thoracic kyphosis: range in normal subjects. Am J Roentgenol 134:979–983
16. Voutsinas SA, MacEwan GD (1986) Saggital profiles of the spine. Clin Orthop 210:235–242
17. Cobb JR (1948) Outline for the study of scoliosis. Instr Course Lect 5:261–268
18. Singer KP, Edmondston SJ, Day RE et al (1994) Computer-assisted curvature assessment and Cobb angle determination of the thoracic kyphosis. Spine 19:1381–1384
19. Harrison DE, Cailliet R, Harrison DD et al (2002) Reliability of Centroid, Cobb and Harrison posterior tangent methods: which to choose for analysis of thoracic kyphosis. Spine 26:E227–E234
20. Lundon KM, Li AM, Bibershtein S (1998) Interrater and intrarater reliability in the measurement of kyphosis in postmenopausal women with osteoporosis. Spine 23:1978–1985
21. Milne JS, Lauder IJ (1976) The relationship of kyphosis to the shape of vertebral bodies. Ann Hum Biol 3:173–179
22. Ohlén G, Spangfort E, Tingvall G (1989) Measurement of spinal sagital configuration and mobility with Debrummer's kyphometer. Spine 14(6):580–583
23. Korovessis P, Petsinis G, Papazisis Z et al (2001) Prediction of thoracic kyphosis using the Debrunner kyphometer. J Spinal Disord 14(1):67–72
24. Ettinger B, Black DM, Palermo L et al (1994) Kyphosis in older women and its relation to back pain, disability and osteopenia: the study of osteoporotic fractures. Osteoporos Int 4:55–60
25. Milne JS, Lauder IJ (1974) Age effects in kyphosis and lordosis in adults. Ann Hum Biol 1:327–337
26. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1(1846):307–310
27. Landis JR, Koch GG (1977) The measurement of observer agreement for the categorical data. Biometrics 33:159–174
28. Kado DM, Christianson L, Palermo L et al (2006) Comparing a supine radiologic versus standing clinical measurement of kyphosis in older women: the Fracture Intervention Trial. Spine 31(4):463–467
29. Briggs AM, Wrigley TV, Tully EA et al (2007) Radiographic measures of thoracic kyphosis in osteoporosis: Cobb and vertebral centroid angles. Skeletal Radiol 36:761–767
30. Mac-Thiong JM, Pinel-Giroux FM, de Guise JA, Labelle H (2007) Comparison between constrained and non-constrained Cobb techniques for the assessment of thoracic kyphosis and lumbar lordosis. Eur Spin J 16:1325–1331
31. Potter BK, Rosner MK, Lehman RA Jr et al (2005) Reliability of end, neutral and stable vertebrae identification in adolescent idiopathic scoliosos. Spine 30(14):1658–1663
32. Streiner DL, Norman GR (2006) "Precision" and "accuracy": two terms that are neither. J Clin Epidemiol 59(4):327–330