

Intra- and inter-reader reliability of semi-automated quantitative morphometry measurements and vertebral fracture assessment using lateral scout views from computed tomography

Y. M. Kim · S. Demissie · R. Eisenberg ·
E. J. Samelson · D. P. Kiel · M. L. Bouxsein

Received: 10 August 2010 / Accepted: 17 December 2010
© International Osteoporosis Foundation and National Osteoporosis Foundation 2011

Abstract

Summary Intra- and inter-reader reliability of semi-automated quantitative vertebral morphometry measurements was determined using lateral computed tomography (CT) scout views. The method requires less time than conventional morphometry. Reliability was excellent for vertebral height measurements, good for height ratios, and comparable to semi-quantitative grading by radiologists for identification of vertebral fractures.

Y. M. Kim (✉) · R. Eisenberg · M. L. Bouxsein
Center for Advanced Orthopaedic Studies,
Beth Israel Deaconess Medical Center,
Boston, MA, USA
e-mail: ymkim73@gmail.com

Y. M. Kim · R. Eisenberg · M. L. Bouxsein
Department of Orthopaedic Surgery, Harvard Medical School,
Boston, MA, USA

Y. M. Kim
Department of Internal Medicine, NHIC Ilsan Hospital,
Goyang-shi, Gyeonggi-do, Korea

S. Demissie
Department of Biostatistics,
Boston University School of Public Health,
Boston, MA, USA

E. J. Samelson · D. P. Kiel
Institute for Aging Research, Hebrew SeniorLife,
Boston, MA, USA

E. J. Samelson · D. P. Kiel
Department of Medicine, Harvard Medical School,
Boston, MA, USA

E. J. Samelson · D. P. Kiel
Division of Gerontology, Beth Israel Deaconess Medical Center,
Boston, MA, USA

Introduction Underdiagnosis and undertreatment of vertebral fracture (VFX) is a well-known problem worldwide. Thus, new methods are needed to facilitate identification of VFX. This study aimed to determine intra- and inter-reader reliability of semi-automated quantitative vertebral morphometry based on shape-based statistical modeling (SpineAnalyzer, Optasia Medical, Cheadle, UK).

Methods Two non-radiologists independently assessed vertebral morphometry from CT lateral scout views at two time points in 96 subjects (50 men, 46 women, 70.3 ± 8.9 years) selected from the Framingham Heart Study Offspring and Third Generation Multi-Detector CT Study. VFXs were classified based solely on morphometry measurements using Genant's criteria. Intraclass correlation coefficients (ICCs), root mean squared coefficient of variation (RMS CV) and kappa (k) statistics were used to assess reliability.

Results We analyzed 1,246 vertebrae in 96 subjects. The analysis time averaged 5.4 ± 1.7 min per subject (range, 3.2–9.1 min). Intra- and inter-reader ICCs for vertebral heights were excellent (>0.95) for all vertebral levels combined. Intra- and inter-reader RMS CV for height measurements ranged from 2.5% to 3.9% and 3.3% to 4.4%, respectively. Reliability of vertebral height ratios was good to fair. Based on morphometry measurements alone, readers A and B identified 51–52 and 46–59 subjects with at least one prevalent VFX, respectively, and there was a good intra- and inter-reader agreement ($k=0.59$ – 0.69) for VFX identification.

Conclusions Semi-automated quantitative vertebral morphometry measurements from CT lateral scout views are convenient and reproducible, and may facilitate assessment of VFX.

Keywords Computed tomography · Reliability · Semi-automated vertebral morphometry · Vertebral fracture

Abbreviations

ABQ	Algorithm-based approach for the qualitative identification of vertebral fracture
BMI	Body mass index
CI	Confidence interval
CT	Computed tomography
CV	Coefficient of variation
d_B	Biconcave deformity percentage
d_C	Crush deformity percentage
d_W	Wedge deformity percentage
DXA	Dual energy X-ray absorptiometry
h_A	Height anterior
h_M	Height mid
h_P	Height posterior
ICCs	Intraclass correlation coefficients
K	Kappa
min	Minutes
MRX	Morphometric radiography
MXA	Morphometric X-ray absorptiometry
QCT	Quantitative computed tomography
QM	Quantitative vertebral morphometry measurement
r_B	Biconcave ratio
r_C	Crush ratio
r_W	Wedge ratio
RMS	Root mean squared CV
CV	
SQ	Semi-quantitative
s	Seconds
VFx	Vertebral fracture

Introduction

Vertebral fractures are the hallmark of osteoporosis, being the most prevalent osteoporotic fracture. Moreover, the presence of vertebral fracture is among the strongest risks for future subsequent vertebral and non-vertebral fracture [1–3]. Vertebral fractures cause direct functional impairment as well as future disabling fractures and increased mortality [4–6]. Despite the high prevalence and significant impact, underdiagnosis and undertreatment of vertebral fractures is a worldwide problem [7]. Less than one third of individuals with vertebral fracture receive medication or proper treatment [8–11]. In particular, mild fractures are often overlooked or unrecognized on routine lateral chest or lateral spine radiographs [12, 13]. New methods are urgently needed to improve the accuracy and efficiency of identifying vertebral fractures so that proper therapeutic intervention can be initiated.

There are three general approaches for identifying prevalent vertebral fractures from lateral spinal radiographs, lateral dual-energy X-ray absorptiometry (DXA), or lateral

scout views from quantitative computed tomography (QCT) imaging: (1) visual semi-quantitative (SQ) grading developed by Genant et al. [14], (2) algorithm-based approach for the qualitative identification of vertebral fracture (ABQ) developed by Jiang et al. [15], and (3) quantitative vertebral morphometry measurement (QM), which was first developed in the 1960s [16]. Although the SQ and ABQ methods are currently considered as the gold standard methods for vertebral fracture assessment, they have limitations, including the requirement for highly experienced readers and modest reproducibility between readers, particularly for fractures of mild severity [17, 18]. In comparison, since the QM method relies on direct measurements of vertebral dimensions achieved by placement of six points on the vertebral body by a trained reader, disadvantages include the major time commitment to place six points on each vertebral body and potential variability in point placement by different readers. Thus, a combination of SQ and QM approaches may enhance the strengths of each technique while minimizing their limitations, possibly improving the identification of vertebral fractures [19, 20].

Recent development of shape-based statistical modeling technology for semi-automated quantitative morphometry may make QM measurements more feasible. One of these algorithms (SpineAnalyzer, Optasia Medical, Cheadle, UK) not only allows rapid, semi-automated placement of the standard six vertebral morphometry points, but also identifies 95 points to carefully delineate the shape of each vertebral body [21]. Thus, it facilitates QM methods by eliminating the manual annotation of six points on each vertebra, and it also has the potential to provide novel information about vertebral shape. A previous study using lateral radiographs reported excellent accuracy and reproducibility for semi-automated vertebral morphometry measurements using this semi-automated algorithm. In particular, the mean accuracy error of vertebral height measurements was 1.06 ± 1.2 mm, which corresponds to 3.4% of vertebral height on average, and the mean precision error, reflecting inter-observer variability, was 0.61 ± 0.73 mm, corresponding to 2.3% of vertebral height [21]. However, this study was limited, in that it did not report inter- and intra-reader reliability of vertebral heights, height ratios, and vertebral fracture classifications based on the morphometry measurements.

Our long-term goal is to enhance assessment of vertebral fractures by determining the clinical utility of these novel semi-automated quantitative morphometry measurements. Specifically, in this study we determined intra- and inter-reader reliability of semi-automated vertebral morphometry measurements and fracture assessment using lateral scout views from QCT. In addition, we compared the time required to complete the morphometry analysis in subjects with and without fractures.

Methods

Subjects

The participants used in this study are identical to those used by Samelson et al. [22] to determine the reliability of vertebral fracture assessment from lateral QCT scout views using Genant's semi-quantitative algorithm [14]. Specifically, 100 participants were selected from the community-based Framingham Heart Study Offspring and Third Generation Multi-detector computed tomography (MDCT) study [23]. QCT scans of the chest and abdomen were acquired in 3,529 participants in the MDCT study for assessment of coronary and aortic calcium [24]. Subjects selected for the current study included 50 men and 50 women, ranging in age from 50 to 87 years. To ensure an adequate number of individuals with vertebral fractures in our reliability study, a clinical investigator experienced in evaluating lateral spine images (DPK) reviewed the scout films of persons age 70 years and older to identify at least 16 individuals with suspected vertebral fracture. In these 16 individuals, there were 30 vertebral fractures, including 12 mild (SQ 1), 14 moderate (SQ 2), and 4 severe (SQ 3) fractures. Seven individuals had one vertebral fracture, one had two vertebral fractures, three had three vertebral fractures, and three had four vertebral fractures. We then selected a convenience sample of an additional 84 subjects to obtain 100 individuals.

Computed tomography scans

QCT scans were acquired using an eight-slice multi-detector QCT scanner (Lightspeed Ultra, General Electric Medical Systems, Milwaukee, WI), as previously described [24]. The scout views used in this study consisted of frontal and lateral low energy 2D scanograms extending from the upper thoracic (T4) to sacral (S1) vertebral levels (Fig. 1).

Semi-automated quantitative vertebral morphometry

Semi-automated quantitative vertebral morphometry was performed using a model-based shape recognition technology that provides standard six-point morphometry, plus detailed annotation to define the shape of each vertebra between T4 to L4 (Fig. 1). DICOM images were loaded and displayed, and vertebrae of interest from T4 to L4 were

labeled by the operator by manually placing points in the approximate center of each vertebra. Using these points, the algorithm then automatically identifies vertebral body margins, draws contours, and places points for standard six-point morphometry. The operator reviews the images, and if necessary, manually adjusts the point placement. The program computes vertebral heights, height ratios, and deformities indicative of vertebral fracture. The semi-automated measurements can be made on lateral images from various sources, including radiographs, DXA-based VFA, and lateral CT scout views [25, 26].

Study design

To determine inter-and intra-reader reliability, vertebral morphometry measurements were performed by two non-radiologist readers, at two different time points (>2 weeks apart) using the semi-automated algorithm. Thus, four evaluations were independently performed for each subject's scout view image. Readers were blinded to subject identification number and age. In addition, we also determined reproducibility of vertebral heights and height ratios using the unadjusted morphometry point placement (i.e., no operator intervention). In this case, one non-radiologist reader performed vertebral morphometry at two time points, and both times did not apply any manual intervention to vertebral contours or point placement.

Vertebral morphometry measurements and assessment of vertebral fracture

We measured posterior (h_P), anterior (h_A), and mid-vertebral height (h_M). We used height ratios and deformity percentages from six morphometry points to classify fractures as wedge, biconcave, and crush fractures, the latter being based on Black et al. [27]. The equations for vertebral height ratios and deformity percentages are as follows: where h_A is the anterior height for the vertebral body at the current level, h_P is the posterior height for the vertebral body at the current level, h_M is the mid height for the vertebral body at the current level, $h_{P_{i-1}}$ is the posterior height for the vertebral body at the level below, $h_{P_{i+1}}$ is the posterior height for the vertebral body at the level above, $h_{A_{i-1}}$ is the anterior height for the vertebral body at the level below, and $h_{A_{i+1}}$ is the anterior height for the vertebral body at the level above.

$$\text{wedge ratio}(r_W) = h_A/h_P$$

$$\text{biconcave ratio}(r_B) = h_M/h_P$$

$$\text{crush ratio}(r_C) = \min[\max(h_{P_i}/h_{P_{i-1}}, h_{A_i}/h_{A_{i-1}}), \max(h_{P_i}/h_{P_{i+1}}, h_{A_i}/h_{A_{i+1}})]$$

$$\text{wedge deformity}(d_W) = 100 \times [1 - r_W]$$

$$\text{biconcave deformity}(d_B) = 100 \times [1 - r_B]$$

$$\text{crush deformity}(d_C) = 100 \times [1 - r_C]$$

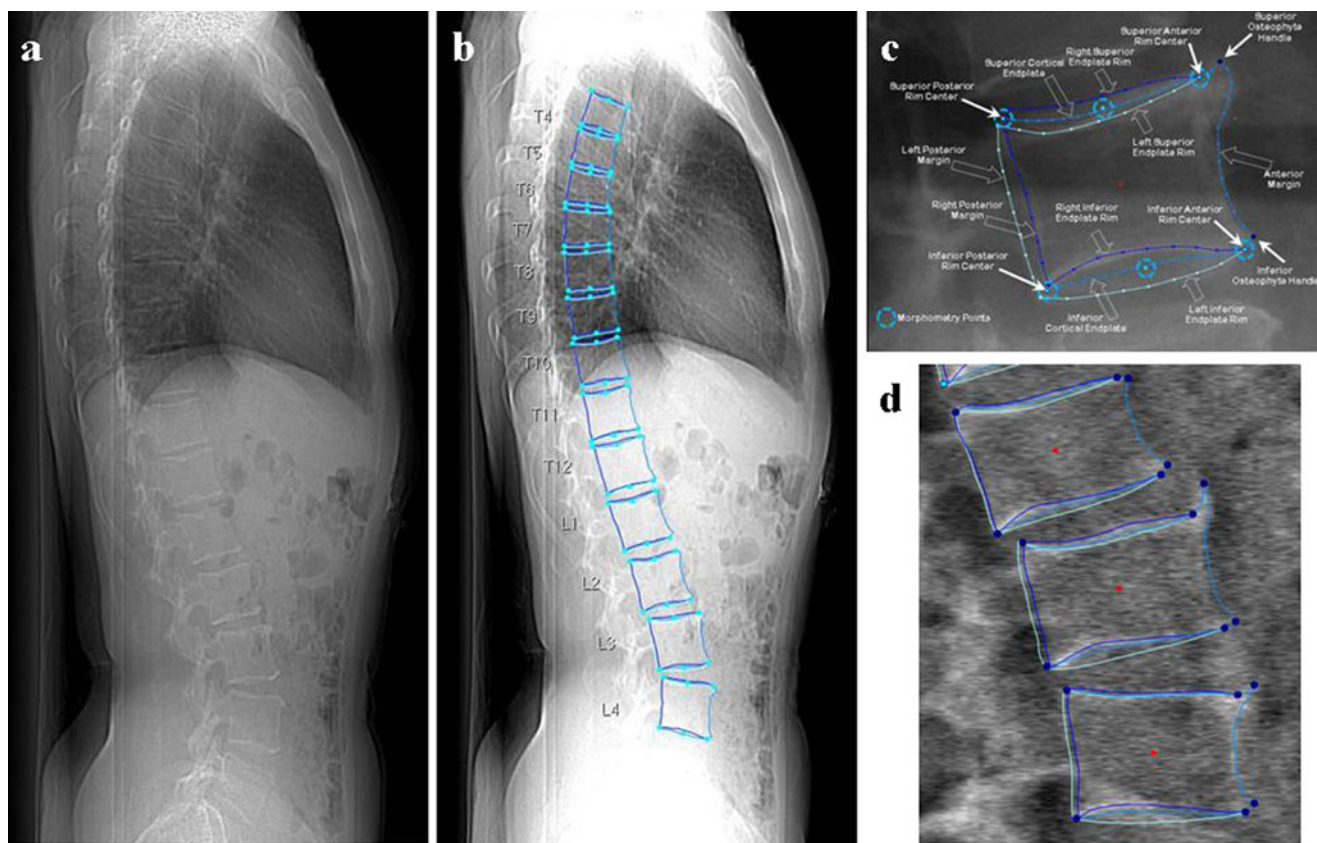


Fig. 1 Method used to conduct semi-automated quantitative morphometry measurements (SpineAnalyzer, Optasia Medical, Cheadle, UK). **a** CT lateral scout view; **b** operator identification of individual

vertebra from T4 to L4; **c** automated placement of standard quantitative morphometry points (*large circles*) and 95 points for shape definition (*smaller dots*); and **d** shape contours of lumbar area

Vertebral fractures were identified based by deformity percentages derived from morphometry alone, using Genant's semi-quantitative scale [14] as a guide; grade 0 (<20% deformity), grade 1 ($\geq 20\%$ deformity), grade 2 ($\geq 25\%$ deformity), and grade 3 ($\geq 40\%$ deformity).

The time to perform morphometry measurements was measured by one non-radiologist reader in 25 randomly selected subjects for both the unadjusted morphometry point placement and the semi-automated approach, in which the operator adjusts morphometry point placement as needed.

Statistical analysis

We computed intraclass correlations (ICCs), and the root mean squared coefficient of variation (RMS CV) for the inter- and intra-reader differences in the vertebral heights and height ratios (wedge, biconcave, and crush) for the semi-automated approach. We also computed intra-reader reliability for the two trials of unadjusted measurements, and for the comparison of unadjusted vs. semi-automated (i.e., manually adjusted) morphometry point placement. To compute RMS CV, we used the method recommended by Glüer et

al. [28]. To determine reliability of vertebral fracture assessment based on deformity percentages, we estimated agreement corrected for chance using a simple kappa (k) statistic and associated 95% confidence interval (CI) [29]. We computed k for intra-reader agreement between time 1 and time 2 separately for each of the two readers, and k for inter-reader agreement between reader A and reader B separately for each of the two time points. We computed the k values for two dichotomous definitions of prevalent vertebral fracture: (1) grades 1–3 (mild, moderate, and severe) versus grade 0 (normal), and (2) grades 2–3 (moderate and severe) versus grades 0–1 (normal and mild). Unreadable vertebrae ($n=35$) were classified as normal (grade 0).

Analyses were performed on a vertebra-specific level. We considered $k > 0.75$ as excellent agreement, 0.40–0.75 as fair to good agreement, and < 0.40 as poor agreement beyond that expected by chance as characterized by Fleiss [30]. We also conducted a stratified analysis to evaluate potential differences in agreement by spinal region, categorized as T4–6, T7–9, T10–12, and L1–4, as previously reported [22, 31]. We used Student's t test to compare the time to complete morphometric analysis in subjects with at

least one prevalent mild vertebral fracture vs. those with no fracture, and in subjects with at least one prevalent moderate or severe fracture vs. those with no fracture. We compared the prevalence of vertebral fracture identification by both readers and both time points by vertebral level. All statistical analyses were conducted using SPSS (SPSS Inc., Chicago, IL, USA), and SAS (SAS Institute Inc., Cary, NC, USA).

Results

Demographic data of participants

Among 100 participants, 4 women were excluded due to poor image quality. The average BMI of four women excluded was $42.0 \pm 7.4 \text{ kg/m}^2$ (34.5–52.1), indicating that the poor image quality was likely due to excessive weight or obesity. Two vertebrae were not able to be analyzed by both readers at both time points, and were excluded from the analysis. In addition, 35 vertebrae were unreadable by at least one reader at one time point. All together, 4,949 vertebrae (99.1%) out of a total of 4,992 were analyzable using the semi-automated algorithm. Demographic data and age distribution for the 50 men and 46 women included in the reliability study are shown in Table 1. The participants' mean age was 70.3 ± 8.9 years, ranging from 50 to 87 years.

Reproducibility of vertebral morphometry measurements using the algorithm without manual contour adjustment

To evaluate the reproducibility of the morphometry algorithm without manual contour adjustments, we computed the ICCs for the morphometry done at two time points, and also compared this to ICCs for unadjusted morphometry points vs. semi-automated (i.e., manual intervention by the operator's discretion) (reader A, time 1) in the same 96 participants. The ICCs for vertebral height measurements for unadjusted time 1 vs. unadjusted time 2 ranged from 0.97 to 0.98 (Table 2) and were similar to ICCs for vertebral heights comparing the unadjusted to the semi-automated approach (ICC=0.96 to 0.97) (Table 2). In both

Table 1 Demographic data of study subjects (mean±SD)

	Men (N=50)	Women (N=46)	Total (N=96)
Age (years)	70±9	71±9	70±9
50–59 (n)	6 (12%)	7 (15%)	13 (14%)
60–69 (n)	16 (32%)	13 (28%)	29 (30%)
≥70 (n)	28 (56%)	26 (57%)	54 (56%)
Height (cm)	174±7	159±7	167±10
Weight (kg)	85±14	68±12	77±16
BMI (kg/m ²)	28±4	27±4	27±4

Table 2 ICCs of vertebral heights and height ratios, unadjusted-time 1 vs. unadjusted-time 2 and unadjusted vs. semi-automated approach

T4-L4	ICC (95% CI)	Unadjusted-time 1 vs. unadjusted-time 2			Unadjusted vs. semi-automated		
		No fracture ^a (n=1,190)		Fracture ^a (n=56)	All (n=1,246)		Fracture ^a (n=56)
<i>h_P</i>		0.974 (0.971, 0.977)	0.975 (0.972, 0.978)	0.914 (0.854, 0.950)	0.972 (0.969, 0.975)	0.973 (0.969, 0.976)	0.947 (0.909, 0.969)
<i>h_A</i>		0.978 (0.976, 0.981)	0.979 (0.976, 0.981)	0.919 (0.861, 0.952)	0.961 (0.956, 0.965)	0.965 (0.960, 0.968)	0.738 (0.548, 0.848)
<i>h_M</i>		0.978 (0.976, 0.980)	0.979 (0.976, 0.981)	0.920 (0.864, 0.953)	0.965 (0.961, 0.969)	0.969 (0.965, 0.972)	0.826 (0.700, 0.899)
<i>r_W</i>		0.792 (0.768, 0.814)	0.782 (0.756, 0.806)	0.641 (0.388, 0.790)	0.644 (0.602, 0.681)	0.651 (0.609, 0.689)	-0.192 (-1.055, 0.308)
<i>r_B</i>		0.718 (0.684, 0.747)	0.712 (0.677, 0.743)	0.335 (-0.134, 0.610)	0.408 (0.338, 0.471)	0.415 (0.345, 0.478)	-0.655 (-1.853, 0.040)
<i>r_C</i>		0.577 (0.527, 0.622)	0.571 (0.519, 0.617)	0.616 (0.346, 0.775)	0.565 (0.514, 0.611)	0.541 (0.486, 0.591)	0.764 (0.594, 0.863)

^a Fracture defined by vertebral deformity ≥20% using analyses from auto-time 1; *h_A*, height anterior; *h_M*, height mid, *r_W*, wedge ratio; *r_B*, biconcave ratio; *r_C*, crush ratio

cases, ICCs showed that the vertebral height measurements were highly reliable, but the vertebral height ratios were slightly less reliable than the vertebral heights. Moreover, the unadjusted algorithm for morphometry point placement is less reliable in subjects with fractures, providing rationale for the approach in which the operator adjusts morphometry points after visual inspection of the images (Figs. 2 and 3).

Time to conduct semi-automated vertebral morphometry measurements

The time needed to conduct semi-automated vertebral morphometry measurements averaged 5.4 min \pm 1.7 min (range, 3.2–9.1 min) per subject. In comparison, the average time needed to place morphometry points using the algorithm with no operator adjustments of vertebral contours was 45.6 \pm 4.5 s (38.7–56.8 s). There was no difference in analysis time for the semi-automated approach between subjects with no fracture and those with only mild vertebral fracture. However, the time needed to perform semi-automated morphometry measurements was approximately 2 min longer in subjects with moderate or severe fracture compared to those with no fracture (6.7 \pm 1.6 min vs. 4.6 \pm 1.3 min, $p=0.002$). The time required to complete the semi-automated morphometry analysis was independent of sex, age, height, weight, and BMI (data not shown).

Reproducibility of vertebral morphometry by the semi-automated algorithm (i.e., manual adjustment of point placement by operator)

Intra- and inter-reader ICCs for anterior, mid, and posterior vertebral heights for all vertebral levels combined were excellent, ranging from 0.96 to 0.98 (Table 3). ICCs were also excellent at distinct spinal regions (T4-9, T10-12, and

L1-4), ranging from 0.87 to 0.96. Intra- and inter-reader RMS CV values ranged from 2.5% to 3.9% and 3.3% to 4.4%, respectively.

For vertebral height ratios, *intra-reader* ICCs for wedge ratio were good to excellent for all vertebrae together and for distinct spinal regions (T4-9, T10-12, and L1-4), with ICCs ranging from 0.62 to 0.83 (Table 4). However, biconcave and crush ratio measurements were less reliable with ICCs ranging from 0.36 to 0.73 for biconcave ratio and 0.48 to 0.63 for crush ratio. Likewise, *inter-reader* ICCs for wedge ratio were also good to excellent for all vertebrae together and for distinct spinal regions, ranging from 0.59 to 0.80, but the inter-reader reliability of biconcave and crush ratio measurements were only fair to good, with ICCs ranging from 0.38 to 0.72 for biconcave ratio and 0.42 to 0.62 for crush ratio. Intra- and inter-reader RMS CV for the various height ratios ranged from 3.6% to 5.8% and 4.1% to 5.4%, respectively (Table 4). However, unlike the comparable reliability of vertebral heights in the T4-6 region with other regions (Table 3), the intra- and inter-reader reliability was worse for vertebral height ratios at T4-6 compared to other spine regions (Table 4).

Reliability of morphometric vertebral deformities by the semi-automated algorithm

Based on morphometry measurements alone, readers A and B identified 52 and 59 subjects at time 1, and 51 and 46 subjects at time 2 with at least one prevalent vertebral fracture, respectively (Table 5). The total number of vertebrae classified as fractured ranged from 6.2% to 8.7%, including 91 and 108 by readers A and B at time 1, respectively, and 94 and 77 at time 2 by readers A and B, respectively, with the majority of fractures (53–63%) being mild (SQ 1, i.e., morphometric deformity $\geq 20\%$, but $< 25\%$)

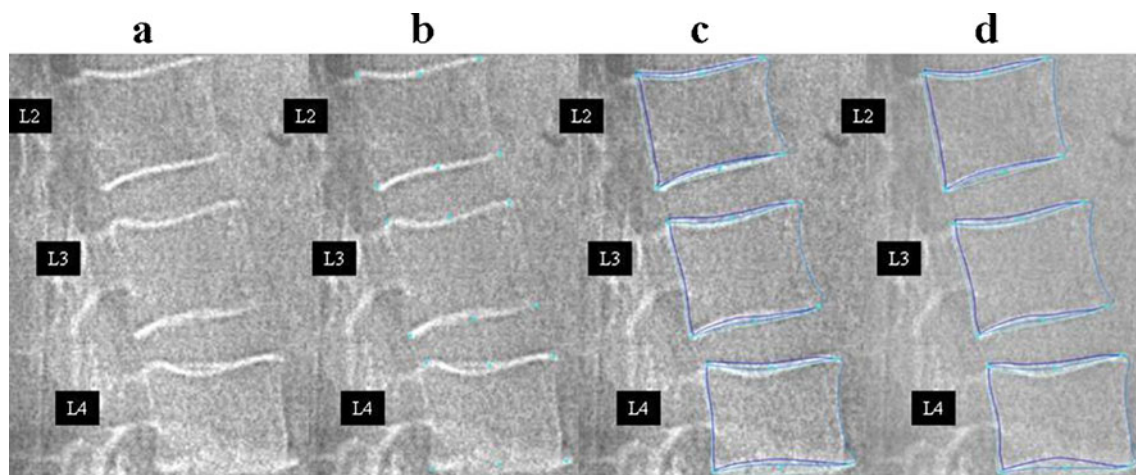


Fig. 2 Unadjusted algorithm which needed no manual adjustment. **a** Original lateral scout view from CT scan, **b** automated placement of morphometry points, and **c** automated placement of contour lines

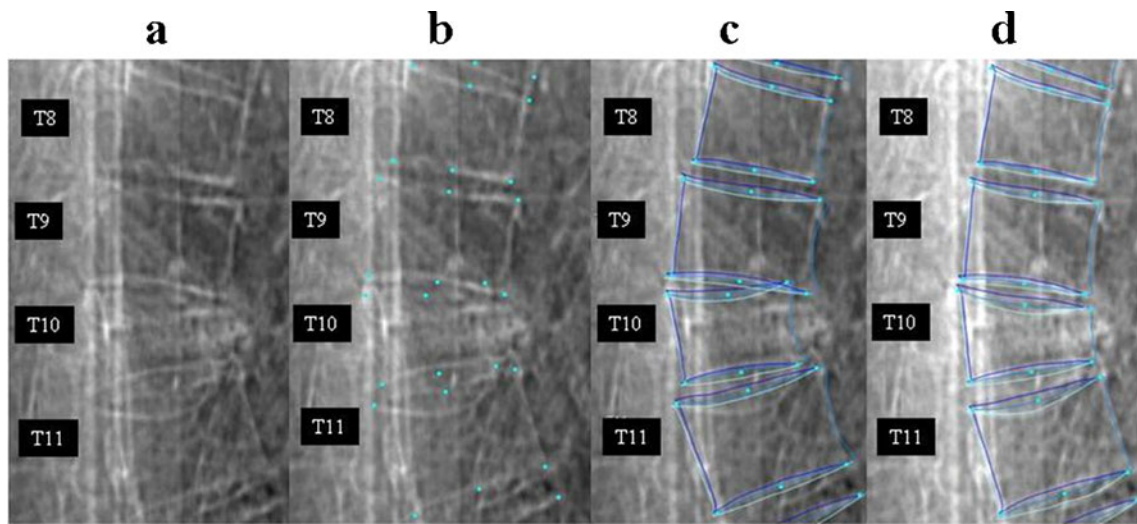


Fig. 3 Unadjusted algorithm which needed manual adjustment. **a** Original lateral scout view from CT scan, **b** automated placement of morphometry points, and **c** automated placement of contour lines,

showing poor placement for fractured vertebra at T10; and **d** contour lines after manual adjustment of points on T10

(Table 5). The proportion of wedge fractures (73.2%) was much higher than that of biconcave (15.7%) or crush (11.1%) fractures. The distribution of prevalent fractures by spinal location was bimodal, with peak frequencies occurring at T7-8 and T11-12 (Fig. 4). Whereas the spatial

distribution of wedge fractures mimicked the total bimodal distribution, biconcave fractures were congregated within lower thoracic and lumbar vertebrae, whereas crush fractures were distributed throughout the vertebra levels. In terms of fracture distribution by deformity types at

Table 3 Intra- and inter-reader reliability of vertebral height measurements by semi-automated morphometry

		Intra-reader				Inter-reader			
		A		B		Time 1		Time 2	
		ICCs	RMS CV (%)	ICCs	RMS CV (%)	ICCs	RMS CV (%)	ICCs	RMS CV (%)
All (T4-L4)	h_P	0.97	2.8	0.96	3.4	0.96	3.6	0.96	3.6
	h_A	0.98	2.7	0.98	3.2	0.97	3.8	0.97	3.8
	h_M	0.98	2.7	0.97	3.4	0.96	4.1	0.97	4.0
T4-9	h_P	0.93	2.8	0.89	3.9	0.89	3.8	0.89	3.8
	h_A	0.96	2.5	0.91	3.8	0.91	3.8	0.91	3.7
	h_M	0.95	2.5	0.91	3.5	0.91	3.9	0.92	4.0
T4-6	h_P	0.92	2.8	0.88	3.8	0.88	3.9	0.86	4.1
	h_A	0.94	2.7	0.89	3.8	0.88	3.9	0.87	4.0
	h_M	0.93	2.7	0.89	3.8	0.88	4.3	0.88	4.6
T7-9	h_P	0.91	2.8	0.86	3.8	0.86	3.8	0.90	3.4
	h_A	0.96	2.2	0.92	3.5	0.91	3.7	0.92	3.4
	h_M	0.95	2.3	0.91	3.2	0.91	3.5	0.93	3.2
T10-12	h_P	0.95	2.6	0.92	3.2	0.92	3.3	0.93	3.4
	h_A	0.95	2.7	0.95	2.9	0.94	3.4	0.94	3.5
	h_M	0.95	2.7	0.93	3.5	0.91	4.3	0.92	4.2
L1-4	h_P	0.91	2.9	0.91	3.1	0.88	3.6	0.89	3.4
	h_A	0.92	2.9	0.94	2.8	0.89	4.0	0.90	4.0
	h_M	0.92	2.9	0.91	3.2	0.87	4.4	0.91	3.9

h_P , height posterior; h_A , height anterior; h_M , height mid

Table 4 Intra- and inter-reader reliability of vertebral height ratios by semi-automated morphometry

		Intra-reader				Inter-reader				
		A		B		Time 1		Time 2		
		ICCs	RMS CV (%)	ICCs	RMS CV (%)	ICCs	RMS CV (%)	ICCs	RMS CV (%)	
All (T4-L4)	r_W	0.81	4.5	0.75	5.1	0.75	5.3	0.77	5.0	
	r_B	0.68	3.9	0.60	4.4	0.59	4.6	0.63	4.2	
	r_C	0.63	4.2	0.56	4.7	0.53	4.8	0.55	4.6	
	T4-9	r_W	0.72	4.4	0.62	5.3	0.59	5.4	0.63	5.1
		r_B	0.49	3.6	0.36	4.7	0.38	4.6	0.41	4.0
		r_C	0.57	4.0	0.48	4.9	0.51	4.6	0.44	4.7
	T4-6	r_W	0.64	4.5	0.47	5.5	0.45	5.6	0.48	5.2
		r_B	0.43	3.7	0.24	5.0	0.30	4.9	0.20	4.4
		r_C	0.54	4.1	0.46	5.0	0.45	4.8	0.38	5.1
T7-9	r_W	0.78	4.3	0.71	5.1	0.68	5.2	0.71	5.0	
	r_B	0.53	3.5	0.45	4.3	0.46	4.2	0.56	3.6	
	r_C	0.59	3.9	0.49	4.7	0.55	4.3	0.48	4.4	
T10-12	r_W	0.83	3.9	0.74	4.8	0.76	4.7	0.80	4.2	
	r_B	0.67	3.8	0.64	4.0	0.63	4.1	0.59	4.2	
	r_C	0.62	4.6	0.58	5.1	0.62	4.9	0.62	4.6	
L1-4	r_W	0.78	5.1	0.77	5.0	0.77	5.4	0.76	5.4	
	r_B	0.73	4.2	0.71	4.3	0.66	4.9	0.72	4.4	
	r_C	0.63	4.1	0.59	4.2	0.42	5.0	0.55	4.3	

r_W , wedge ratio; r_B , biconcave ratio; r_C , crush ratio

specific vertebral levels, wedge fractures comprised most of the prevalent fractures in thoracic and thoracolumbar area; whereas biconcave and crush fractures comprised most of the prevalent fractures in the L2-4 region (data not shown).

Examining all vertebral levels together, we found good intra- ($k=0.59$ to 0.69) and inter-reader agreement ($k=0.67$) for vertebral fracture defined by a deformity of $\geq 20\%$ (\geq SQ 1) (Table 6). When we compared the agreement for vertebral fracture defined by a deformity of $\geq 25\%$ (i.e., \geq SQ 2), intra-reader agreements was higher with k of 0.58 to 0.77 , but

slightly lower for inter-reader agreement with k of 0.61 to 0.64 (Table 6). Inter- and intra-reader agreement was fair at T4-6, but was good to excellent at other spinal regions (T7-9, T10-12, and L1-4) (Table 6).

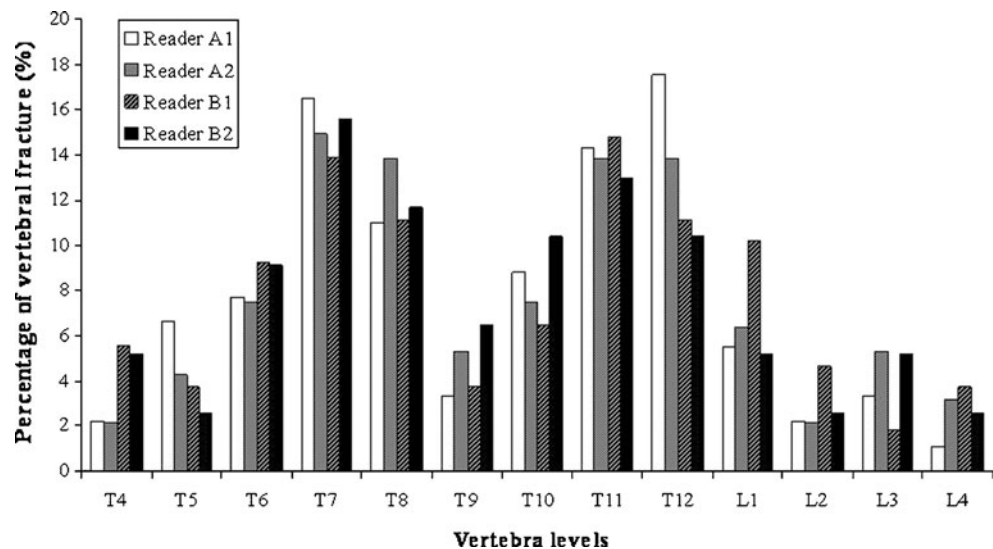
Discussion

In this study, we determined intra- and inter-reader reliability of semi-automated vertebral morphometry meas-

Table 5 Number of subjects and vertebrae according to semi-quantitative grade, determined solely from quantitative morphometry measurements

	Subjects ($n=96$)				Vertebrae ($n=1246$)			
	Reader A		Reader B		Reader A		Reader B	
	Time 1 No. (%)	Time 2 No. (%)	Time 1 No. (%)	Time 2 No. (%)	Time 1 No. (%)	Time 2 No. (%)	Time 1 No. (%)	Time 2 No. (%)
SQ 0	44 (45.8)	45 (46.9)	37 (38.5)	50 (52.1)	1155 (92.7)	1152 (92.5)	1138 (91.3)	1169 (93.8)
SQ 1	26 (27.1)	25 (26.0)	29 (30.2)	21 (21.9)	56 (4.5)	59 (4.7)	63 (5.1)	41 (3.3)
SQ 2	18 (18.8)	19 (19.8)	23 (24.0)	21 (21.9)	26 (2.1)	27 (2.2)	37 (3.0)	32 (2.6)
SQ 3	8 (8.3)	7 (7.3)	7 (7.3)	4 (4.1)	9 (0.7)	8 (0.6)	8 (0.6)	4 (0.3)
Total SQ \geq 1	52 (54.2)	51 (53.1)	59 (61.5)	46 (47.9)	91 (7.3)	94 (7.5)	108 (8.7)	77 (6.2)
Total SQ \geq 2	26 (27.1)	26 (27.1)	30 (31.3)	25 (26.0)	35 (2.8)	35 (2.8)	45 (3.6)	36 (2.9)

Fig. 4 Prevalence of vertebral fracture (deformity $\geq 20\%$) by vertebral level. The highest prevalence was seen at T7-8 and T11-12. Reader A1, reader A-time 1; reader A2, reader A-time 2; reader B1, reader B-time 1; reader B2, reader B-time 2



urements and morphometry-based fractures using lateral CT scout views. The semi-automated algorithm provided excellent intra- and inter-reader reliability for vertebral height measurements, along with good to fair reliability for vertebral height ratios. Reliability for vertebral fracture assessment based solely on quantitative morphometry was also good and was comparable to previous reports for SQ vertebral fracture grading by radiologists. Furthermore, the average time to complete the semi-automated morphometry analysis was approximately 9 min 40 s less than previously reported for manual morphometry analysis [32].

Our evaluation of the reproducibility of measurements derived from the unadjusted morphometry points showed that the ICCs for the two unadjusted analyses were not equal to 1, demonstrating some variation in the algorithm's point placement due to variation wherein the operator places the initial seed point in the middle of the vertebral body. The ICC for unadjusted analyses would only be equal to 1 if the initialization points were placed at identical locations every time. The ICCs from the unadjusted morphometry points were worse in subjects with fractures than those without fractures, providing strong rationale for operator review of point placement, and adjustment as needed.

Although we found excellent intra- and inter-reader reliability for semi-automated vertebral height measurements, inter- and intra-reader reliability of vertebral height ratios and associated agreement for vertebral fracture classification was good to fair. Poorer reliability of height ratios is anticipated, since the ratio reflects the error in each of the individual height measurements used to compute the ratio. Furthermore, our classification scheme for vertebral fractures relied on thresholds for height ratios, meaning that a very small (and clinically insignificant) variation in a height ratio from 24.8% to 25.1% would lead to two different fracture classifications. Such a phenomenon

explains in part why *k* statistics for vertebral fracture were worse than agreement for height measurements, and also why the *k* scores did not improve even when the fracture definition was changed to include only moderate and severe fractures (i.e., deformity $\geq 25\%$), as is customarily seen in studies that utilize semi-quantitative visual assessment of fractures [22, 31].

In general, we observed the better reliability for wedge and biconcave fractures than for crush fractures. This may be because intra-vertebral height ratios are used to define wedge and biconcave fractures. In contrast, inter-vertebral height ratios are used to determine crush fractures, and this may have contributed to the inferior reliability for these fracture types. In addition, the relatively low prevalence of crush fractures may have contributed to their lower intra- and inter-reader reliability.

In comparing our results to previous reports, it is important to note that the reliability of vertebral heights and height ratios varies according to the type of imaging modality, the spinal location, and subject characteristics (i.e., osteoporotic vs. normal) [33, 34]. Specifically, the reliability of morphometric measurements from radiographs is generally better than morphometric measurements from lateral dual-energy X-ray absorptiometry scans due to superior spatial resolution provided by radiographs [33, 34]. The reliability of quantitative morphometry measurements is better in persons who do not have osteoporosis than in those with low bone density or vertebral fractures [33, 34]. In the current study, we used QCT lateral scout views from a mixed sample of normal and osteoporotic subjects and found reliability for vertebral height measurements to be comparable to prior reports [34, 35]. Intra- and inter-reader reliability of vertebral height ratios from the current study were slightly worse than those of reported for lateral radiographs but comparable to those reported for lateral dual-energy X-ray absorptiometry scans [34]. Prior reports did not systemati-

Table 6 Intra- and inter-reader agreement for vertebral fracture classification, kappa (k) statistics (95% CI)

	SQ \geq 1 vs. SQ=0						SQ \geq 2 vs. SQ \leq 1					
	Intra-reader			Inter-reader			Intra-reader			Inter-reader		
	A	B	Time 2	Time 1	Time 2	Time 1	A	B	Time 2	Time 1	Time 2	
All (T4-L4)	0.69 (0.61, 0.77)	0.59 (0.50, 0.67)	0.67 (0.58, 0.75)	0.67 (0.59, 0.75)	0.67 (0.58, 0.75)	0.61 (0.49, 0.74)	0.77 (0.65, 0.88)	0.58 (0.45, 0.71)	0.64 (0.51, 0.77)	0.61 (0.49, 0.74)	0.64 (0.51, 0.77)	
T4-9	0.71 (0.59, 0.82)	0.54 (0.41, 0.67)	0.67 (0.55, 0.79)	0.63 (0.51, 0.75)	0.67 (0.55, 0.79)	0.50 (0.30, 0.69)	0.72 (0.54, 0.90)	0.43 (0.24, 0.63)	0.64 (0.45, 0.83)	0.50 (0.30, 0.69)	0.64 (0.45, 0.83)	
T4-6	0.47 (0.24, 0.71)	0.39 (0.17, 0.61)	0.52 (0.28, 0.76)	0.54 (0.34, 0.75)	0.52 (0.28, 0.76)	0.42 (0.09, 0.74)	0.44 (0.03, 0.85)	0.15 (-0.13, 0.44)	0.28 (-0.16, 0.72)	0.42 (0.09, 0.74)	0.28 (-0.16, 0.72)	
T7-9	0.81 (0.70, 0.93)	0.63 (0.48, 0.78)	0.73 (0.60, 0.87)	0.68 (0.54, 0.82)	0.73 (0.60, 0.87)	0.54 (0.30, 0.78)	0.83 (0.66, 0.99)	0.55 (0.32, 0.78)	0.73 (0.54, 0.92)	0.54 (0.30, 0.78)	0.73 (0.54, 0.92)	
T10-12	0.68 (0.54, 0.81)	0.73 (0.60, 0.86)	0.64 (0.49, 0.79)	0.78 (0.67, 0.89)	0.64 (0.49, 0.79)	0.70 (0.50, 0.90)	0.72 (0.50, 0.93)	0.72 (0.53, 0.91)	0.65 (0.43, 0.88)	0.70 (0.50, 0.90)	0.65 (0.43, 0.88)	
L1-4	0.66 (0.45, 0.87)	0.45 (0.24, 0.66)	0.70 (0.51, 0.90)	0.53 (0.32, 0.74)	0.70 (0.51, 0.90)	0.77 (0.51, 1.00)	0.93 (0.80, 1.00)	0.72 (0.42, 1.00)	0.61 (0.30, 0.92)	0.77 (0.51, 1.00)	0.61 (0.30, 0.92)	

cally evaluate whether reproducibility of vertebral height measurements depends on spinal region; however, at least one study [33] reported similar results as ours, with slightly worse reproducibility of height measurements in the upper thoracic regions than elsewhere. The poorer reproducibility is likely due to the difficulty of visualizing vertebrae in the upper thoracic region.

In comparing our results for reliability for vertebral fracture identification to those previously reported by Samelson et al. [22] for vertebral fracture assessment by standard SQ readings, we found that despite the use of different techniques for fracture identification (i.e., quantitative morphometry by non-radiologists vs. visual semi-quantitative scoring by trained radiologists), the two studies had comparable inter- and intra-reader reliability for vertebral fracture assessment. The distribution of vertebral fractures was also similar for the two approaches, with bimodal peaks at the mid-thoracic and thoracolumbar levels. However, vertebral fracture assessment solely by quantitative morphometry identified nearly twice as many fractures (SQ \geq 1) as the SQ method. The quantitative morphometry measurements may have been more sensitive than SQ reading for mild fractures. However, it is more likely that the increased number of fractures identified by purely quantitative morphometry is because the deformity-based classification scheme does not exclude non-fracture deformities of congenital, developmental, and degenerative origins. Ultimately, it may be best to combine QM methods with visual SQ grading to optimize sensitivity and accuracy of fracture determination. Alternatively, it may be possible to combine the morphometry data with ‘machine learning’ algorithms to develop computer-aided diagnosis for vertebral fractures, as proposed by Roberts et al. [36]. Altogether, the availability of semi-automated, reproducible, accurate vertebral morphometry data provides a great opportunity for future development of novel algorithms to improve detection of vertebral fractures.

Our study has a few limitations that must be considered when interpreting the results. Our reproducibility data are based on repeat analyses of the same scans, rather than analysis of duplicate scan acquisitions. Our vertebral morphometry measures and fracture assessments were based on lateral CT scout views, which are currently not used in routine practice for assessment of vertebral fractures. However, QCT is being used more commonly in general and in osteoporosis research studies. CT scans of the trunk that are acquired for other clinical reasons may be useful for assessment of vertebral fractures. In particular, midline sagittal reformations of CT scan data are highly useful for identifying vertebral fractures [37, 38], and indeed, our own fracture detection would likely have been enhanced if we had used this approach. In addition, vertebral fracture identification by semi-automated mor-

phometry has not yet been directly compared to traditional visual semi-quantitative methods. These data are needed to define the optimal clinical use for the automated morphometry measurements.

The current study also had a number of strengths. In particular, non-radiologist readers performed the semi-automated vertebral morphometry measurements, thereby establishing the practical utility of semi-automated vertebral morphometry by technical staff and/or physicians who may not have advanced training in radiology. Also, we conducted both intra- and inter-reader reliability of semi-automated vertebral morphometry using a sample enriched with subjects that had fractures, thereby providing reliability estimates in the expected population that the technique would be used clinically (i.e., those suspected of vertebral fracture). The subjects selected for this study may not represent the general population, as there were more subjects with vertebral fracture than would be expected for an age- and sex-matched sample of the general population.

In summary, underreporting of vertebral fractures is common worldwide [12, 13]. This is problematic, since both clinical and morphometric fractures are associated with significant morbidity and are strong risk factors for future fracture. Moreover, numerous treatments that reduce future fracture risk, even in subjects with prevalent fractures, are available. Thus, new methods that facilitate and enhance the detection of vertebral fractures may improve clinical management of patients with osteoporosis. The results from this study suggest that semi-automated vertebral morphometry is a feasible, reliable, and quick method that may complement current methods for identifying vertebral fractures and also promote development of novel automated algorithms to enhance vertebral fracture identification.

Acknowledgments This work was supported by grants from the National Institute for Arthritis, Musculoskeletal, and Skin Diseases (NIAMS) R01AR053986, NIAMS and the National Institute on Aging (NIA) R01AR/AG041398, and by the National Heart, Lung, and Blood Institute (NHLBI) Framingham Heart Study (NIH/NHLBI Contract N01-HC-25195). We acknowledge the assistance of Peter Steiger, PhD of Optasia Medical.

Conflicts of interest None.

References

- van Helden S, Cals J, Kessels F, Brink P, Dinant GJ, Geusens P (2006) Risk of new clinical fractures within 2 years following a fracture. *Osteoporos Int* 17:348–354
- Klotzbuecher CM, Ross PD, Landsman PB, Abbott TA 3rd, Berger M (2000) Patients with prior fractures have an increased risk of future fractures: a summary of the literature and statistical synthesis. *J Bone Miner Res* 15:721–739
- Kanis JA, Johnell O, De Laet C, Johansson H, Oden A, Delmas P, Eisman J, Fujiwara S, Garnero P, Kroger H, McCloskey EV, Mellstrom D, Melton LJ, Pols H, Reeve J, Silman A, Tenenhouse A (2004) A meta-analysis of previous fracture and subsequent fracture risk. *Bone* 35:375–382
- Edmond SL, Kiel DP, Samelson EJ, Kelly-Hayes M, Felson DT (2005) Vertebral deformity, back symptoms, and functional limitations among older women: the Framingham Study. *Osteoporos Int* 16:1086–1095
- Melton LJ 3rd (2003) Adverse outcomes of osteoporotic fractures in the general population. *J Bone Miner Res* 18:1139–1141
- Lau E, Ong K, Kurtz S, Schmier J, Edidin A (2008) Mortality following the diagnosis of a vertebral compression fracture in the Medicare population. *J Bone Joint Surg Am* 90:1479–1486
- Delmas PD, van de Langerijt L, Watts NB, Eastell R, Genant H, Grauer A, Cahall DL, IMPACT Study Group (2005) Underdiagnosis of vertebral fractures is a worldwide problem: the IMPACT study. *J Bone Miner Res* 20(4):557–563
- Freedman BA, Potter BK, Nesti LJ, Giuliani JR, Hampton C, Kuklo TR (2008) Osteoporosis and vertebral compression fractures-continued missed opportunities. *Spine J* 8:756–762
- Siris ES, Bilezikian JP, Rubin MR, Black DM, Bockman RS, Bone HG, Hochberg MC, McClung MR, Schnitzer TJ (2003) Pins and plaster aren't enough: a call for the evaluation and treatment of patients with osteoporotic fractures. *J Clin Endocrinol Metab* 88:3482–3486
- Lems WF (2007) Clinical relevance of vertebral fractures. *Ann Rheum Dis* 66:2–4
- Lentle BC, Brown JP, Khan A, Leslie WD, Levesque J, Lyons DJ, Siminoski K, Tarulli G, Josse RG, Hodsman A (2007) Recognizing and reporting vertebral fractures: reducing the risk of future osteoporotic fractures. *Can Assoc Radiol J* 58:27–36
- Majumdar SR, Kim N, Colman I, Chahal AM, Raymond G, Jen H, Siminoski KG, Hanley DA, Rowe BH (2005) Incidental vertebral fractures discovered with chest radiography in the emergency department: prevalence, recognition, and osteoporosis management in a cohort of elderly patients. *Arch Intern Med* 165:905–909
- Gehlbach SH, Bigelow C, Heimisdottir M, May S, Walker M, Kirkwood JR (2000) Recognition of vertebral fracture in a clinical setting. *Osteoporos Int* 11:577–582
- Genant HK, Wu CY, van Kuijk C, Nevitt MC (1993) Vertebral fracture assessment using a semiquantitative technique. *J Bone Miner Res* 8(9):1137–1148
- Jiang G, Eastell R, Barrington NA, Ferrar L (2004) Comparison of methods for the visual identification of prevalent vertebral fracture in osteoporosis. *Osteoporos Int* 15:887–896
- Herxthal LM (1968) Measurement of anterior vertebral compressions and biconcave vertebrae. *Am J Roentgenol Radium Ther Nucl Med* 103(3):635–644
- Prevrhal S, Krege JH, Chen P, Genant H, Black DM (2009) Teriparatide vertebral fracture risk reduction determined by quantitative and qualitative radiographic assessment. *Curr Med Res Opin* 25(4):921–928
- Ferrar L, Jiang G, Schousboe JT, DeBold CR, Eastell R (2008) Algorithm-based qualitative and semiquantitative identification of prevalent vertebral fracture: agreement between different readers, imaging modalities, and diagnostic approaches. *J Bone Miner Res* 23(3):417–424
- Genant HK, Jergas M, Palermo L, Nevitt M, Valentin RS, Black D, Cummings SR (1996) Comparison of semiquantitative visual and quantitative morphometric assessment of prevalent and incident vertebral fractures in osteoporosis The Study of Osteoporotic Fractures Research Group. *J Bone Miner Res* 11(7):984–996
- Grados F, Fechtenbaum J, Flipo E, Kolta S, Roux C, Fardellone P (2009) Radiographic methods for evaluating osteoporotic vertebral fractures. *Joint Bone Spine* 76(3):241–247

21. Brett A, Miller CG, Hayes CW, Krasnow J, Ozanian T, Caminis J, Block JE, van Kuijk C (2009) Development of a clinical workflow tool to enhance the detection of vertebral fractures: accuracy and precision evaluation. *Spine* 34(22):2437–2443
22. Samelson EJ, Christiansen BA, Demissie S, Broe KE, Zhou Y, Meng CA, Yu W, Cheng X, O'Donnell CJ, Hoffmann U, Genant HK, Kiel DP (2010) Reliability of vertebral fracture assessment using multidetector CT lateral scout views: the Framingham Osteoporosis Study. *Osteoporos Int* May 21 (in press)
23. Moselewski F, O'Donnell CJ, Achenbach S, Ferencik M, Massaro J, Nguyen A, Cury RC, Abbara S, Jang IK, Brady TJ, Hoffmann U (2005) Calcium concentration of individual coronary calcified plaques as measured by multidetector row computed tomography. *Circulation* 111:3236–3241
24. Pou KM, Massaro JM, Hoffmann U, Lieb K, Vasan RS, O'Donnell CJ, Fox CS (2009) Patterns of abdominal fat distribution: the Framingham Heart Study. *Diabetes Care* 32:481–485
25. Cootes TF, Edwards GJ, Taylor CJ (2001) Active appearance models. *IEEE Trans Pattern Anal Mach Intell* 23(6):681–885
26. Roberts M, Cootes T, Pacheco E, Adams J (2007) Quantitative vertebral fracture detection on DXA images using shape and appearance models. *Acad Radiol* 14:1166–1178
27. Black DM, Palermo L, Nevitt MC, Genant HK, Epstein R, San Valentin R, Cummings SR (1995) Comparison of methods for defining prevalent vertebral deformities: the study of osteoporotic fractures. *J Bone Miner Res* 10(6):890–902
28. Gluer CC, Blake G, Lu Y, Blunt BA, Jergas M, Genant HK (1995) Accurate assessment of precision errors: how to measure the reproducibility of bone densitometry techniques. *Osteoporos Int* 4(4):262–270
29. Cohen J (1960) A coefficient of agreement for nominal scales. *Educ Psychol Meas* 20:37–46
30. Fleiss J (1981) The measurement of interrater agreement. *Statistical method for rates and proportions*. John Wiley and Sons, Inc., New York, pp 211–236
31. Takeda M, Wu CY, Lang TF, Genant HK (1998) Vertebral fracture assessment using the lateral scoutview of computed tomography in comparison with radiographs. *Osteoporos Int* 8:197–203
32. Guglielmi G, Palmieri F, Placentino MG, D'Errico F, Stoppino LP (2009) Assessment of osteoporotic vertebral fractures using specialized workflow software for 6-point morphometry. *Eur J Radiol* 70:142–148
33. Ferrar L, Jiang G, Eastell R (2001) Short-term precision for morphometric X-ray absorptiometry. *Osteoporos Int* 12:710–715
34. Rea JA, Chen MB, Li J, Potts E, Fan B, Blake GM, Steiger P, Smith IG, Genet HK, Fogelman I (1999) Morphometric X-ray absorptiometry and morphometric radiography of the spine: a comparison of analysis precision in normal and osteoporotic subjects. *Osteoporos Int* 9:536–544
35. Wu CY, Li J, Jergas M, Genant HK (1995) Comparison of semiquantitative and quantitative techniques for the assessment of prevalent and incident vertebral fractures. *Osteoporos Int* 5:354–370
36. Roberts MG, Pacheco EMB, Mohankumar R, Cootes TF, Adams JE (2010) Detection of vertebral fractures in DXA VFA images using statistical models of appearance and a semi-automatic segmentation. *Osteoporos Int* 21:2037–2046
37. Woo EK, Mansoubi H, Alyas F (2008) Incidental vertebral fractures on multidetector CT images of the chest: prevalence and recognition. *Clin Radiol* 63:160–164
38. Williams AL, Al-Busaidi A, Sparrow PJ, Adams JE, Whitehouse RW (2009) Under-reporting of osteoporotic vertebral fractures on computed tomography. *Eur J Radiol* 69:179–183