Published in final edited form as:

Osteoporos Int. 2012 March ; 23(3): 1007–1016. doi:10.1007/s00198-011-1774-z.

Identification of prevalent vertebral fractures using CT lateral scout views: a comparison of semi-automated quantitative vertebral morphometry and radiologist semi-quantitative grading

Y. M. Kim,

Center for Advanced Orthopaedic Studies, Beth Israel Deaconess Medical Center, Boston, MA, USA. Department of Orthopedic Surgery, Harvard Medical School, Boston, MA, USA. Department of Internal Medicine, Mizmedi Hospital, Seoul, South Korea

S. Demissie,

Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA

H. K. Genant,

Synarc and University of California San Francisco, San Francisco, CA, USA

X. Cheng,

Department of Radiology, Peking University, Jishuitan Hospital, Beijing, China

W. Yu,

Department of Radiology, Peking Union Medical College Hospital, Beijing, China

E. J. Samelson,

Institute for Aging Research, Hebrew SeniorLife, Boston, MA, USA. Department of Medicine, Harvard Medical School, Boston, MA, USA

D. P. Kiel, and

Institute for Aging Research, Hebrew SeniorLife, Boston, MA, USA. Department of Medicine, Harvard Medical School, Boston, MA, USA

M. L. Bouxsein

Center for Advanced Orthopaedic Studies, Beth Israel Deaconess Medical Center, Boston, MA, USA. Department of Orthopedic Surgery, Harvard Medical School, Boston, MA, USA

M. L. Bouxsein: mbouxsei@bidmc.harvard.edu

Abstract

Summary—We compared vertebral fracture assessment by semi-automated quantitative vertebral morphometry measurements with the conventional semi-quantitative (SQ) grading using lateral CT scout views. The semi-automated morphometry method showed good to excellent agreement with the visual SQ grading by radiologists for identification of vertebral fractures.

Introduction—Semi-automated quantitative vertebral morphometry (QM) measurements may enhance management of osteoporosis patients by providing an efficient means to identify vertebral fractures (VFX). We compared identification of prevalent VFX by semi-automated QM to SQ grading.

Methods—A non-radiologist performed semi-automated QM from CT lateral scout views in 200 subjects (102 men, 98 women, 65.8±8.9 years) selected from the Framingham Heart Study Multidetector CT Study. VFx were classified in the QM approach based on using Genant's criteria for deformities, and compared with conventional SQ grading performed by experienced radiologists as the gold standard. The kappa (k) statistics, percent agreement (% Agree), sensitivity (SE), specificity (SP), positive predictive value (PPV), and negative predictive value (NPV) were computed.

Results—Among 200 subjects, 57 had mild and 41 had moderate or severe VFx by visual SQ grading. Per-person analyses showed excellent agreement between the two methods, with $k=0.780$. The % Agree ranged from 86.7% to 91.2%, the SE was 81.3%–96%, and the SP was 86.5%–92%. Among 2,588 vertebrae analyzed, 107 had mild and 49 had moderate or severe VFx by visual SQ grading. Per-vertebra analyses revealed good agreement, with $k=0.580$. Agreement between the methods tended to be highest in L1–L4 region. Agreement and validity measures were higher when only moderate and severe fractures were included.

Conclusion—The semi-automated quantitative vertebral morphometry measurements from CT lateral scout views provided good to excellent agreement with the standard SQ grading for assessment of prevalent vertebral fractures.

Keywords

Computed tomography; Quantitative vertebral morphometry; Semi-automated vertebral morphometry; Semi-quantitative; Validity; Vertebral fracture

Introduction

Vertebral fractures are the most common type of osteoporotic fracture, occurring in 30% to 50% of people over the age of 50 years [1, 2]. Vertebral fractures are associated with profound morbidity, such as pain and decreased function, and increased mortality. Moreover, the presence of a vertebral fracture is a strong predictor of future fracture at other skeletal sites, including the hip and radius [3–5]. As many as 20% of women with a prevalent vertebral fracture suffer a new fracture within 1 year [6]. However, despite the marked morbidity associated with vertebral fractures and their association with future fracture risk, prevalent vertebral fractures are often unrecognized on radiographs or quantitative computed tomography (QCT) scans, with as many as 40% to 87% of vertebral fractures unreported by standard clinical review [7–9]. As a number of therapies have been proven to reduce the risk of new vertebral fractures, methods are needed to improve the accuracy and efficiency of identifying vertebral fractures to improve fracture risk assessment and so that proper therapeutic intervention can be initiated.

There are three general approaches to identify vertebral fractures: (1) visual identification of fracture; (2) visual identification using either the semi-quantitative (SQ) grading developed by Genant et al. [10] or the algorithm-based approach for the qualitative identification of vertebral fracture (ABQ) developed by Jiang et al. [11]; and (3) quantitative vertebral morphometry (QM), which was first developed in the 1960s [12]. Although Genant's SQ grading method has been considered as the gold-standard method for vertebral fracture assessment, and its use is recommended by the International Society for Clinical Densitometry (ISCD) for diagnosing fractures by DXA-based vertebral fracture assessment (VFA) [13], it does have limitations, including the requirement for highly experienced readers and modest reproducibility between readers, particularly for mild vertebral fractures [14, 15].

In comparison, vertebral morphometry is a quantitative method to identify vertebral fractures based on the measurement of distinct vertebral dimensions, calculating relative changes (or differences) in vertebral height as indicators of fracture. Its potential advantage over SQ or ABQ methods is that it calculates deformity percentages by directly measured vertebral heights. However, since the traditional QM method relies on the measurement of vertebral dimensions achieved by placement of six morphometry points on the vertebral body, disadvantages include the large time commitment to place six points on each vertebral body, potential variability in point-placement by different readers, and lack of ability to distinguish non-fracture vertebral deformities from fractures.

Altogether, there is still no universally agreed upon gold standard for vertebral fracture diagnosis, and there continues to be disagreement over the optimal approach for identifying vertebral fractures in clinical practice, clinical research, and clinical trials. It is possible that combination of visual SQ grading and more feasible QM approaches may enhance the strengths of each technique while minimizing their limitations, possibly improving the identification of vertebral fractures [16–18].

Recently, there has been renewed interest in QM due to technical developments and introduction of algorithms for semi-automated placement of vertebral morphometry points. Thus, vertebral morphometry has been used for VFA by dual-energy X-ray Absorptiometry (DXA) [13]. More recently, semi-automated vertebral morphometry by shape-based modeling techniques was introduced, and found to be suitable for not only radiographs but also DXA-based VFA, and QCT lateral scout views [19, 20]. Using a semi-automated algorithm for vertebral morphometry, we previously reported good to excellent intra- and inter-reader reliability for vertebral height measurements assessed by non-radiologist readers [20].

Our long-term goal is to improve assessment of vertebral fractures by determining the clinical utility of these novel semi-automated quantitative morphometry measurements. The objective of this study is to compare vertebral fracture assessment by semi-automated quantitative vertebral morphometry measurements with the conventional SQ grading using lateral CT scout views.

Methods

Study subjects

Subjects were selected from the community-based Framingham Heart Study Offspring and Third Generation Multidetector Computed Tomography (MDCT) Study [21–23]. QCT scans of the chest and abdomen were acquired in 3,529 participants (35–87 year) in the MDCT Study for assessment of vascular calcification. Volumetric computed tomography scans were obtained using an eight-slice multidetector computed tomography scanner (Light-speed Ultra/Plus, General Electric Medical Systems, Milwaukee, WI) operating at a tube voltage 120 kVp, tube current 320/400 mA (<220/>220 lb body weight), and gantry rotation of 500 ms. Scans had a nominal in-plane voxel size of 0.68 mm and a slice thickness of 2.5 mm. The scout views used in the current study consisted of frontal and lateral low-energy 2D scanograms extending from the upper thoracic (T4) to sacral (S1) vertebral levels (Fig. 1). A hydroxyapatite phantom (Image Analysis, Columbia, KY) was placed under each individual and scanned concurrently to allow conversion of Hounsfield units to equivalent concentration of calcium hydroxyapatite (grams per cubic centimeter). Integral volumetric bone density (grams per cubic centimeter) of the third lumbar vertebrae (L3) was measured from the lumbar scan using previously published algorithms [24].

We used prevalent vertebral fracture status for each participant, determined by radiologists' semi-quantitative evaluation (see below for methods), and aimed to select a subset of participants, aged 50 years and older, including 30 men and 30 women with prevalent mild vertebral fractures (SQ 1), 30 men and 30 women with prevalent moderate or severe vertebral fractures (SQ 2 or 3); and for each person with fracture, an age- and sex-matched control with no prevalent vertebral fractures. For subjects with multiple vertebral fractures, the vertebral level with the worst fracture severity was used to classify the subject. Whereas we were able to select the desired number of subjects for prevalent SQ 1 fractures, there were only 22 men and 20 women with prevalent SQ 2 or 3 fractures. Thus, the total sample included 104 men (30 with SQ 1, 22 with SQ 2/3, and 52 age-matched with no fracture) and 100 women (30 with SQ 1, 20 with SQ 2/3, and 50 age matched with no fracture), ranging in age from 50 to 87 years. Body weight (kilograms) and height (centimeter) were measured and used to calculate body mass index (BMI, kilograms per square meter).

Prevalent vertebral fracture assessment by visual semi-quantitative method

Lateral CT scout views were used to identify prevalent vertebral fractures in all 3,529 participants using Genant's semi-quantitative algorithm [10]. We previously reported good reliability ($k=0.59$ to 0.69) for assessment of prevalent vertebral fractures applying this method to lateral CT scout views [20]. Two experienced radiologists graded the lateral scout views for prevalent vertebral fractures: no fracture (SQ 0), mild (SQ 1), moderate (SQ 2) or severe (SQ 3) fracture. Radiologists were blinded to subject age, and each evaluated approximately half of participants. CT images identified with a vertebral fracture of any severity were re-evaluated by both radiologists together to generate a consensus fracture score based on visual SQ grading.

Prevalent vertebral fracture assessment by semi-automated quantitative morphometry

Semi-automated quantitative vertebral morphometry was performed by a non-radiologist clinician using a model-based shape recognition technology (SpineAnalyzer, Optasia Medical, Cheadle, UK), as described in previous reports [19, 20]. Briefly, the operator identifies the center of each vertebral body from T4 to L4, and then the algorithm provides detailed annotation to define the shape of each vertebra, from which it derives standard six-point morphometry (Fig. 1). The operator then reviews the annotation and morphometry points, and adjusts them if necessary. Minor adjustments are frequently needed in subjects with moderate to severe vertebral deformities, whereas few adjustments are needed in subjects with no deformities. The average time to analyze a single subject is approximately 5 min [19, 20]. The six morphometry points are then used to compute vertebral heights, vertebral height ratios, and deformity percentages indicative of vertebral fracture by the semi-automated algorithm based solely on standard six-point morphometry. Specifically, we measured posterior (h_p), anterior (h_A), and mid-vertebral height (h_M), and used vertebral height ratios and deformity percentages to classify fracture types such as wedge, biconcave, and crush fractures, the latter being based on Black et al. [25]. Vertebral fractures were identified based on deformity percentages derived from morphometry alone, using the thresholds suggested by Genant's SQ scale as a guide: no fracture (SQ 0, $<20\%$ deformity), mild (SQ 1, 20% deformity $<25\%$), moderate (SQ 2, 25% deformity $<40\%$), and severe (SQ 3, deformity 40%).

Statistical analysis

We compared demographic data such as age, height, weight, BMI, and integral vBMD stratified by sex and by visual SQ grading using analysis of variance and post hoc testing. We compared agreement between vertebral fracture assessment using semi-automated QM measurements and visual SQ grading done by the radiologists using a kappa statistic and associated 95% confidence interval (CI) [26]. 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 [27].

Analyses were performed considering the person as the unit of analysis and the vertebral level as the unit of analysis. For analysis conducted on a person level, we classified each individual who had more than one fracture according to the fracture of greatest severity. For the per-vertebra analysis, we evaluated potential differences in agreement by spinal regions by conducting a stratified analysis by spinal region: T4–T9, T10–T12, and L1–L4, as previously reported [28, 29]. We used two dichotomous definitions of prevalent vertebral fracture: [1] SQ 1 (mild, moderate, and severe) vs. SQ 0 (no fracture), and [2] SQ 2 (moderate and severe) vs. SQ 0 (no fracture). In the comparison of SQ 2 (moderate and severe) vs. SQ 0 (no fracture), SQ1 (mild fractures) were excluded from analysis. We also computed the percent agreement (% Agree), sensitivity (SE), specificity (SP), positive predictive value (PPV), and negative predictive value (NPV) with the SQ grading as the gold standard. The equations for predictive values are as follows where TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

$$\begin{aligned}\text{Sensitivity (SE)} &= \text{TP} / (\text{TP} + \text{FN}) \\ \text{Specificity (SP)} &= \text{TN} / (\text{FP} + \text{TN}) \\ \text{Positive predictive value (PPV)} &= \text{TP} / (\text{TP} + \text{FP}) \\ \text{Negative predictive value (NPV)} &= \text{TN} / (\text{FN} + \text{TN}) \\ \text{Percent agreement (\% Agree)} &= (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \times 100\end{aligned}$$

Statistical analyses were conducted using SPSS (SPSS Inc., Chicago, IL, USA) and p value <0.05 was considered significant.

Results

Subject description and demographic data

Among 204 subjects, two men and two women were excluded due to scoliosis or poor scan quality, leading to a final sample of 200 subjects. One hundred and two subjects had no fracture by visual SQ grading (52 men and 50 women), 57 subjects (29 men and 28 women) were identified with a mild fracture (SQ 1), and 41 subjects (21 men and 20 women) were identified with a moderate or severe fracture (SQ 2 or 3). Of the 2,600 vertebrae evaluated by the readers, 12 vertebrae were unreadable by all readers, and thus, excluded from the analyses. Among the remaining 2,588 vertebrae available for each of the readings, 8 vertebrae (0.3%) from the semi-automated QM and 61 vertebrae (2.4%) from consensus visual SQ grading were unreadable and classified as normal (SQ 0), respectively. Thus, 2,432 vertebrae were SQ 0 by visual grading (1,233 M; 1,199 W), 107 were SQ 1 (62 M; 45 W), and 49 were SQ 2 or 3 (24 M; 25 W). The mean age of the 200 subjects was 65.8 ± 8.9 years. Women with moderate and severe vertebral fractures were shorter and older than those with mild vertebral fractures, and had lower integral bone density than those with no fracture or mild vertebral fractures (Table 1). Other variables were not significantly different among the three groups.

Per-person analysis between semi-automated QM measurements and the consensus visual SQ grading—The semi-automated morphometry-based method identified fewer mild vertebral fractures and more severe vertebral fractures compared to the radiologist consensus SQ readings (Table 2). Per-person analysis comparing semi-automated QM and the consensus visual SQ grading for prevalent fractures showed good to excellent agreement when comparing SQ 0 vs. SQ 1. In particular, $k=0.780$, 0.824 , and 0.734 for T4–

L4 in all subjects, men, and women, respectively (Table 3). The percent agreement between morphometry and visual grading was 86.7% to 91.2%, the sensitivity was 81.3% to 96%, and the specificity was 86.7% to 91.2%. The positive predictive value was slightly higher in women, but the negative predictive value was slightly better in men. Per-person agreement analysis for SQ 0 vs. SQ 2 fractures showed excellent agreement between the methods, with $k=0.824$ to 0.923 (Table 3).

Per-vertebra analysis between the semi-automated QM measurements and the consensus visual SQ grading—Per-vertebra analysis revealed fair to good agreement between the two methods comparing SQ 0 vs. SQ 1, and did not vary substantially by vertebral region (Table 4). For example, for T4–L4 in all subjects, men and women, $k=0.580$, 0.594 , and 0.560 , respectively. The overall percent agreement, specificity, and negative predictive value were better, but the sensitivity and positive predictive value were worse compared to the per-person analysis. Among the different spinal regions, the k , sensitivity, and negative predictive value tended to be better in the L1–L4 region in all subjects, and in both men and women. Similar to what we observed in the per-person analyses, the agreement and validity parameters improved when a fracture definition of SQ 0 vs. SQ 2 was considered, but was still only fair to good with $k=0.659$.

Discussion

In this study, we aimed to determine the potential clinical utility of semi-automated QM assessment of prevalent vertebral fractures by comparing it to conventional visual SQ vertebral fracture grading using lateral CT scout views. We found that the semi-automated QM provided excellent agreement with SQ identification of prevalent vertebral fractures in the per-person analysis, and fair to good agreement for the per-vertebrae analysis.

We previously reported intra- and inter-reader reliability of semi-automated vertebral morphometry measurements and morphometry-based fractures using lateral CT scout views [20]. The semi-automated algorithm provided excellent intra- and inter-reader reliability for vertebral height measurements (intraclass correlation coefficients= 0.96 – 0.98). The semi-automated algorithm also provided good reliability for vertebral fracture assessment based solely on QM ($k=0.59$ to 0.69) that was comparable with previous reports for SQ vertebral fracture grading by radiologists [29]. We also confirmed the rationale for performing semi-automated measurements (i.e., those with operator intervention to adjust morphometry points as necessary) rather than fully automated measurements by demonstrating poorer reliability when the morphometry points were not adjusted, particularly in vertebrae with moderate or severe fractures [20].

There have been several studies showing reasonable agreement for the identification of prevalent and incident vertebral fractures by quantitative morphometry of radiographs or lateral DXA images versus the visual SQ grading [31–33], and one prior study comparing the SQ assessment of vertebral fractures using CT lateral scout views to that with conventional radiography [28]. However, little is known about the validity of applying semi-automated quantitative morphometry to CT lateral scout views in comparison with the visual SQ grading of the same images. In the current study, per-person analysis showed excellent agreement ($k=0.73$ to 0.82) between the two techniques, with the percent agreement ranging from 87% to 91%. The k scores and percent agreement were higher than per-person analyses in previous studies comparing prevalent vertebral fracture detection using Genant's SQ criteria applied to DXA-based VFA versus lateral radiographs in postmenopausal women [34, 35]. Whereas agreement between two different imaging techniques (DXA-VFA vs. radiographs) may be anticipated to be lower than the agreement between two different

methods applied to the same image data (current study), the results here are comparable to the agreement reported in the aforementioned studies.

In general, we found fair to good agreement (k score= 0.58) between the QM-based approach for identification of prevalent vertebral fractures and the visual SQ reading in by-vertebra analyses. There are no prior data comparing semi-automated vertebral morphometry with SQ grading on a per-vertebrae basis, so we compared our results to previous reports on standard vertebral morphometry compared with SQ grading applied to radiographs. Although the k scores of per-vertebra analyses were lower than those of per-person analyses in our study, they were comparable to a prior study (k scores from 0.65 to 0.82) when four different morphometry analyses with commonly used cutoff thresholds were compared to SQ reading on radiographs [32]. It is noteworthy that our results were comparable to this previous study, despite the poorer resolution of CT lateral scout views compared to radiographs. Among the different spinal regions, L1–L4 showed the highest k scores and percent agreement in all subjects and both genders, similar to previous reports [29, 36]. Moreover, the percent agreement, specificity, and negative predictive value were generally higher in per-vertebra analyses compared with per-subject analyses which suggest that this new semi-automated QM algorithm showed higher power to exclude true negatives in the per-vertebra analysis, whereas it diagnosed more true positives in the per-person analysis.

Prior studies have investigated different deformity thresholds for defining vertebral fracture according to morphometric criteria, and reported k scores between quantitative morphometric-defined vertebral fractures and SQ-methods ranging from 0.53 to 0.75 and agreement of 89.6% to 96.4%, results which are similar to our data generated with deformity thresholds based on Genant's criteria [18, 37]. To determine whether agreement between the two approaches for vertebral fracture identification would be better if only moderate and severe fractures were considered, we also analyzed SQ 0 vs. SQ 2 in per-subject and per-vertebra analyses. We found higher k scores and other validity parameters such as percent agreement, sensitivity, specificity, and negative predictive value compared with those of SQ 0 vs. SQ 1 in both per-subject and per-vertebra analyses. This is expected as mild vertebral fractures are challenging to detect by most methods. Thus, the semi-automated algorithm used in our study showed relatively reasonable validity using a commonly used threshold adopted from Genant's SQ grading. Increasing the threshold of deformation used to identify a fracture might lead to higher accuracy than the standard cutoff points, though additional work is needed to explore this strategy.

There are a few limitations that must be considered when interpreting the results. Our vertebral morphometry measures and fracture assessments were based on lateral CT scout views, which are not yet currently used in routine practice for assessment of vertebral fractures. QCT scans are being used more commonly for other clinical reasons, and may be useful for identification of vertebral fractures in clinical practice as well as in osteoporosis research studies. Unlike many previous studies that used radiographs, in the current study the vertebra levels were not marked on the images, and therefore the identification of specific vertebral levels may have differed between the radiologist SQ and the semi-automated QM readings. This might have contributed to the poorer agreement found using the per-vertebra analysis compared to the per-subject analysis. To explore this, we carefully reviewed the level-by-level fracture identification. We identified ten subjects in whom the vertebral levels likely differed between the SQ and morphometry assessments. After correcting this mismatch, the per-vertebrae k scores improved from $k=0.58$ to $k=0.64$ (95% CI; 0.58, 0.70) for SQ 0 vs. SQ 1 and from $k=0.66$ to $k=0.73$ (95% CI; 0.63, 0.82) for SQ 0 vs. SQ 2, suggesting that indeed there was some mismatching of vertebral levels between the two techniques. It should also be noted that good agreement between two techniques

does not necessarily imply accuracy, particularly when there is no accepted gold-standard method, so the results should be interpreted with caution. Lastly, for the semi-automated algorithm we identified the prevalent vertebral fractures solely by quantitative morphometry, and therefore did not exclude any non-fracture deformities that would presumably not been identified as fractures by the radiologist SQ readings. This is a well-known, inherent limitation of current morphometry-based approaches, yet despite this, agreement between the two methods was reasonably good.

The optimal clinical use of the semi-automated morphometry algorithm has not been defined. However, because it can be performed by operators with limited training in a reasonable amount of time, these semi-automated morphometry measurements may be best used as a complement to SQ-based approaches to identify vertebral fractures; for example, the triage approach which was applied by Black et al. [25] and McCloskey et al. [38]. The excellent negative predictive value suggests that this semi-automated morphometry may be useful in confirming vertebrae that are normal, thereby reducing the number of images that need to be reviewed by a reader with expertise in visual assessment of vertebral fractures. Furthermore, the quantitative morphometry measurements could be used in conjunction with semi-automated computer-vision based algorithms to identify fractures, as previously demonstrated [39–44].

The current study also has a number of strengths. Participants were members of a well-characterized community-based cohort making results generalizable to a Caucasian population. We did, however, include a greater proportion of subjects with fractures than would be found in a random population-based sample of similar age. This was done to increase the number of fracture cases in the study, thereby improving the accuracy of the reliability measurements. The range of fracture severity in our study allowed us to evaluate performance of the semi-automated QM algorithm in detecting both mild and moderate fractures. In addition, like our previous reliability study, a non-radiologist reader performed the semi-automated vertebral morphometry, thereby establishing the practical utility of semi-automated vertebral morphometry by technical staff and/or physicians at clinics who may not have advanced training in radiology.

Both clinical and morphometric vertebral fractures result in significant morbidity and are strong risk factors for future fracture, and thus, current underreporting of vertebral fractures is problematic and new methods to facilitate and enhance the detection of vertebral fractures are needed. The current results indicate good to excellent agreement between semi-automated vertebral morphometry-based assessment of prevalent vertebral fractures and an established semi-quantitative method. Semi-automated vertebral morphometry may complement current methods for identifying vertebral fractures.

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, K01AR053118, and by the National Heart, Lung, and Blood Institute (NHLBI) Framingham Heart Study (NIH/NHLBI Contract N01-HC-25195). We acknowledge Peter Steiger, PhD of Optasia Medical for assistance with installation and maintenance of software.

Abbreviations

ABQ	Algorithm-based approach for the qualitative identification of vertebral fracture
BMI	Body mass index

CI	Confidence interval
cm	Centimeter
CT	Computed tomography
CV	Coefficient of variation
DXA	Dual-energy X-ray Absorptiometry
k	Kappa
kg	Kilogram
MRX	Morphometric radiography
MXA	Morphometric X-ray absorptiometry
NPV	Negative predictive value
PPV	Positive predictive value
QCT	Quantitative computed tomography
QM	Quantitative vertebral morphometry
SE	Sensitivity
SP	Specificity
SQ	Semi-quantitative
VFA	Vertebral fracture assessment
VFx	Vertebral fracture

References

1. 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. A meta-analysis of previous fracture and subsequent fracture risk. *Bone*. 2004; 35:375–382. [PubMed: 15268886]
2. Melton LJ 3rd, Kan SH, Frye MA, Wahner HW, O'Fallon WM, Riggs BL. Epidemiology of vertebral fractures in women. *Am J Epidemiol*. 1989; 129:1000–1011. [PubMed: 2784934]
3. Klotzbuecher CM, Ross PD, Landsman PB, Abbott TA 3rd, Berger M. Patients with prior fractures have an increased risk of future fractures: a summary of the literature and statistical synthesis. *J Bone Miner Res*. 2000; 15:721–739. [PubMed: 10780864]
4. Melton LJ 3rd, Lane AW, Cooper C, Eastell R, O'Fallon WM, Riggs BL. Prevalence and incidence of vertebral deformities. *Osteoporos Int*. 1993; 3:113–119. [PubMed: 8481586]
5. van Helden S, Cals J, Kessels F, Brink P, Dinant GJ, Geusens P. Risk of new clinical fractures within 2 years following a fracture. *Osteoporos Int*. 2006; 17:348–354. [PubMed: 16378167]
6. Lindsay R, Silverman SL, Cooper C, Hanley DA, Barton I, Broy SB, Licata A, Benhamou L, Geusens P, Flowers K, Stracke H, Seeman E. Risk of new vertebral fracture in the year following a fracture. *Jama*. 2001; 285:320–323. [PubMed: 11176842]
7. Gehlbach SH, Bigelow C, Heimisdottir M, May S, Walker M, Kirkwood JR. Recognition of vertebral fracture in a clinical setting. *Osteoporos Int*. 2000; 11:577–582. [PubMed: 11069191]
8. Majumdar SR, Kim N, Colman I, Chahal AM, Raymond G, Jen H, Siminoski KG, Hanley DA, Rowe BH. 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*. 2005; 165:905–909. [PubMed: 15851642]

9. Williams AL, Al-Busaidi A, Sparrow PJ, Adams JE, Whitehouse RW. Under-reporting of osteoporotic vertebral fractures on computed tomography. *Eur J Radiol.* 2009; 69:179–183. [PubMed: 17913429]
10. Genant HK, Wu CY, van Kuijk C, Nevitt MC. Vertebral fracture assessment using a semiquantitative technique. *J Bone Miner Res.* 1993; 8:1137–1148. [PubMed: 8237484]
11. Jiang G, Eastell R, Barrington NA, Ferrar L. Comparison of methods for the visual identification of prevalent vertebral fracture in osteoporosis. *Osteoporos Int.* 2004; 15:887–896. [PubMed: 15071725]
12. Hurxthal LM. Measurement of anterior vertebral compressions and biconcave vertebrae. *Am J Roentgenol Radium Ther Nucl Med.* 1968; 103:635–644.
13. Schousboe JT, Vokes T, Broy SB, Ferrar L, McKiernan F, Roux C, Binkley N. Vertebral fracture assessment: the 2007 ISCD official positions. *J Clin Densitom.* 2008; 11:92–108. [PubMed: 18442755]
14. Ferrar L, Jiang G, Schousboe JT, DeBold CR, Eastell R. Algorithm-based qualitative and semiquantitative identification of prevalent vertebral fracture: agreement between different readers, imaging modalities, and diagnostic approaches. *J Bone Miner Res.* 2008; 23:417–424. [PubMed: 17967136]
15. Prevral S, Kregg JH, Chen P, Genant H, Black DM. Teriparatide vertebral fracture risk reduction determined by quantitative and qualitative radiographic assessment. *Curr Med Res Opin.* 2009; 25:921–928. [PubMed: 19250060]
16. Grados F, Fechtenbaum J, Flipon E, Kolta S, Roux C, Fardellone P. Radiographic methods for evaluating osteoporotic vertebral fractures. *Joint Bone Spine.* 2009; 76:241–247. [PubMed: 19196531]
17. Black DM, Cummings SR, Stone K, Hudes E, Palermo L, Steiger P. A new approach to defining normal vertebral dimensions. *J Bone Min Res.* 1991; 6:883–892.
18. Genant HK, Jergas M, Palermo L, Nevitt M, Valentin RS, Black D, Cummings SR. 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.* 1996; 11:984–996. [PubMed: 8797120]
19. Brett A, Miller CG, Hayes CW, Krasnow J, Ozanian T, Abrams K, Block JE, van Kuijk C. Development of a clinical workflow tool to enhance the detection of vertebral fractures: accuracy and precision evaluation. *Spine (Phila Pa 1976).* 2009; 34:2437–2443. [PubMed: 19829258]
20. Kim YM, Demissie S, Eisenberg R, Samelson EJ, Kiel DP, Bouxsein ML. Intra- and inter-reader reliability of semi-automated quantitative morphometry measurements and vertebral fracture assessment using lateral scout views from computed tomography. *Osteoporos Int.* 2011; 10.1007/s00198-011-1530-4
21. Rosito GA, Massaro JM, Hoffmann U, Ruberg FL, Mahabadi AA, Vasan RS, O'Donnell CJ, Fox CS. Pericardial fat, visceral abdominal fat, cardiovascular disease risk factors, and vascular calcification in a community-based sample: the Framingham Heart Study. *Circulation.* 2008; 117:605–613. [PubMed: 18212276]
22. Moselewski F, O'Donnell CJ, Achenbach S, Ferencik M, Massaro J, Nguyen A, Cury RC, Abbata S, Jang IK, Brady TJ, Hoffmann U. Calcium concentration of individual coronary calcified plaques as measured by multidetector row computed tomography. *Circulation.* 2005; 111:3236–3241. [PubMed: 15956133]
23. Pou KM, Massaro JM, Hoffmann U, Lieb K, Vasan RS, O'Donnell CJ, Fox CS. Patterns of abdominal fat distribution: the Framingham Heart Study. *Diabetes Care.* 2009; 32:481–485. [PubMed: 19074995]
24. Lang T, LeBlanc A, Evans H, Lu Y, Genant H, Yu A. Cortical and trabecular bone mineral loss from the spine and hip in long-duration spaceflight. *J Bone Miner Res.* 2004; 19:1006–1012. [PubMed: 15125798]
25. Black DM, Palermo L, Nevitt MC, Genant HK, Epstein R, San Valentin R, Cummings SR. Comparison of methods for defining prevalent vertebral deformities: the Study of Osteoporotic Fractures. *J Bone Miner Res.* 1995; 10:890–902. [PubMed: 7572313]
26. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas.* 1960; 20:37–46.

27. Fleiss, J. Statistical Method for Rates and Proportions. John Wiley and Sons, Inc; New York: 1981. The measurement of interrater agreement; p. 211-236.
28. Takada M, Wu CY, Lang TF, Genant HK. Vertebral fracture assessment using the lateral scoutview of computed tomography in comparison with radiographs. *Osteoporos Int.* 1998; 8:197–203. [PubMed: 9797902]
29. 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, Boussein ML. Reliability of vertebral fracture assessment using multidetector CT lateral scout views: the Framingham Osteoporosis Study. *Osteoporos Int.* 2011; 22:1123–1131. [PubMed: 20495902]
30. Kim SH, Choi HS, Rhee Y, Kim KJ, Lim SK. Prevalent vertebral fractures predict subsequent radiographic vertebral fractures in postmenopausal Korean women receiving antiresorptive agent. *Osteoporos Int.* 2011; 22:781–787. [PubMed: 20533028]
31. Ferrar L, Jiang G, Barrington NA, Eastell R. Identification of vertebral deformities in women: comparison of radiological assessment and quantitative morphometry using morphometric radiography and morphometric X-ray absorptiometry. *J Bone Miner Res.* 2000; 15:575–585. [PubMed: 10750573]
32. Grados F, Roux C, de Vernejoul MC, Utard G, Sebert JL, Fardellone P. Comparison of four morphometric definitions and a semiquantitative consensus reading for assessing prevalent vertebral fractures. *Osteoporos Int.* 2001; 12:716–722. [PubMed: 11605736]
33. Hospers IC, van der Laan JG, Zeebregts CJ, Nieboer P, Wolffenbuttel BH, Dierckx RA, Kreeftenberg HG, Jager PL, Slart RH. Vertebral fracture assessment in supine position: comparison by using conventional semiquantitative radiography and visual radiography. *Radiology.* 2009; 251:822–828. [PubMed: 19380691] *Osteoporos Int.* 2012; 23:1007–1016. 1015. [PubMed: 21927924]
34. Rea JA, Chen MB, Li J, Blake GM, Steiger P, Genant HK, Fogelman I. Morphometric X-ray absorptiometry and morphometric radiography of the spine: a comparison of prevalent vertebral deformity identification. *J Bone Miner Res.* 2000; 15:564–574. [PubMed: 10750572]
35. Fuerst T, Wu C, Genant HK, von Ingersleben G, Chen Y, Johnston C, Econs MJ, Binkley N, Vokes TJ, Crans G, Mitlak BH. Evaluation of vertebral fracture assessment by dual X-ray absorptiometry in a multicenter setting. *Osteoporos Int.* 2009; 20:1199–1205. [PubMed: 19083074]
36. Chapurlat RD, Duboeuf F, Marion-Audibert HO, Kalpakcioglu B, Mitlak BH, Delmas PD. Effectiveness of instant vertebral assessment to detect prevalent vertebral fracture. *Osteoporos Int.* 2006; 17:1189–1195. [PubMed: 16758133]
37. Wu CY, Li J, Jergas M, Genant HK. Comparison of semiquantitative and quantitative techniques for the assessment of prevalent and incident vertebral fractures. *Osteoporos Int.* 1995; 5:354–370. [PubMed: 8800786]
38. McCloskey EV, Vasireddy S, Threlkeld J, Eastaugh J, Parry A, Bonnet N, Beneton M, Kanis JA, Charlesworth D. Vertebral fracture assessment (VFA) with a densitometer predicts future fractures in elderly women unselected for osteoporosis. *J Bone Miner Res.* 2008; 23:1561–1568. [PubMed: 18505372]
39. Roberts MG, Cootes TF, Adams JE. Vertebral shape: automatic measurement with dynamically sequenced active appearance models. *Med Image Comput Comput Assist Interv.* 2005; 8:733–740. [PubMed: 16686025]
40. de Bruijne M, Lund MT, Tanko LB, Pettersen PP, Nielsen M. Quantitative vertebral morphometry using neighbor-conditional shape models. *Med Image Comput Comput Assist Interv.* 2006; 9:1–8. [PubMed: 17354867]
41. Roberts M, Cootes TF, Adams JE. Vertebral morphometry: semiautomatic determination of detailed shape from dual-energy X-ray absorptiometry images using active appearance models. *Invest Radiol.* 2006; 41:849–859. [PubMed: 17099422]
42. Roberts M, Cootes T, Pacheco E, Adams J. Quantitative vertebral fracture detection on DXA images using shape and appearance models. *Acad Radiol.* 2007; 14:1166–1178. [PubMed: 17889334]

43. de Bruijne M, Lund MT, Tanko LB, Pettersen PC, Nielsen M. Quantitative vertebral morphometry using neighbor-conditional shape models. *Med Image Anal.* 2007; 11:503–512. [PubMed: 17720611]
44. Roberts MG, Pacheco EM, Mohankumar R, Cootes TF, Adams JE. Detection of vertebral fractures in DXA VFA images using statistical models of appearance and a semiautomatic segmentation. *Osteoporos Int.* 2010; 21:2037–2046. [PubMed: 20135093]

Fig. 1. Method used to conduct semi-automated quantitative morphometry measurements (SpineAnalyzer, Optasia Medical, Cheadle, UK). **a** CT lateral scout view of the vertebrae from T4 to L4; **b** shape contours of the vertebrae from T4 to L4; **c** semi-automated placement of standard quantitative morphometry points of T12; **d** 95 points placed for shape contours of T12

Table 1

Subject description stratified by sex and by visual SQ grading by the radiologists

	SQ 0	SQ 1	SQ 2
<i>All (n=200)</i>			
Number (n)	102 (51.0%)	57 (28.5%)	41 (20.5%)
Sex (men%)	51.0	50.9	51.2
Age (years)	65.6±8.9	64.2±7.8	68.6±10.0 ^{**}
Height (cm)	169.5±10.7	167.9±8.6	166.9±12.7
Weight (kg)	82.4±16.9	78.3±14.4	80.0±20.1
BMI (kg/m ²)	28.8±5.3	27.7±4.6	28.4±5.2
L3 Integral vBMD (g/cm ³)	0.168±0.040	0.158±0.031	0.142±0.039 ^{*,**}
<i>Men (n=102)</i>			
Number (n)	52 (51.0%)	29 (28.4%)	21 (20.6%)
Age (years)	65.2±8.0	64.6±6.7	66.7±9.7
Height (cm)	177.0±7.7	174.4±6.0	177.0±8.7
Weight (kg)	89.6±15.7	87.0±10.4	91.7±17.7
BMI (kg/m ²)	28.7±5.3	28.6±3.4	29.1±4.1
L3 Integral vBMD (g/cm ³)	0.174±0.039	0.162±0.026	0.163±0.038
<i>Women (n=98)</i>			
Number (n)	50 (51.0%)	28 (28.6%)	20 (20.4%)
Age (years)	66.0±9.8	63.8±9.0	70.6±10.2 ^{**}
Height (cm)	161.3±6.7	161.3±5.2	156.2±5.4 ^{*,**}
Weight (kg)	74.8±14.6	69.3±12.3	68.2±14.9
BMI (kg/m ²)	28.8±5.3	26.7±5.4	27.6±6.0
L3 Integral vBMD (g/cm ³)	0.162±0.041	0.154±0.036	0.118±0.024 ^{*,**}

*
 $p < 0.05$ comparing SQ 0 vs. SQ 2,**
 $p < 0.05$ comparing SQ 1 vs. SQ 2 by pairwise tests

Table 2

Number of subjects in SQ grouping, SQ 0, SQ 1, and SQ 2 by semi-automated vertebral morphometry or consensus SQ grading by the radiologists

Semi-automated QM	Consensus QM grading			
	SQ 0	SQ 1	SQ 2	Total
All				
SQ 0	91	8	3	102
SQ 1	9	25	4	38
SQ 2	2	24	34	60
Total	102	57	41	200
Men				
SQ 0	45	0	2	47
SQ 1	6	15	2	23
SQ 2	1	14	17	32
Total	52	29	21	102
Women				
SQ 0	46	8	1	55
SQ 1	3	10	2	15
SQ 2	1	10	17	28
Total	50	28	20	98

Table 3

Per-subject agreement between the consensus visual semi-quantitative (SQ) grading by the radiologists and semi-automated quantitative morphometry (QM) for prevalent vertebral fracture identification

Consensus SQ grading vs. Semi-automated QM						
T4-L4	k (95% CI)	% Agree	SE (%)	SP (%)	PPV (%)	NPV (%)
All subjects (n=200)						
SQ 0 vs. SQ 1	0.780 (0.694, 0.866)	89.0	89.2	88.8	89.2	88.8
SQ 0 vs. SQ 2	0.905 (0.823, 0.987)	96.2	91.9	97.8	94.4	96.8
Men (n=102)						
SQ 0 vs. SQ 1	0.824 (0.714, 0.934)	91.2	96.0	86.5	87.3	95.7
SQ 0 vs. SQ 2	0.887 (0.762, 1.012)	95.4	89.5	97.8	94.4	95.7
Women (n=98)						
SQ 0 vs. SQ 1	0.734 (0.601, 0.867)	86.7	81.3	92.0	90.7	83.6
SQ 0 vs. SQ 2	0.923 (0.819, 1.027)	96.9	94.4	97.9	94.4	97.9

k kappa statistics, % Agree percent agreement, SE sensitivity, SP specificity, PPV positive predictive value, NPV negative predictive value

Table 4

Per-vertebra agreement between the consensus visual semi-quantitative (SQ) grading by the radiologists and semi-automated quantitative morphometry (QM) for prevalent vertebral fracture identification

Consensus SQ grading vs. semi-automated QM						
Fractures included	k (95% CI)	Agree (%)	SE (%)	SP (%)	PPV (%)	NPV (%)
All subjects (n=2588)						
SQ 0 vs. SQ 1						
T4-L4	0.580 (0.517, 0.643)	94.6	69.2	96.3	54.3	98.0
T4-T9	0.579 (0.489, 0.669)	94.3	63.1	96.7	58.9	97.2
T10-T12	0.550 (0.436, 0.664)	92.3	78.6	93.4	47.1	98.3
L1-L4	0.621 (0.484, 0.758)	96.8	73.3	97.8	56.4	98.9
SQ 0 vs. SQ 2						
T4-L4	0.659 (0.555, 0.763)	98.5	81.8	98.8	56.3	99.7
Men (n=1319)						
SQ 0 vs. SQ 1						
T4-L4	0.594 (0.512, 0.676)	94.1	75.6	95.4	53.3	98.2
T4-T9	0.615 (0.499, 0.731)	94.1	67.3	96.4	62.3	97.1
T10-T12	0.543 (0.398, 0.688)	90.8	87.0	91.2	44.4	98.9
L1-L4	0.615 (0.431, 0.799)	96.5	85.7	96.9	50.0	99.5
SQ 0 vs. SQ 2						
T4-L4	0.659 (0.512, 0.806)	98.5	85.7	98.7	54.5	99.7
Women (n=1269)						
SQ 0 vs. SQ 1						
T4-L4	0.560 (0.462, 0.658)	95.2	61.4	97.2	55.8	97.7
T4-T9	0.527 (0.382, 0.672)	94.5	57.1	96.9	54.1	97.3
T10-T12	0.558 (0.376, 0.740)	93.9	68.4	95.6	52.0	97.8
L1-L4	0.630 (0.428, 0.832)	97.1	62.5	98.7	66.7	98.4
SQ 0 vs. SQ 2						
T4-L4	0.659 (0.512, 0.806)	98.5	78.3	98.9	58.1	99.6

k kappa statistics, % Agree percent agreement, SE sensitivity, SP specificity, PPV positive predictive value, NPV negative predictive value