High precision semiautomated computed tomography measurement of lumbar disk and vertebral heights

Sovira Tan^{a)}

National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health,
Building 10 CRC Room 4-1339, Bethesda, Maryland 20892

Jianhua Yao and Lawrence Yao

Radiology and Imaging Sciences, National Institutes of Health, Building 10, Bethesda, Maryland 20892

Michael M. Ward

National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health,
Building 10 CRC Room 4-1339, Bethesda, Maryland 20892

(Received 11 May 2012; revised 31 October 2012; accepted for publication 7 November 2012;
published 18 December 2012)

Purpose: Evaluation of treatments of many spine disorders requires precise measurement of the heights of vertebral bodies and disk spaces. The authors present a semiautomated computer algorithm measuring those heights from spine computed tomography (CT) scans and evaluate its precision.

Methods: Eight patients underwent two spine CT scans in the same day. In each scan, five thoracolumbar vertebral heights and four disk heights were estimated using the algorithm. To assess precision, the authors computed the differences between the height measurements in the two scans, coefficients of variation (CV), and 95% limits of agreement. Intraoperator and interoperator precisions were evaluated. For local vertebral and disk height measurement (anterior, middle, posterior) the algorithm was compared to a manual mid-sagittal plane method.

Results: The mean (standard deviation) interscan difference was as low as 0.043 (0.031) mm for disk heights and 0.044 (0.043) mm for vertebral heights. The corresponding 95% limits of agreement were $[-0.085, 0.11]$ and $[-0.10, 0.12]$ mm, respectively. Intraoperator and interoperator precision was high, with a maximal CV of 0.30%. For local vertebral and disk heights, the algorithm improved upon the precision of the manual mid-sagittal plane measurement by as much as a factor of 6 and 4, respectively.

Conclusions: The authors evaluated the precision of a novel computer algorithm for measuring vertebral body heights and disk heights using short term repeat CT scans of patients. The 95% limits of agreement indicate that the algorithm can detect small height changes of the order of 0.1 mm.

© 2013 American Association of Physicists in Medicine. [<http://dx.doi.org/10.1118/1.4769412>]

Key words: intervertebral disk height, vertebral height, semiautomated measurement, precision/repeatability, CT.

I. INTRODUCTION

Reliable measurement of intervertebral disk and vertebral height may benefit many clinical and research applications. Disk height loss is associated with disk degeneration and low back pain.^{1,2} The influence of disk height on the mechanical properties of the spine and of mechanical loading on disk height are subjects of considerable interest.³⁻⁵ Disk height measurements are used to optimize the fit of implants and prostheses^{6,7} and to assess the outcome of surgical interventions.^{8,9} Similarly, vertebral height has been used to assess the effectiveness of height restoration procedures.^{10,11} Vertebral height measurement is also important in assessment of vertebral fractures. Loss of vertebral height characterizes vertebral compression fractures, a hallmark of osteoporosis,¹²⁻¹⁵ and also a consequence of metastasis.¹⁶

High precision is particularly important when small time dependent changes are of interest. For disk height, these include diurnal variations¹⁷ and age related changes.^{18,19} In osteoporosis research, studies have monitored the progression of

vertebral deformities over years.^{20,21} A recent study reported variations in vertebral dimensions (including vertebral height) in a group of women over the course of 6 yr.²² Clinical trials have also measured vertebral heights over years to assess the efficacy of osteoporosis treatments.^{23,24}

Disk height has traditionally been measured using radiography and magnetic resonance imaging (MRI), while vertebral height has been measured using radiography and increasingly dual emission x-ray absorptiometry (DXA). The limitations associated with radiography (use of 2D images to assess 3D structures and impaired depiction of key anatomy by superimposed structures) may compromise its precision. DXA also produces 2D images, and has the additional drawback of lower resolution but does not suffer from central projection distortion.

Computed tomography (CT) has also been investigated for disk and vertebral height measurement.^{6,7,9,10,13} CT and MRI provide complete 3D information about anatomical structures. CT can achieve higher resolution and is better suited to bone visualization. Bauer *et al.* reported that CT compared

favorably to radiography for fracture detection in osteoporosis provided the slice thickness was 3 mm or less.¹³ However, use of CT increases radiologists' reading time. Computer driven automation can mitigate this problem while improving intrareader and inter-reader variability. Various degrees of automation have been attained for radiography,^{25–27} DXA,²⁸ and MRI.¹⁸ For CT, an automated method has recently been proposed, but because this method performs measurements only on the mid-sagittal slice, it reduces to a 2D measure.²⁹ We present an algorithm that takes advantage of the full 3D information and is mostly automatic.

A precision study allows one to quantify variability associated with repeat measurements, which is crucial for determining what can reliably be considered true change. We present a precision study of our approach using same day repeat scans. Most precision or reproducibility studies reported in the literature are limited to an investigation of intrareader and inter-reader variability, with subjects imaged only once. Reader variability is evaluated but not the variability introduced by the imaging modality. However, each modality has its limitations, distortions, and artifacts. These can depend on the position of the patient relative to the imaging system, variations in which are usually unavoidable in longitudinal studies. Precision studies that assess the impact of imaging variability require repeat scans of the subjects, with time between scans short enough to exclude any real change.

II. MATERIALS AND METHODS

II.A. Patients and scanning protocol

The study protocol was approved by the institutional review board and subjects provided written informed consent. They were eight patients with ankylosing spondylitis (7 men, 1 woman) with a mean age of 57 yr (ranging from 32 to 70 yr). Five patients were scanned on Philips Brilliance 64 and three were scanned on a GE Lightspeed Ultra. The former is 64-detector row scanner, while the latter is an older 8-detector row scanner. For both scanners voltage and current parameters were 120 kVp, 300 mAs, respectively. Slice thickness was 1.5 and 1.25 mm, respectively, for the Philips and GE. Spacing between slices was 0.7 and 0.625 mm for the Philips and GE, respectively. For each scan, three reconstructions were made at field of view (FOV) diameters 25, 40, and 48 cm, corresponding to axial in-plane pixel sizes of about 0.49, 0.78, and 0.94 mm, which we call high, medium, and low resolutions. The time between the 2 scans was a few minutes. After the first scan, patients were made to get out of the scanner and stand up before lying down again for the second scan. This ensured the variation is within the range expected in a longitudinal study. Each patient was scanned from the middle of the T10 vertebra to the middle of the L4 vertebra so that for each of them four intervertebral disk spaces (IDS) and five vertebral bodies could be processed. All patients were scanned with a QCT phantom (Mindways Software Inc., Austin, TX). The QCT phantom consists of five tubes of K_2HPO_4 of different densities.

II.B. Methods

II.B.1. Computer algorithm

Figure 1 provides an overview of the complete algorithm which consists of three main steps. First, vertebral bodies are segmented in the CT scans. The end plates of the vertebral bodies are then extracted. Finally, vertebral and disk heights are computed as distances between end plates.

II.B.1.a. Segmentation of vertebral bodies. Many image processing techniques have previously been applied to the segmentation of vertebral bodies in CT.^{30–33} Our algorithm is based on level sets. Level sets are evolving contours or surfaces that can expand, contract, and even split or merge.³⁴ For the purpose of segmentation they are designed to deform so as to match an object of interest and stop at its edges. In the present work, the main level set used was the geodesic active contour (GAC). Its evolution equation can be written as³⁵

$$\frac{d\psi}{dt} = \alpha g(\vec{x})c|\nabla\psi| + \beta g(\vec{x})\kappa|\nabla\psi| + \gamma \nabla g(\vec{x})\nabla\psi. \quad (1)$$

The evolving contour is encoded as the zero level set of the distance function $\psi(\vec{x}, t)$. In other words points that verify $\psi(\vec{x}, t) = 0$ form the contour. By convention the distance is

FIG. 1. Overview of the complete computer algorithm. In (e) the right hand side of the top vertebral surface was cut out to allow a lateral view inside the vertebral body (g).

negative for points inside the contour and positive for the ones outside:

$$\psi(\vec{x}, t) = \pm d, \quad (2)$$

where d is the distance from point $\vec{x}$ to the zero level set contour. Our algorithm is 3D and $\vec{x}$ here represents the three space coordinates.

The first term on the right-hand side of Eq. (1) is the propagation term that makes the contour move with velocity c . The second term, the curvature term, controls the smoothness of the contour using the mean curvature κ . The third term, the advection term, locks the contour to the boundary. The spatial function $g(\vec{x})$, derived from the images to be segmented, contains information about the objects' boundaries and is called the speed function. The parameters α , β , and γ weight the importance of each term. Those parameters were optimized using a previously published method.³⁶

The speed function $g(\vec{x})$ should ideally have values close to 1 where there are no boundaries (so that the level set can expand rapidly) and values close to 0 where boundaries are present (so that the level set stops). This can be achieved, for instance, by writing³⁷

$$g(\vec{x}) = 1 - \frac{1}{1 + \exp\left(-\frac{I-\xi}{\eta}\right)}, \quad (3)$$

where I is the gradient magnitude of the grey level image at voxel $\vec{x}$. The two parameters ξ and η are typically computed using the equations³⁷

$$\eta = \frac{K_1 - K_2}{6}, \quad \xi = \frac{K_1 + K_2}{2}, \quad (4)$$

where K_1 is the minimum gradient magnitude value along the object's boundary and K_2 is the average gradient magnitude inside the object where the level set is initialized. Those definitions ensure that the level set advances over internal gradients but stops at the minimum gradient along the boundary, as Eq. (3) maps gradient values up to K_2 to approximately 1 and gradient values equal or larger than K_1 to approximately 0. An algorithm for evaluating K_1 and K_2 has been described.³⁶

To segment the different parts of a vertebral body (trabecular and cortical bone), successive level sets were used. The segmentation starts from a seed manually placed roughly in the center of the vertebral body. The user approximately selects the mid-sagittal slice by maximizing the visualization of the spinous process on the sagittal reformation. On the mid-sagittal slice, the user roughly locates the center of the vertebral body, which provides the z-coordinate of the seed. The user is then presented with the axial slice corresponding to this z-coordinate. The user again approximately locates the center of the vertebral body, which provides the x- and y-coordinates of the seed. As a preprocessing step, Gaussian smoothing with a standard deviation of 0.5 mm is applied to the image.

The level set equation is solved iteratively using a Runge-Kutta discretization³⁸ of Eq. (1):

$$\begin{aligned} \psi_{t+1} = & \psi_t + (\alpha g(\vec{x})c|\nabla\psi_t| + \beta g(\vec{x})\kappa|\nabla\psi_t| \\ & + \gamma \nabla g(\vec{x})\nabla\psi_t)\Delta t. \end{aligned} \quad (5)$$

Figure 1(b) shows an example of final segmentation result. Once a vertebral body is segmented in the CT volume, a triangular mesh is generated to represent its surface [Fig. 1(c)]. Triangular meshes are a common way to model surfaces in computer graphics.³⁹ Surfaces are discretized into small triangles facilitating the visualization of 3D objects.

II.B.1.b. Segmentation of vertebral end plates. We extract the end plates from the triangular meshes representing the surfaces of the vertebrae using the same level set as Eq. (1), but transposed from the Cartesian domain of rectangular grids to the domain of a surface mesh. Whereas the most relevant feature of rectangular grids of CT scans are grey level gradients, the useful features on a surface mesh are curvature measures (the vertebral body surface is more curved at the ridgelines than on the end plates). The curvature measure we used is curvedness (C):⁴⁰

$$C = \sqrt{\frac{\kappa_1^2 + \kappa_2^2}{2}}, \quad (6)$$

where κ_1 and κ_2 are the principal curvatures. This measure can be computed locally at each vertex of the triangles that form the mesh. The more curved the local surface is, the larger C is. The equations for the speed function remain the same as Eqs. (3) and (4), except that grey level gradients are replaced by gradients of curvature measures. To implement this level set, we used a previously published technique³⁶ but with a new seeding strategy. The previous algorithm relied on the same manually placed seed that initiated the segmentation of the vertebral body. This seed was projected on the surface mesh along the z-axis. The sites of the projections on the mesh were taken as the seeds for the surface level sets. From those seeds two approximate 10 mm radius circles were formed on the upper and lower end plates, respectively. These initial contours were then allowed to grow until they met the ridgelines of the end plates.³⁶ Our new seeding strategy removes the dependency on the manually placed seed. By exploiting the characteristic geometry of the vertebral body, we also sought to make the initial contour closer to the end plates' ridgelines. As a consequence, fewer iterations are needed to complete the segmentation. The new seeding technique reduces the number of iterations from 80 to just 30.

The new seeding strategy is based on the observation that the end plates form large areas of near flat surface. Detecting flat areas on the mesh can be performed using the curvedness measure. In a previous study analyzing the distribution of C values on vertebral end plates,³⁶ we found that the distribution peaked around the value 0.1. Using 0.1 as a threshold, we consider vertices with C values of 0.1 or less to be on a locally flat patch of surface. As a shorthand, we will call vertices on a flat patch of surface "flat-area vertices." Fig. 2(a) shows flat-area vertices (white) on a vertebral body. It can be seen that the condition, $C \leq 0.1$, is not sufficient to distinguish end plate vertices from vertices on the side of the vertebral body that also contains large locally flat areas. However, another condition can be used to discriminate between the end plates and the side of the vertebral body. Due to the geometry of the spine and vertebral body as well as the

FIG. 2. New seeding technique. (a) Flat-area vertices. (b) Flat-area vertices on regions of the surface oriented more horizontally than vertically. (c) New seed. (d) Seed obtained with the previous algorithm. Results after 30 iterations for (e) the new seed and (f) the previous seed.

patient's position on the scanner table, the end plates tend to be oriented horizontally while the side of the vertebral body is more vertically oriented. On the mesh, we compute the local normal $\vec{n}$ to the surface at each vertex. We record all vertices verifying

$$C \leq 0.1 \quad \text{and} \quad |n_z| > |n_x| \quad \text{and} \quad |n_z| > |n_y|. \quad (7)$$

The additional conditions on the normal select vertices on locations where the local surface is oriented more horizontally than vertically. Figure 2(b) shows that those requirements eliminate most of the vertices on the side of the vertebral body. The remaining flat-area vertices can then be clustered using a simple connected component labeling algorithm. The two largest clusters [Fig. 2(c)] correspond to the upper and lower end plates. Those clusters are used as the initialization for the level set. For comparison, we show the initialization obtained with the previous algorithm [Fig. 2(d)]. The new initialization is much closer to the complete end plate. The rest of the algorithm follows the original algorithm.³⁶ Figure 2(e) shows the segmented end plate from the new initialization after 30 iterations. Figure 2(f) shows the state of the segmentation from the old initialization after 30 iterations. Finally, the

direction of the normal to each segmented end plate is evaluated using a least square method.³⁶

II.B.1.c. Computation of vertebral and disk heights. Vertebral heights and disk heights are computed, respectively, as the distances between end plates of the same vertebra and of adjacent vertebrae [Figs. 1(e)]. Distances between end plates are computed as the means of individual distance measures between corresponding points. Such individual distance measures are represented by gray segments of line in Figs. 1(f) and 1(g), which are close-ups of Fig. 1(e). Points on the end plates [white dots in Figs. 1(d) to 1(g)] are the vertices of the triangles that make the surface meshes. They naturally sample the surfaces uniformly and densely. The number of individual segments is of the order of a few thousands. For each point on one end plate, its corresponding counterpart on the opposite end plate is found as the vertex most aligned along the direction of $\vec{N}$, the normal to the two end plates (average of the two normals). Let U be a vertex of the upper plate. Its counterpart L on the lower end plate is defined as the vertex that maximizes the scalar product:

$$S = \vec{UL} \cdot \vec{N}. \quad (8)$$

All vectors being normalized to a unit length, $\vec{UL}$ and $\vec{N}$ are perfectly aligned when S is equal to 1. Any deviation results in a lower scalar product. To avoid line segments not aligned with the normal, a match on the lower end plate was accepted only if S was not lower than 0.99.

II.B.2. Intraoperator and interoperator variability

Although the method is mostly automated, it requires an operator to input seeds for the vertebral body segmentation. We investigated measurement variability associated with this manual input. One operator (S.T.) provided two sets of seeds (with a lapse of one month) for the intraoperator reliability study. Another operator (M.W.) provided one set for the interoperator reliability study. Both operators performed the task for both scans of each patient. We restricted this analysis to the high resolution reconstructions.

The seed placement does not require extensive training. The second operator (M.W.) only performed the task twice. He was trained once on all patients. His second set of seeds was used in the study. The algorithm does not require an accurate location of the vertebral body's center.

II.B.3. Local heights

For the evaluation of vertebral fractures, it is desirable to estimate vertebral height locally at anterior, middle, and posterior positions. This allows the identification of different types of fracture: wedge, biconcave, and crush.¹² We can also use our algorithm to compute heights locally. Figure 3 shows the three areas (anterior, middle, and posterior) within which we can restrict height computation. These local areas are defined in the following manner. We first determine the middle, anterior, and posterior points (Fig. 3). The middle point is computed as the average of all end plate points. The anterior and posterior points are the end plate points along the y-axis

FIG. 3. Anterior, middle, and posterior areas for local height measure. The larger points represent the anterior, middle, and posterior points.

(sagittal direction) furthest from the central point. The anterior, middle, and posterior areas are then defined as the sets of points closest to the anterior, middle, and posterior points, respectively. We evaluated the precision of the algorithm in local areas limited to 15% of the end plate surface. We also estimated the precision of the anterior-posterior ratio (AP) and the middle-posterior ratio (MP).¹² The same areas can be used to compute local disk heights, which are useful for characterizing disk geometry. Disk convexity can be assessed using the ratio of middle disk height to the mean of the anterior and posterior disk heights. Convex and concave disks are associated with ratios larger and smaller than 1, respectively.²

We also measured local heights manually on the mid-sagittal slice of the high resolution CT scans using the Carestream PACS software (Carestream, Rochester, NY) as shown in Fig. 4. We compared the precision of this manual method to that of the automated algorithm.

II.B.4. Robustness to Gaussian noise

We tested the algorithm's robustness to the addition of Gaussian noise to simulate scanning at lower current levels. Using the inverse proportional relation between noise variance and current level, we added the specific amount of Gaussian noise to simulate current levels of 200, 100, and 50 mAs. Noise variance was estimated as the mean of the five squared standard deviations in the grey levels of the five tubes of the QCT phantom. Noise variance was first computed in the original image with the actual current level of 300 mAs. To simulate current levels of 200, 100, and 50 mAs, the final noise

FIG. 4. Manual measurement of local vertebral heights (left) and local disk heights (right) on the mid-sagittal plane.

variances after added Gaussian noise had to be, respectively, 3/2, 3, and 6 times the original noise variance.

The precision of the algorithm was estimated for each of the three levels of noise. We also investigated how the added Gaussian noise affected the measured heights. We computed the mean difference between vertebral and disk heights from the original image and the same heights when noise is added. This study was restricted to the medium resolution reconstructions. We did not use the high resolution reconstructions because, with its more limited field of view, the QCT phantom was not always visible.

II.B.5. Accuracy

In the absence of a gold standard, we compared the algorithm's measures with manual measures on the mid-sagittal slice of the eight high resolution reconstructions of the first scan. This is restricted to local vertebral and disk height measures. While in Sec. II.B.3, manual local measures were performed by joining corners (anterior and posterior heights) and midpoints of end plates (middle height), here, the operator traced the local heights along a subjective estimation of the normals to the end plates in conformity with the algorithm. It is important to note that the cortical bone has a thickness. For vertebral heights, the operator drew the lines from mid cortical bone of the superior end plate to mid cortical bone of the inferior end plate of the same vertebral body. However, for disk heights, the lines started just outside the cortical bone.

In addition, while in Sec. II.B.3 the algorithm measured local heights in small neighborhoods, here, the computed heights were restricted to the middle, anterior, and posterior points (Fig. 3) in conformity with the manual method.

II.B.6. Statistical analysis

For the study of the variability between scans, we computed the differences in height measurements between the first and second scans. For the intraoperator and interoperator variability study, the differences were for the same scan but with different seeds. We took the absolute value of the differences. The mean (standard deviation) of the differences and root-mean-square coefficients of variation (CV) constitute our measure of precision. We used Bland-Altman analysis to detect systematic error trends and evaluate the 95% agreement limits. Intraclass correlation (ICC) was also evaluated. To test if precision varied by scanner and image resolution, we performed an analysis of variance on the absolute value of the differences, with adjustment for patient and vertebra or disk level, using SAS software (version 9.2, Cary, N.C.).

For the accuracy study, we computed the correlation between manual and automated measures as well as the ICCs. We also computed the mean (standard deviation) difference between manual and automated measures, which was performed on signed rather than absolute value in order to detect potential systematic biases. Analysis of variance was also performed to compare manual and automated heights, with adjustment for patient and vertebral or disk level.

TABLE I. Precision of vertebral height measurements.

Vertebral height measurements						
	n	1st scan mean (mm)	Differences in absolute value mean $\pm$ SD (mm)	ICC	95% limits of agreement (mm)	rmsCV (%)
All	120	26.3	0.058 $\pm$ 0.052	0.99	[−0.14, 0.16]	0.21
64-row	75	26.0	0.044 $\pm$ 0.034	0.99	[−0.10, 0.12]	0.15
8-row	45	26.8	0.081 $\pm$ 0.067	0.99	[−0.19, 0.22]	0.28
High	40	26.3	0.057 $\pm$ 0.052	0.99	[−0.12, 0.17]	0.21
Medium	40	26.3	0.057 $\pm$ 0.053	0.99	[−0.15, 0.16]	0.21
Low	40	26.3	0.061 $\pm$ 0.051	0.99	[−0.15, 0.16]	0.21

III. RESULTS

When all measures are considered, the overall mean differences (standard deviations) were small, 0.058 (0.052) and 0.056 (0.049) mm for vertebral and disk height (Tables I and II, respectively). For vertebral heights, the number of samples is $n = 8$ (patients) $\times 5$ (vertebrae) = 40 for each image resolution, $n = 3$ (resolutions) $\times 5$ (patients) $\times 5$ (vertebrae) = 75 for the 64-detector row scanner, $n = 3$ (resolutions) $\times 3$ (patients) $\times 5$ (vertebrae) = 45 for the 8-detector row scanner, and $n = 3$ (resolutions) $\times 8$ (patients) $\times 5$ (vertebrae) = 120 for all measurements. For disk heights, the number of samples is $n = 8$ (patients) $\times 4$ (IDSs) = 32 for each image resolution, $n = 3$ (resolutions) $\times 5$ (patients) $\times 4$ (IDSs) = 60 for the 64-detector row scanner, $n = 3$ (resolutions) $\times 3$ (patients) $\times 4$ (IDSs) = 36 for the 8-detector row scanner, and $n = 3$ (resolutions) $\times 8$ (patients) $\times 4$ (IDSs) = 96 for all measurements. The vertebral and disk height measures for the first and second scans (high resolution reconstructions) can be seen in graphical form in the supplementary material.⁴⁶

The corresponding CVs are 0.21% and 0.94%. For all sets of measurements, the ICC was 0.99. Figures 5 and 6 show the Bland-Altman plots for all vertebral and disk heights, stratified by scanner (64- or 8-detector row) and by resolutions (high, medium, and low). The 95% confidence intervals are narrow and the differences do not seem to vary with the mean heights.

III.A. 64-detector row versus 8-detector row

To investigate if more advanced technology improved precision, we stratified the difference measures by scanner (64- and 8-detector row). The mean differences (standard deviations)

obtained for the older 8-detector row scanner are nearly twice as large as those of the 64-detector row scanner. Analysis of variance, with all resolutions pooled together, showed that the difference in precision between scanners was statistically significant for disk heights ($p = 0.015$) but not for vertebral heights ($p = 0.074$).

III.B. Influence of image resolution

To investigate the influence of the three image resolutions we stratified the results into the corresponding three groups. Results from the two scanners were pooled together. From Tables I and II (last three rows), for vertebral heights, high, medium, and low resolutions have similar precision. For disk heights, high resolution was slightly more precise than medium resolution or low resolution. However, analysis of variance revealed no statistically significant difference in precision between resolutions for either disk ($p = 0.32$) or vertebral ($p = 0.90$) heights.

III.C. Intraoperator and interoperator precision

Both intraoperator and interoperator measures had better precision than interscan measures with maximal CVs of 0.30% and 0.12% for disk and vertebral heights, respectively (Table III). For vertebral heights, the number of samples is $n = 2$ (scans) $\times 8$ (patients) $\times 5$ (vertebrae) = 80. For disk heights, the number of samples is $n = 2$ (scans) $\times 8$ (patients) $\times 4$ (IDSs) = 64. Intraoperator precision is better than interoperator precision for vertebral heights but they are virtually identical for disk heights (CV of 0.29% and 0.30%). ICC was very

TABLE II. Precision of disk height measurements.

Disk height measurements						
	n	1st scan mean (mm)	Differences in absolute value mean $\pm$ SD (mm)	ICC	95% limits of agreement (mm)	rmsCV (%)
All	96	6.82	0.056 $\pm$ 0.049	0.99	[−0.14, 0.15]	0.94
64-row	60	6.69	0.043 $\pm$ 0.031	0.99	[−0.085, 0.11]	0.62
8-row	36	7.04	0.078 $\pm$ 0.063	0.99	[−0.20, 0.19]	1.3
High	32	6.82	0.046 $\pm$ 0.039	0.99	[−0.12, 0.12]	0.78
Medium	32	6.82	0.062 $\pm$ 0.050	0.99	[−0.14, 0.17]	1.0
Low	32	6.81	0.060 $\pm$ 0.054	0.99	[−0.15, 0.16]	1.0

FIG. 5. Bland-Altman plots for vertebral heights, stratified by scanner and resolution.

high for both intraoperator and interoperator measures and for both disk and vertebral heights.

III.D. Local heights

Tables IV and V present results for local vertebral and disk heights, respectively. For vertebral heights, the number of samples is $n = 3$ (resolutions) $\times 8$ (patients) $\times 5$ (vertebrae) = 120. For disk heights, the number of samples is $n = 3$ (resolutions) $\times 8$ (patients) $\times 4$ (IDSs) = 96. Measurements for middle vertebral height have the best precision (CV of 0.33%), followed by posterior (0.40%), and anterior (0.44%). For disks, local middle heights also exhibit the best precision (1.27%). The 95% limits of agreement indicate that the algorithm can detect meaningful local height variations as small as one third of a millimeter.

III.E. Precision of manual measurements versus precision of automated measurements for the local heights on the mid-sagittal plane

Tables VI and VII compare precision results for local height measurements and their ratios using a mid-sagittal

plane manual method and the algorithm. For vertebral heights, the number of samples is $n = 8$ (patients) $\times 5$ (vertebrae) = 40. For disk heights, the number of samples is $n = 8$ (patients) $\times 4$ (IDSs) = 32. The 95% limits of agreement indicate that the manual method can detect meaningful local height variations of the order of a millimeter. A comparison of the CVs obtained by the algorithm and the manual technique suggests that the algorithm is 3 to 6 times and 2 to 3 times more precise for local vertebral and disk heights, respectively.

III.F. Robustness to Gaussian noise

Three levels of Gaussian noise, simulating current levels of 200, 100, and 50 mAs, were added to the medium resolution reconstructions of all patients (Fig. 7). Vertebral and disk heights were measured and the CVs between first and second scan were computed. Precision remained high for all noise levels. The CVs increased only slightly from 0.213% (300 mAs) to 0.265% (50 mAs) for vertebral heights and from 1.01% (300 mAs) to 1.18% (50 mAs) for disk heights (Fig. 8).

FIG. 6. Bland-Altman plots for disk heights, stratified by scanner and resolution.

However, the addition of Gaussian noise caused the measured heights to deviate from the original measurements. The mean difference (in absolute value) between vertebral and disk heights from the original image and the same heights when noise is added increased with de-

creasing simulated current levels. Those differences remained small and increased only from 0.024 (200 mAs) to 0.10 mm (50 mAs) for vertebral heights and from 0.030 (200 mAs) to 0.11 mm (50 mAs) for disk heights (Fig. 9).

TABLE III. Intraoperator and interoperator variability and agreement.

		n	Differences in absolute value mean +/- SD (mm)	ICC	95% limits of agreement (mm)	rmsCV (%)
Intra	Vertebra	80	0.017 +/- 0.023	0.99	[-0.058, 0.054]	0.077
	Disk	64	0.019 +/- 0.019	0.99	[-0.052, 0.053]	0.29
Inter	Vertebra	80	0.024 +/- 0.038	0.99	[-0.084, 0.094]	0.12
	Disk	64	0.018 +/- 0.021	0.99	[-0.052, 0.055]	0.30

TABLE IV. Precision of local vertebral height measurements.

Local vertebral height measurements					
		Differences in absolute value			rmsCV (%)
	n	mean +/- SD	ICC	95% limits of agreement	
Anterior	120	0.12 +/- 0.11 mm	0.99	[-0.33, 0.31] mm	0.44
Middle	120	0.083 +/- 0.077 mm	0.99	[-0.21, 0.23] mm	0.33
Posterior	120	0.12 +/- 0.092 mm	0.99	[-0.26, 0.33] mm	0.40
AP ratio	120	0.0063 +/- 0.0056	0.98	[-0.018, 0.014]	0.62
MP ratio	120	0.0050 +/- 0.0042	0.97	[-0.013, 0.012]	0.49

TABLE V. Precision of local disk height measurements.

Local disk height measurements					
		Differences in absolute value			rmsCV
	n	mean +/- SD	ICC	95% limits of agreement	(%)
Anterior	96	0.11 +/- 0.085 mm	0.99	[-0.20, 0.30] mm	1.59
Middle	96	0.092 +/- 0.071 mm	0.99	[-0.20, 0.25] mm	1.27
Posterior	96	0.13 +/- 0.10 mm	0.99	[-0.33, 0.32] mm	2.34
Convexity ratio	96	0.020 +/- 0.020	0.98	[-0.055, 0.056]	1.69

TABLE VI. Comparison between the precision of the vertebral height measurements obtained with the computer algorithm and a manual mid-sagittal method.

		Local vertebral height measurements					
		n	Differences in absolute value		95% limits of agreement	rmsCV (%)	
			mean	+/- SD			ICC
Anterior	Manual	40	0.56	+/- 0.35 mm	0.97	[-1.31, 1.26] mm	1.94
	Algorithm	40	0.12	+/- 0.11 mm	0.99	[-0.33, 0.29] mm	0.43
Middle	Manual	40	0.50	+/- 0.37 mm	0.96	[-1.19, 1.24] mm	1.82
	Algorithm	40	0.078	+/- 0.078 mm	0.99	[-0.17, 0.24] mm	0.32
Posterior	Manual	40	0.39	+/- 0.31 mm	0.98	[-1.10, 0.89] mm	1.31
	Algorithm	40	0.13	+/- 0.091 mm	0.99	[-0.25, 0.34] mm	0.41
AP ratio	Manual	40	0.023	+/- 0.017	0.94	[-0.055, 0.057]	2.32
	Algorithm	40	0.0065	+/- 0.0052	0.98	[-0.018, 0.013]	0.62
MP ratio	Manual	40	0.019	+/- 0.016	0.86	[-0.046, 0.052]	2.05
	Algorithm	40	0.0050	+/- 0.0041	0.97	[-0.013, 0.012]	0.49

TABLE VII. Comparison between the precision of the disk height measurements obtained with the computer algorithm and a manual mid-sagittal method.

		Local disk height measurements				
		n	Differences in absolute value			rmsCV (%)
			mean +/- SD	ICC	95% limits of agreement	
Anterior	Manual	32	0.46 +/- 0.43 mm	0.98	[-0.75, 1.38] mm	4.96
	Algorithm	32	0.13 +/- 0.077 mm	0.99	[-0.23, 0.32] mm	1.49
Middle	Manual	32	0.29 +/- 0.24 mm	0.99	[-0.57, 0.81] mm	3.19
	Algorithm	32	0.1 +/- 0.077 mm	0.99	[-0.24, 0.26] mm	1.39
Posterior	Manual	32	0.41 +/- 0.32 mm	0.92	[-0.80, 1.14] mm	7.47
	Algorithm	32	0.17 +/- 0.13 mm	0.99	[-0.39, 0.44] mm	2.28
Convexity ratio	Manual	32	0.078 +/- 0.049	0.95	[-0.20, 0.15]	4.90
	Algorithm	32	0.019 +/- 0.023	0.98	[-0.058, 0.059]	1.80

FIG. 7. Gaussian noise added to (a) the original image to simulate scanning with lower current levels of (b) 200, (c) 100, and (d) 50 mAs. The current level in the original image is 300 mAs.

III.G. Accuracy

Manual and computed local measures exhibited very good correlations ranging from 0.87 to 0.98 (Fig. 10). Agreement was also very good with ICCs ranging from 0.87 to 0.98. For vertebral heights, the analysis of variance detected a significant bias. Computed measures were about 1 mm larger than

manual measures (Table VIII). Disk heights exhibited a mean adjusted difference between -0.07 and 0.35 mm and no bias (Table IX). For vertebral heights, the number of samples is $n = 8$ (patients) $\times$ 5 (vertebrae) = 40. For disk heights, the number of samples is $n = 8$ (patients) $\times$ 4 (IDSs) = 32.

III.H. Computing time

We evaluated the mean durations of the three parts of the algorithm: 33 s for the segmentation of one vertebral body, 27 s for the simultaneous segmentation of two end plates, and 5.3 s for the computation of the distance of between two end plates. In total, it takes about 1 min to compute the height of one vertebra. The computer used was a desktop with a processor speed of 3.16 GHz. The computer code has not been optimized for speed.

IV. DISCUSSION

We have presented what, to the best of our knowledge, is the first precision study of disk and vertebral height measurement in CT involving short term repeat scans in patients. The algorithm has excellent precision and can detect height changes as small as one tenth of a millimeter. A comparison

FIG. 8. Precision of the algorithm in the presence of added Gaussian noise simulating lower current levels. The current level in the original image is 300 mAs.

FIG. 9. Mean differences between height measurements with no added noise and with added Gaussian noise simulating lower current levels. Error bars represent the standard deviations. The current level in the original image is 300 mAs.

with manual mid-sagittal plane measurements shows that the algorithm can be as much as 6 times more precise. We investigated the influence of several factors including scanner technology (64- versus 8-detector row) and in-plane image resolution. The 64-detector row scanner was found to perform significantly better than the 8-detector row scanner for disk heights. No significant difference was found between different resolutions for either disk or vertebral heights. The method has high intraoperator and interoperator precision with ICC of 0.99 for both and CVs ranging from 0.077% to 0.30%.

Our study included eight patients. A one day double CT scan entails a considerable amount of radiation. It is very difficult to find volunteers willing to submit to such an exposure for a study that advances knowledge but offers no clinical benefit to them. However, we computed both vertebral and disk heights. Adding disk and vertebral heights, we evaluated nine measures per subject. Because our algorithm can identify vertebral end plates, measuring disk or vertebral height is computationally identical. Our algorithm was therefore tested on 72 height samples. Additional work with a larger cohort would be useful to further establish the method's precision.

All patients had ankylosing spondylitis and had CT scans as part of a study to measure syndesmophytes (new bone formation at the vertebral rim). The presence of syndesmophytes did not substantially compromise the precision of our technique. Similar paravertebral ossifications commonly occur in degenerative spondylosis, and could compromise the precision of conventional methods of vertebral or disc height measurement based on radiography or even CT.

We showed that the algorithm was robust to Gaussian noise. This robustness is likely due to the Gaussian smoothing that constitutes the preprocessing step that prepares the segmentation of the vertebral bodies. However, the added noise caused a deviation from the initial height measurements with no supplementary noise. This suggests that in a longitudinal study it would be best to keep scanning and reconstruction parameters constant to ensure similar levels of noise. If that is not practicable, it is possible that a smoothness equalization procedure⁴¹ could reduce the discrepancy in measurements introduced by differing levels of noise.

The accuracy study revealed a significant difference between manual and computed local vertebral heights. On average, computed local vertebral heights were about 1 mm larger

FIG. 10. Correlation between manual measures and computed measures for local vertebral and disk heights.

TABLE VIII. Difference between manual and automated local mid-sagittal vertebral height measurements.

		Vertebral height			
	n	Difference (mean +/- SD)	Adjusted difference	95% CI (mm)	p
Anterior	40	-0.58 +/- 0.74 mm	-0.573	-1.178, 0.022	0.06
Middle	40	-0.95 +/- 0.62 mm	-0.948	-1.496, -0.40	0.005
Posterior	40	-0.95 +/- 0.86 mm	-0.955	-1.48, -0.429	0.004

TABLE IX. Difference between manual and automated local mid-sagittal disk height measurements.

	n	Disk height			
		Difference (mean $\pm$ SD)	Adjusted difference	95% CI (mm)	<i>p</i>
Anterior	32	-0.072 $\pm$ 0.65 mm	-0.07	-0.78, 0.64	0.65
Middle	32	0.36 $\pm$ 0.68 mm	0.35	-0.26, 0.97	0.22
Posterior	32	0.012 $\pm$ 0.59 mm	0.01	-0.35, 0.38	0.94

than manual measures. For the manual measures, the operator drew lines from mid cortical bone to mid cortical bone. On the other hand, the computer algorithm segmented the entire thickness of the cortical bone. Computed vertebral heights were therefore from outer cortical bone boundary to the outer cortical bone boundary of the opposite end plate. The difference in the amount of cortical bone included in the two types of measures could account for the bias detected by the analysis of variance. This hypothesis is further supported by the fact that no significant difference was detected between manual and computed disk heights. In this case the operator drew the lines from outer cortical bone boundary to the outer cortical bone boundary of the adjacent end plate, which is consistent with the computer algorithm. This study highlights the problems facing manual measurement. The choice of measurement end points can be less than straightforward. The perception of the cortical bone thickness itself can be dependent upon the magnification factor chosen by the operator when visualizing the vertebra.

Because vertebral and disk heights have traditionally been measured using other modalities (radiography, MRI, and DXA), it is useful to compare our results to those previously reported for other modalities. We limit the comparison to investigations that performed *in vivo* short term repeat scans. To the best of our knowledge the most recent comparable results are those of Frobin *et al.*⁴² for radiography, Boos *et al.*¹⁷ for MRI, and Crabtree *et al.*⁴³ for DXA. Measurements were performed manually in all three studies.

Important differences in materials and methods make meaningful comparisons difficult. Since none of those studies investigated the effects of different image resolutions, we did not average results from our three reconstructions here but only used the results of the high resolution images. Each group of researchers used a different methodology to evaluate the precision of their method. For comparison purposes in this discussion, we adopted their precision metrics.

Boos *et al.* (MRI) measured only disk heights using the standard deviation of the signed differences between interscan measurements as the measure of precision. Applying this measure to our data, our CT method yields a standard deviation of 0.061 mm for disk height, compared with 0.36 mm for MRI (see supplementary material).⁴⁶ This represents an improvement of about a factor of 6. Frobin *et al.* (radiography) measured both vertebral and disk heights and used CVs as their precision metric. Our CT method yields CVs of 0.21% and 0.78% for vertebral and disk heights, respectively, compared to 2.2% and 4.15% for radiography (see supplementary material).⁴⁶ Crabtree *et al.* (DXA) measured local vertebral

heights and their ratios only. They evaluated variability using the root mean square of errors and CV. They reported four sets of results, one for each of three sites plus an additional reading for a center that used two settings on the DXA system. We list the range (minimum and maximum) of those measures (see supplementary material).⁴⁶ Comparing the CVs, our CT method is about 4 to 6 times more precise than the best DXA results and about 6 to 8 times more precise than the worst DXA results.

An attractive feature of our algorithm is that it reduces interoperator variability to levels lower than interscan variability. Boos *et al.* (MRI) reported that inter-reader variability could be as high as interscan variability. Crabtree *et al.* (DXA) reported that inter-reader variability surpassed interscan variability for one DXA system (Hologic QDR 4500A).

The superior precision of CT over radiography and DXA can be explained by the fact that CT captures the complete 3D structure of objects while radiography and DXA only provide a 2D projection. The additional information makes CT height estimates more robust. Our results also show that the CT algorithm has higher precision than MRI measurements. CT images usually have higher resolution than MRI and offer better visualization of bone. The superior interoperator repeatability of our CT algorithm could be due to automation.

However, the trade-off associated with CT's superior precision and interoperator repeatability is greater radiation exposure. The four modalities have different strengths and weaknesses and should be selected according to the application's specific needs. CT would be most suitable when the highest precision is essential.

Routine clinical screening for vertebral fractures typically involves imaging both the entire thoracic and lumbar spine. Although CT could also be used to scan both the entire lumbar and thoracic spine, this would entail a larger dose of radiation. For this reason, our method may not be suited to routine practice to detect fractures. For the moment, it might be best regarded as a research instrument, for instance, for pilot studies of new treatments where sensitivity to change is very helpful because it would require fewer subjects and shorter trial duration.

Currently, for longitudinal studies, there are many definitions of incident fracture based on radiographs. The most authoritative seems to be Black *et al.*'s, who define an incident fracture as height reduction of at least 20%–25%.⁴⁴ This is sometimes combined with the requirement for an absolute height reduction of at least 3–4 mm.

Although Black *et al.* do not explicitly mention the precision of vertebral height measurement in radiographs, it is

implicitly taken into account as the rate of false positives in the determination of the 20%–25% threshold. Black *et al.* arrived at this threshold by minimizing the number of participants required to significantly differentiate two groups, the placebo group and the group receiving medication, in a hypothetical clinical trial. Using a definition of incident fractures based on a reduction in vertebral height of less than 20% required more participants because the imprecision of x ray measurements made the two groups harder to differentiate (less distinction at this threshold between true positive fractures and false positive fractures).

With our highly precise CT method, a definition of vertebral fracture as height reduction of 10% or even 5% may be possible. Assuming a mean vertebral height of 25 mm, 5% represents 1.25 mm, which is well above the 95% limits of agreement of our least precise measure [anterior height (−0.33, 0.31) mm]. The number of participants in a clinical trial could be kept low and between-group differences could be detected sooner, resulting in shorter trials.

Like CT, MRI also can also provide 3D visualization of the spine. However, CT usually has better resolution and very high contrast between cortical bone and soft tissue. Image acquisition time is much longer for MRI and patient movement can result in motion artifact on the obtained image. These factors have made the automatic segmentation of vertebrae by a computer algorithm more feasible in CT than in MRI. CT also has higher availability and lower cost. However, MRI is a more versatile diagnostic tool, with excellent soft tissue contrast. The scope of our method is more limited. It focuses on studies where vertebral height and disk height are the main outcomes.

Currently, disk height is often estimated using MRI. Local heights (anterior, middle, and posterior) are measured in the mid-sagittal slice. The Dabbs method (adapted from radiography to the mid-sagittal slice in MRI) takes the mean of the anterior and posterior heights.⁴⁵ Because cortical bone typically does not show on MRI, due to low water content, those disk measurements have to be performed between the trabecular bone of adjacent vertebral bodies. This approach therefore likely overestimates disk height. For the design of disk implants/prostheses, a more accurate measure is desirable and CT is preferred. Our method performs the measurements in 3D space rather than just at the mid-sagittal slice. This is made possible by automation as slice by slice manual measurement would be prohibitively expensive in terms of manpower. For research purposes, our method is more precise than the method proposed by Boos *et al.*¹⁷ for MRI and would be able to detect smaller changes, for instance, in studies of interventions to slow or reverse disk narrowing, or to identify risk factors for disk degeneration more accurately.

Our algorithm provides a powerful submillimeter precision instrument for research applications. Although a submillimeter vertebral or disk height reduction may not be clinically meaningful, our method provides a tool for investigating associations that could not be resolved using a less precise measurement technique. For instance, we could investigate if a vertebral height reduction of 1 mm is a predictor of a verte-

bral fracture as it is now defined (vertebral height reduction of at least 20% and 4 mm). More precise measures allow better/more accurate identification of clinical correlates of disc or vertebral height loss. For example, if we were to predict disk height based on age, sex, height, occupation, and smoking, with crude measures, only the strongest predictors would be found to be significantly associated; with more precise measures, more predictors would likely be found to be significantly associated, and the estimates of strength of association would be more accurate.

V. CONCLUSION

We have presented a novel computer assisted method for estimating vertebral and disk heights using CT scans. We have evaluated the method's precision using unique data, same day double CT scans of patients. The method exhibited very high precision and can detect height changes as small as one tenth of a millimeter. Longitudinal studies particularly benefit from higher precision as smaller samples and shorter study durations are required for detecting statistically significant changes. This method may therefore be of particular benefit for studies evaluating treatments to maintain or normalize vertebral or disk heights.

ACKNOWLEDGMENT

This research was supported by the Intramural Research Program of NIAMS and CC (NIH).

^{a)}Author to whom correspondence should be addressed. Electronic mail: tanso@mail.nih.gov

¹M. T. Modic and J. S. Ross, "Lumbar degenerative disk disease," *Radiology* **245**, 43–61 (2007).

²C. W. Pfirrmann, A. Metzendorf, A. Elfering, J. Hodler, and N. Boos, "Effect of aging and degeneration on disc volume and shape: A quantitative study in asymptomatic volunteers," *J. Orthop. Res.* **24**, 1086–1094 (2006).

³C. J. Siepe, W. Hitzl, P. Meschede, A. K. Sharma, M. F. Khattab, and M. H. Mayer, "Interdependence between disc space height, range of motion and clinical outcome in total lumbar disc replacement," *Spine* **34**, 904–916 (2009).

⁴T. B. Neuschwander, J. Cutrone, B. R. Macias, S. Cutrone, G. Murthy, H. Chambers, and A. R. Hargens, "The effect of backpacks on the lumbar spine in children: a standing magnetic resonance imaging study," *Spine* **35**, 83–88 (2010).

⁵T. Zander, P. Krishnakanth, G. Bergmann, and A. Rohlmann, "Diurnal variations in intervertebral disc height affect spine flexibility, intradiscal pressure and contact compressive forces in the facet joints," *Comput. Methods Biomech. Biomed. Eng.* **13**, 551–557 (2010).

⁶S. H. Zhou, I. D. McCarthy, A. H. McGregor, R. R. Coombs, and S. P. Hughes, "Geometrical dimensions of the lower lumbar vertebrae—Analysis of data from digitised CT images," *Eur. Spine J.* **9**, 242–248 (2000).

⁷E. B. van der Houwen, P. Baron, A. G. Veldhuizen, J. G. Burgerhof, P. M. van Ooijen, and G. J. Verkerke, "Geometry of the intervertebral volume and vertebral endplates of the human spine," *Ann. Biomed. Eng.* **38**, 33–40 (2010).

⁸T. L. Schulte, F. Leistra, V. Bullmann, N. Osada, V. Vieth, B. Marquardt, T. Lerner, U. Liljenqvist, and L. Hackenberg, "Disc height reduction in adjacent segments and clinical outcome 10 years after lumbar 360 degrees fusion," *Eur. Spine J.* **16**, 2152–2158 (2007).

⁹M. J. McGirt, S. Eustacchio, P. Varga, M. Vilendecic, M. Trummer, M. Gorensek, D. Ledic, and E. J. Carragee, "A prospective cohort study

- of close interval computed tomography and magnetic resonance imaging after primary lumbar discectomy: Factors associated with recurrent disc herniation and disc height loss," *Spine* **34**, 2044–2051 (2009).
- ¹⁰A. Hiwataishi, T. Moritani, Y. Numaguchi, and P. L. Westesson, "Increase in vertebral body height after vertebroplasty," *AJNR Am. J. Neuroradiol.* **24**, 185–189 (2003).
 - ¹¹G. Sun, P. Jin, M. Li, X. W. Liu, and F. D. Li, "Height restoration and wedge angle correction effects of percutaneous vertebroplasty: Association with intraosseous clefts," *Eur. Radiol.* **21**, 2597–2603 (2011).
 - ¹²H. K. Genant, C. Y. Wu, C. van Kuijk, and M. C. Nevitt, "Vertebral fracture assessment using a semiquantitative technique," *J. Bone Miner. Res.* **8**, 1137–1148 (1993).
 - ¹³J. S. Bauer, D. Müller, A. Ambekar, M. Dobritz, M. Matsuura, F. Eckstein, E. J. Rummeny, and T. M. Link, "Detection of osteoporotic vertebral fractures using multidetector CT," *Osteoporosis Int.* **17**, 608–615 (2006).
 - ¹⁴G. Guglielmi, D. Diacinti, C. van Kuijk, F. Aparisi, C. Krestan, J. E. Adams, and T. M. Link, "Vertebral morphometry: Current methods and recent advances," *Eur. Radiol.* **18**, 1484–1496 (2008).
 - ¹⁵F. Grados, J. Fechtenbaum, E. Flipon, S. Kolta, C. Roux, and P. Fardellone, "Radiographic methods for evaluating osteoporotic vertebral fractures," *Jt., Bone Spine* **76**, 241–247 (2009).
 - ¹⁶J. A. Kanis, E. V. McCloskey, T. Powles, A. H. Paterson, S. Ashley, and T. Spector, "A high incidence of vertebral fracture in women with breast cancer," *Br. J. Cancer* **79**, 1179–1181 (1999).
 - ¹⁷N. Boos, A. Wallin, M. Aebi, and C. Boesch, "A new magnetic resonance imaging analysis method for the measurement of disc height variations," *Spine* **21**, 563–570 (1996).
 - ¹⁸N. Roberts, C. Gratin, and G. H. Whitehouse, "MRI analysis of lumbar intervertebral disc height in young and older populations," *J. Magn. Reson. Imaging* **7**, 880–886 (1997).
 - ¹⁹Z. Shao, G. Rompe, and M. Schiltenswolf, "Radiographic changes in the lumbar intervertebral discs and lumbar vertebrae with age," *Spine* **27**, 263–268 (2002).
 - ²⁰J. Reeve, M. Lunt, D. Felsenberg, A. J. Silman, C. Scheidt-Nave, G. Poor, C. Gennari, K. Weber, R. Lorenc, P. Masaryk, J. B. Cannata, J. Dequeker, D. M. Reid, H. A. Pols, L. I. Benevolenskaya, J. J. Stepan, T. Miazgowski, A. Bhalla, J. B. Armas, R. Eastell, A. Lopes-Vaz, G. Lyrilis, I. Jajic, A. D. Woolf, D. Banzer, W. Reisinger, C. J. Todd, B. Felsch, S. Havelka, K. Hoszowski, J. Janott, O. Johnell, H. H. Raspe, O. B. Yershova, J. A. Kanis, G. Armbrrecht, J. D. Finn, W. Gowin, T. W. O'Neill, and E. P. O. S. Group, "Determinants of the size of incident vertebral deformities in European men and women in the sixth to ninth decades of age: The European Prospective Osteoporosis Study (EPOS)," *J. Bone Miner. Res.* **18**, 1664–1673 (2003).
 - ²¹M. G. Donaldson, L. Palermo, J. T. Schousboe, K. E. Ensrud, M. C. Hochberg, and S. R. Cummings, "FRAX and risk of vertebral fractures: The fracture intervention trial," *J. Bone Miner. Res.* **24**, 1793–1799 (2009).
 - ²²S. Kolta, S. Kerkeni, C. Travert, W. Skalli, R. Eastell, C. C. Glüer, and C. Roux, "Variations in vertebral body dimensions in women measured by 3D-XA: A longitudinal in vivo study," *Bone* **50**, 777–783 (2012).
 - ²³R. Lindsay, J. Nieves, C. Formica, E. Henneman, L. Woelfert, V. Shen, D. Dempster, and F. Cosman, "Randomised controlled study of effect of parathyroid hormone on vertebral-bone mass and fracture incidence among postmenopausal women on oestrogen with osteoporosis," *Lancet* **350**, 550–555 (1997).
 - ²⁴D. M. Black, P. D. Delmas, R. Eastell, I. R. Reid, S. Boonen, J. A. Cauley, F. Cosman, P. Lakatos, P. C. Leung, Z. Man, C. Mautalen, P. Mesenbrink, H. Hu, J. Caminis, K. Tong, T. Rosario-Jansen, J. Krasnow, T. F. Hue, D. Sellmeyer, E. F. Eriksen, S. R. Cummings, and H. P. F. Trial, "Once-yearly zoledronic acid for treatment of postmenopausal osteoporosis," *N. Engl. J. Med.* **356**, 1809–1822 (2007).
 - ²⁵J. E. Iglesias and M. de Bruijne, "Semiautomatic segmentation of vertebrae in lateral x-rays using a conditional shape model," *Acad. Radiol.* **14**, 1156–1165 (2007).
 - ²⁶G. Guglielmi, F. Palmieri, M. G. Placentino, F. D'Errico, and L. P. Stopino, "Assessment of osteoporotic vertebral fractures using specialized workflow software for 6-point morphometry," *Eur. J. Radiol.* **70**, 142–148 (2009).
 - ²⁷S. Kasai, F. Li, J. Shiraishi, Q. Li, and K. Doi, "Computerized detection of vertebral compression fractures on lateral chest radiographs: Preliminary results with a tool for early detection of osteoporosis," *Med. Phys.* **33**, 4664–4674 (2006).
 - ²⁸M. G. Roberts, E. M. Pacheco, R. Mohankumar, T. F. Cootes, and J. E. Adams, "Detection of vertebral fractures in DXA VFA images using statistical models of appearance and a semi-automatic segmentation," *Osteoporosis Int.* **21**, 2037–2046 (2010).
 - ²⁹S. Ghosh, R. S. Alomari, V. Chaudhary, and G. Dhillon, "Automatic lumbar vertebra segmentation from clinical CT for wedge compression fracture diagnosis," *Proc. SPIE Med. Imaging* **7963**, 796303 (2011).
 - ³⁰J. Yao, S. D. O'Connor, and R. M. Summers, "Automated spinal column extraction and partitioning," *Proceedings of the IEEE International Symposium on Biomedical Imaging* (Arlington, VA, 2006), pp. 390–393.
 - ³¹A. Mastmeyer, K. Engelke, C. Fuchs, and W. A. Kalender, "A hierarchical 3D segmentation method and the definition of vertebral body coordinate systems for QCT of the lumbar spine," *Med. Image Anal.* **10**, 560–577 (2006).
 - ³²M. Hardisty, L. Gordon, P. Agarwal, T. Skrinkas, and C. Whyne, "Quantitative characterization of metastatic disease in the spine. Part I. Semiautomated segmentation using atlas-based deformable registration and the level set method," *Med. Phys.* **34**, 3127–3134 (2007).
 - ³³M. S. Aslan, A. Ali, H. Rara, and A. A. Farag, "An automated vertebra identification and segmentation in CT images," in *Proceedings of the IEEE International Conference on Image Processing* (Hong Kong, 2010), pp. 233–236.
 - ³⁴J. A. Sethian, *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision and Materials Science* (Cambridge University Press, Cambridge, England, 1999).
 - ³⁵V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *Int. J. Comput. Vis.* **22**, 61–79 (1997).
 - ³⁶S. Tan, J. Yao, M. M. Ward, L. Yao, and R. M. Summers, "Computer aided evaluation of ankylosing spondylitis using high-resolution CT," *IEEE Trans. Med. Imaging* **27**, 1252–1267 (2008).
 - ³⁷L. Ibanez, W. Schroeder, L. Ng, and J. Cates, *The ITK Software Guide* (Kitware Inc., Clifton Park, NY, 2003).
 - ³⁸W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing* (Cambridge University Press, Cambridge, England, 2007).
 - ³⁹W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3D surface construction algorithm," *ACM SIGGRAPH Comput. Graphics* **21**, 163–169 (1987).
 - ⁴⁰J. J. Koenderink, *Solid Shape* (MIT, Cambridge, MA, 1990).
 - ⁴¹S. Tan, J. Yao, L. Yao, and M. M. Ward, "Improved precision of syndesmo-phyte measurement for the evaluation of ankylosing spondylitis using CT: A phantom and patient study," *Phys. Med. Biol.* **57**, 4683–4704 (2012).
 - ⁴²W. Frobin, P. Brinckmann, M. Biggemann, M. Tillotson, and K. Burton, "Precision measurement of disc height, vertebral height and sagittal plane displacement from lateral radiographic views of the lumbar spine," *Clin. Biomech. (Bristol, Avon)* **12**(Suppl 1), S1–S63 (1997).
 - ⁴³N. Crabtree, J. Wright, A. Walgrove, J. Rea, L. Hanratty, M. Lunt, I. Fogelman, R. Palmer, M. Vickers, J. E. Compston, and J. Reev, "Vertebral morphometry: Repeat scan precision using the Lunar Expert-XL and the Hologic 4500A. A study for the 'WISDOM' RCT of hormone replacement therapy," *Osteoporosis Int.* **11**, 537–543 (2000).
 - ⁴⁴D. M. Black, L. Palermo, M. C. Nevitt, H. K. Genant, L. Christensen, and S. R. Cummings, "Defining incident vertebral deformity: A prospective comparison of several approaches. The Study of Osteoporotic Fractures Research Group," *J. Bone Miner. Res.* **14**, 90–101 (1999).
 - ⁴⁵V. M. Dabbs and L. G. Dabbs, "Correlation between disc height narrowing and low-back pain," *Spine* **15**, 1366–1369 (1990).
 - ⁴⁶See supplemental material at <http://dx.doi.org/10.1118/1.4769412> which graphically shows height measurements for the two repeat scans and compares the precision of CT height measurements with height measurements in MRI, radiography and DXA.