Title

Phantomless calibration of CT scans for measurement of BMD and bone strength—Inter-operator reanalysis precision

Authors

Lee, David C
Hoffmann, Paul F
Kopperdahl, David L
et al.

Publication Date

2017-10-01

DOI

10.1016/j.bone.2017.07.029

Peer reviewed

HHS Public Access

Author manuscript

Bone. Author manuscript; available in PMC 2018 October 01.

Published in final edited form as:

Bone. 2017 October ; 103: 325–333. doi:10.1016/j.bone.2017.07.029.

Phantomless Calibration of CT Scans for Measurement of BMD and Bone Strength — Inter-Operator Reanalysis Precision

David C. Lee, PhD[1], **Paul F. Hoffmann, BS**[1], **David L. Kopperdahl, PhD**[1], and **Tony M. Keaveny, PhD**[2,3]

[1]O.N. Diagnostics LLC, Berkeley, CA, USA

[2]Department of Mechanical Engineering, University of California, Berkeley, CA, USA

[3]Department of Bioengineering, University of California, Berkeley, CA, USA

Abstract

Patient-specific phantomless calibration of computed tomography (CT) scans has the potential to simplify and expand the use of pre-existing clinical CT for quantitative bone densitometry and bone strength analysis for diagnostic and monitoring purposes. In this study, we quantified the inter-operator reanalysis precision errors for a novel implementation of patient-specific phantomless calibration, using air and either aortic blood or hip adipose tissue as internal calibrating reference materials, and sought to confirm the equivalence between phantomless and (traditional) phantom-based measurements. CT scans of the spine and hip for 25 women and 15 men (mean ± SD age of 67 ± 9 years, range 41–86 years), one scan per anatomic site per patient, were analyzed independently by two analysts using the VirtuOst software (O.N. Diagnostics, Berkeley, CA). The scans were acquired at 120 kVp, with a slice thickness/increment of 3 mm or less, on nine different CT scanner models across 24 different scanners. The main parameters assessed were areal bone mineral density (BMD) at the hip (total hip and femoral neck), trabecular volumetric BMD at the spine, and vertebral and femoral strength by finite element analysis; other volumetric BMD measures were also assessed. We found that the reanalysis precision errors for all phantomless measurements were less than 0.5%, which was as good as for phantom calibration. Regression analysis indicated equivalence of the phantom- versus phantomless-calibrated measurements (slope not different than unity, $R^2 \geq 0.98$). Of the main parameters assessed, non-significant paired mean differences (n=40) between the two measurements ranged from 0.6% for hip areal BMD to 1.1% for mid-vertebral trabecular BMD. These results indicate that phantom-equivalent measurements of both BMD and finite element-derived bone strength can be reliably obtained from CT scans using patient-specific phantomless calibration.

Keywords

BMD; bone strength; precision; repeatability; osteoporosis; finite element analysis

Reprints and correspondence: David C. Lee, 2150 Shattuck Ave. STE 610, Berkeley, CA 94704.

1. INTRODUCTION

Quantitative analysis of computed tomography (CT) scans can be performed clinically to identify patients at high risk of fracture [1–3] based on calibrated measurements of bone mineral density (BMD) at the spine and hip, as well as measurements of bone strength when combined with finite element analysis [4–7]. In any quantitative CT analysis, proper calibration of the scan is required to correct for variations in scanner settings and attenuation characteristics [8], any related beam-hardening, and patient-specific characteristics such as body size, all of which can alter the attenuation characteristics [9]. Without such corrections, BMD and bone strength measurements can vary across different CT scanners or with different scan protocols or over time, confounding interpretation and clinical utility.

The most widely used method for calibrating CT scans utilizes an external calibration phantom [10]. However, the need for a phantom, which must be placed under the patient during scanning, adds expense and increases the logistical burden of clinical imaging. Various approaches have been proposed to calibrate without an external calibration phantom. One approach is to not calibrate the scan [11], but this amounts to not performing quantitative densitometry and is therefore questionable for diagnostic or monitoring purposes because attenuation values can vary widely depending on the specific scanner and scan protocol [9, 12]. Another approach is to pre-calibrate a particular CT scanner using a calibration phantom, or to pre-calibrate CT-based BMD measurements via DXA, and then use that pre-calibration for scans of future patients on that or similar scanners [13–16]. While this approach is preferable to not calibrating, both approaches are not specific to the individual patient and thus ignore potentially important calibration issues associated with variations in patient body size and habitus; nor can they be applied retrospectively over any appreciable period of time. A third approach, which represents a patient-specific phantomless calibration, is to utilize the patient's own internal tissues as the calibrating reference materials [17, 18]. One such approach involves sampling the attenuation of a region of muscle, then further processing those attenuation data into components assumed to be associated with pure muscle tissue and pure adipose tissue, thereby providing attenuation data for two known reference materials [19]. However, one limitation of that approach is its poor repeatability [17], presumably due to the challenges of choosing the region of muscle in a repeatable fashion as well as consistently separating out the pure tissue components.

Overcoming these limitations with patient-specific phantomless calibration, we report here on an alternative implementation having improved precision. Already validated clinically against DXA for accurately measuring areal BMD at the hip [2, 3], this technique utilizes as calibrating reference materials the external air that is visible on the patient's CT scan and one of either the patient's blood or adipose tissue adjacent to the bone being assessed. Applying this technique to a diverse cohort of patients and CT scanners, we sought to quantify its inter-operator reanalysis precision for measuring both BMD and finite element-derived bone strength, at both the spine and hip; we also sought to confirm the equivalence of the phantomless versus traditional phantom-based measurements.

2. METHODS

2.1 CT Scans for the Study Sample

The study comprised of a reanalysis of pre-existing research-quality clinical-resolution CT scans that had been analyzed in prior clinical drug trials at O.N. Diagnostics. Permission was obtained from the original sources for reanalysis of the CT scans, and additional internal review board approval for this reanalysis was not necessary due to the retrospective, de-identified nature of the dataset.

We randomly selected a sample of CT scans from a larger pool of scans in order to minimize sources of measurement bias in the selected scans. Eight prior multi-center studies at O.N. Diagnostics had scans that were available for reanalysis. Of those studies, eligible CT scans for inclusion were those that were: 1) acquired on a multi-detector CT scanner at 120 kVp; 2) available for both the spine and hip for an individual patient; 3) imaged with a mineral phantom pair that included an external phantom for calibration and a quality-assurance phantom for beam hardening correction; 4) reconstructed utilizing a slice thickness of 3 mm or less and a standard kernel (GE: standard; Siemens: B30; Philips: B, C; Toshiba: FC12, FC13); 5) without imaging artifacts that would preclude analysis; and 6) were not previously utilized in the development of the phantomless calibration method. Typical exposure values were set according to patient height and weight (ranging from 25–195 mAs for spine scans, and 50–390 mAs for hip scans) or were determined by the scanner's automatic exposure control function (e.g. noise index = 25 HU or quality reference mAs = 160). For any of the studies that assessed longitudinal effects, only baseline scans were included, so no scans reflected any treatment effects. The CT scans for 1032 subjects thus identified were acquired on 58 scanners. From those, we randomly selected 40 subjects — providing sufficient statistical power for a precision study [20, 21] — while limiting the number of scans from any single CT scanner to no more than eight. scans were derived from 24 unique CT scanners, representing nine different CT scanner models (Table 1). Fourteen of these 24 scanners were represented in development of the phantomless calibration method. The cohort consisted of 25 women and 15 men, spanning a wide range of age, weight, height, and body mass index (Table 2).

2.2 Phantom and Phantomless Calibrations

The VirtuOst software (version 2.1, O.N. Diagnostics, Berkeley, CA), written in the Python programming language and utilizing NumPy and SciPy software libraries, was used for both the phantom and phantomless calibrated analyses. Both methods of calibration were performed separately at the spine and hip since attenuation characteristics can differ at each site due to site-specific differences in body habitus. For the phantom calibration, the same type of external calibration phantom (Model 3 Phantom, Mindways Software, Inc., Austin, TX) was utilized for each patient (Figure 1). Following the manufacturer-supplied specifications [22], the attenuation values (Hounsfield Units) of the chambers in the calibration phantom were sampled and the images were calibrated into equivalent-BMD units (mg/cm^3) of a K_2HPO_4-water mixture. The user specified the range along the length of the phantom from which to sample data (avoiding any shading or other artifacts) and then over that range cylinders were automatically registered to each chamber. The final volume of

interest (VOI) was determined by removing an outer layer from each chamber to eliminate volume averaging with the surrounding phantom substrate. To help account for potential beam hardening between the locations of the external calibration phantom and the bone of interest, a quality-assurance torso phantom (Mindways Software, Inc., Austin, TX) was also scanned on top of the external calibration phantom (Figure 1), typically within one day of each subject's CT exam. The resulting quality-assurance phantom scan was calibrated as described above and the measured BMD of its central chamber was used to generate a ratio versus the reference value of BMD for that chamber [23]; this field-uniformity correction ratio was then used to scale the corresponding subject's BMD measurement to correct for any field-uniformity effects, the same ratio used both for the spine and hip scans. Across the 40 scans, values of this ratio varied from 0.92 –1.05. To assess the sensitivity of the overall reanalysis precision error to the measurement of this ratio (which itself contributes some degree of measurement error), we also calculated reanalysis precision with this ratio set to 1.0 for all scans.

For the phantomless calibration, the attenuation values were sampled for external air and either abdominal aortic blood tissue for assessment of the spine or pelvic visceral adipose tissue from the ischioanal fossa for assessment of the hip (Figure 1). Since both reference tissues are near the respective bones of interest, we applied no additional corrections for beam-hardening effects. Segmentation of the aorta was initiated by a single seed point, and then proceeded automatically to identify a curved cylindrical VOI, consisting of variably sized and centered circles on each transverse section, spanning a distance from approximately one vertebral height above and below the center of the target vertebra. Segmentation was performed using gradient-profile algorithms that were independent of absolute attenuation, and the size and location of the circles at each transverse section were chosen to omit regions of calcification and minimize overall heterogeneity in attenuation. Segmentation of the ischioanal fossa, which was fully automated, identified a contiguous volume of attenuation values within a pre-specified range and situated in the appropriate location between the femurs (Figure 1). The resulting thresholded volume was then smoothed to remove rough edges. After the initial segmentation of both tissues, the final VOI was determined by removing an outer layer from each segmented VOI in order to minimize volume averaging with any adjacent tissues. Numerical quality checks were run to test if the final VOI was sufficiently large, the sampled tissue attenuation data were normally distributed, and the average attenuation value was within the expected range. If any check failed, the analyst was alerted and could make manual changes.

The resulting scan-specific attenuation measurements for the external air and the reference tissues were then paired against previously developed reference values of equivalent-BMD (in units of a K_2HPO_4-water mixture) for these materials to provide a scan-specific linear calibration equation. Although the same reference values of equivalent-BMD were used for all scans for a given CT manufacturer, the calibration was specific to each scan since it depended on the scan-specific measured attenuation values of the reference materials. The reference values of equivalent-BMD depended on the CT manufacturer and the kVp setting of the scan, and were developed empirically in a prior calibration process. For that process, we used data from 268 different patients scanned on 31 different CT scanners, spanning four different CT manufacturers (GE Healthcare, Siemens, Philips, Toshiba). The scans were all

acquired at 120 kVp with a slice thickness of 3 mm or less and reconstructed with a standard kernel; exposure values were set according to patient height and weight or were determined by the scanner's automatic exposure control function; table heights were determined by centering the table on the QA or torso phantom. Using these scans, an optimization study was then performed to identify reference values, by CT manufacturer, for the air, blood, and adipose tissues that minimized the difference in measurements of both cortical and trabecular BMD between phantom and phantomless calibrations for the 268 patients. After optimizing, the coefficient of variation (defined here as the standard deviation of the difference in measurements, divided by mean of the underlying measurements) in integral volumetric BMD between phantom and phantomless calibration across all patients ranged across the four manufacturers from 2.9–4.3% (7–10 mg/cm^3) at the hip and from 2.8–4.1% (5–7 mg/cm^3) at spine. Separately, extrapolation of these reference values for 120 kVp to other kVp values was performed using scans of a custom torso phantom consisting of various tissue-equivalent chambers and scanned on 35 different CT scanners (for the four manufacturers) at various settings (80–140 kVp), although performance of the phantomless calibration at different kVp values was not assessed in this study.

To convert units of K_2HPO_4-equivalent areal BMD into units of DXA-equivalent (Hologic) areal BMD, a previously-developed empirical equation (the same for both phantom and phantomless calibration) was used. The resulting type of areal BMD measurements has been validated against DXA measurements [2, 3, 24].

2.3 Bone Strength and BMD Measurements

VirtuOst was also used to measure volumetric BMD and finite element-derived bone strength at the spine and hip, and areal BMD at the hip. For the phantom and phantomless analyses, all image processing and finite element analyses were identical except for the calibration method. For the spine, we measured the mid-vertebral trabecular volumetric BMD and vertebral compressive strength for the L1 vertebra, or if L1 was unavailable, one other level from T12–L3 (Figure 2). Spine areal BMD was not assessed since prior studies have shown that both mid-vertebral trabecular volumetric BMD and vertebral strength are more strongly associated with risk of vertebral fracture than spine areal BMD [25–27]. For the hip, we measured DXA-equivalent areal BMD values for the femoral neck and total hip regions and overall femoral strength for a simulated sideways fall, preferentially from the left proximal femur (Figure 2). All these "clinical" measurements have been validated in various prospective [1, 25, 28] and prevalent [26, 27, 29, 30] fracture-outcome studies, and are available clinically for patient care [2, 3, 31] and meet recommended practice guidelines [32]. To enable further comparisons with the research literature, we also report on the reanalysis precision for various other volumetric BMD measurements, specifically, integral volumetric BMD for both the whole vertebral body and total hip and their respective cortical and trabecular compartments. The *cortical compartment* included all cortical bone and some adjacent endosteal trabecular bone, and was defined as all bone within a fixed distance of the periosteal surface (3 mm for the hip; 2 mm for the spine) plus any other adjacent bone with an apparent-BMD value greater than 1.0 g/cm^3; the *trabecular compartment* was defined as all of the remaining trabecular bone. Henceforth, for brevity, we refer to these compartments as the cortical and trabecular bone, respectively.

All additional image processing for strength and BMD assessments was performed as described elsewhere [1–3]. Briefly, the target bones in the QCT images were segmented (and posterior elements removed for the spine), calibrated, registered into a standardized coordinate system, and then converted into finite element models. The highly automated segmentation techniques for the bone employed a combination of seeded region-growing, various standard morphological algorithms (such as erosion, dilation, and smoothing), active contouring, and calibration-independent adaptive thresholding. After registering the segmented bone into a standardized coordinate system, all regions of interest for the BMD measurements were automatically placed based on the location of specific bone features such as the vertebral endplates, the femoral neck, and the lesser trochanter. The location and orientation of these regions could be adjusted manually if needed. Following precedence [33], the mid-vertebral trabecular volumetric BMD (in mg/cm^3) was sampled within an elliptical region of interest, 8 mm high, located in the middle of the vertebral body in a slightly anterior position (Figure 2). For the hip areal BMD measurements, the segmented femur was automatically registered into the recommended DXA orientation, projected in the anterior-posterior direction into a frontal plane, and then processed to generate DXA-like femoral neck and total hip regions of interest (Figure 2).

After all image processing, the images were converted into finite element models using 1.0 mm (spine) or 1.5 mm (hip) cube-shaped, eight-noded brick elements, both re-sampled resolutions producing in the range of 40,000 elements per finite element model. Element-specific elastic properties (anisotropic for the spine; isotropic for the hip) and elastic-plastic failure properties (higher strength in compression than tension at the hip) were all derived from the calibrated volumetric BMD values [34–36]. After automatically registering each segmented bone to a standardized orientation via rigid-body transformation, boundary conditions were applied to simulate a uniform compressive overload of the spine or a sideways fall of the hip, both via a layer of plastic to mimic laboratory testing conditions, and non-linear finite element analysis was performed to estimate strength (in newtons) at each site.

2.4 Inter-Operator Reanalysis Precision

Two analysts performed all measurements, blinded to each other, for both the phantom and phantomless calibrations, for both the spine and hip, for all 40 subjects. This enabled an inter-observer comparison between the two analysts. For the phantom calibration, this repeated analysis also included re-measurement of the field-uniformity correction ratio from the quality-assurance phantom scans.

2.5 Statistical Analyses

A number of statistical procedures were employed. To assess inter-operator reanalysis precision error for each BMD and strength measurement, both for phantom and phantomless calibrations, we calculated the root-mean-square error [21] between the two analysts' measurements (n=40 pairs, separately for spine and hip), expressed both in absolute units of the measurement (standard deviation, SD_{RMS}) and as a percentage of the mean measurement (coefficient of variation, CV_{RMS}). Any statistically significant difference between the phantom and phantomless precision errors was determined by 95% confidence intervals of

the respective variances using a chi-square distribution [21]. In assessing equivalence between the absolute measurements provided by phantomless and phantom calibration, we tested for any proportional or fixed biases [37] using orthogonal linear regression analysis [38] and Bland-Altman analysis [39], respectively. From the Bland-Altman plots, we calculated the 95% limits of agreement between phantomless versus phantom calibrations; that parameter was calculated as 1.96 times the standard deviation of the paired differences between the phantomless and phantom measurements, centered about any statistically significant fixed bias. We also performed paired t-tests on the phantomless and phantom measurements. All statistics were performed using JMP (Version 9.0, SAS Institute, Inc., Cary, NC).

3. RESULTS

The inter-operator reanalysis precision error for phantomless calibration was no more than 0.5% for all measurements at both the spine and hip, and was at least as low as for phantom calibration (Table 3). The precision error for phantomless calibration ranged from 0.1% for total hip areal BMD and total hip trabecular volumetric BMD to 0.5% for vertebral strength and mid-vertebral trabecular volumetric BMD. Numerically, these precision errors were always either equal to or lower than those for phantom calibration. For all the hip measurements the precision errors were significantly lower ($p<0.05$) for phantomless calibration than for phantom calibration, reflecting that the phantomless calibration was fully automated for the hip, whereas some user input was required for our implementation of the traditional phantom-based calibration. When measurement of the field-uniformity correction ratio was eliminated (i.e. set equal to unity for all analyses) for the traditional phantom-based calibration, precision errors for the phantom calibration decreased by up to 0.2 percentage points (Table 3).

Comparison of the absolute measurements between the phantomless and phantom calibrations confirmed the equivalence of both measurements. For all parameters assessed, with the exception of cortical volumetric BMD, there were no statistically significant (paired) differences between any of the phantom and phantomless measurements, and orthogonal regression showed no proportional bias (slope not different than one) (Table 4, Figure 3); for cortical volumetric BMD, the observed statistical difference was small (1%). Bland-Altman analysis also revealed no trend between the difference in any of the measurements between phantom and phantomless calibration and their average value (Figure 3). For the clinical measurements, the 95% limits of agreement were ±7 mg/cm^3 for mid-vertebral trabecular volumetric BMD, ±550 N for vertebral strength, ±0.037 g/cm^2 for femoral neck areal BMD, ±0.044 g/cm^2 for total hip areal BMD, and ±310 N for femoral strength. Consistent with this level of agreement, all measurements were highly correlated between the two calibration methods ($R^2 \geq 0.98$ for the clinical measures, $R^2 \geq 0.95$ for the other vBMD measures, Table 4).

4. DISCUSSION

External calibration phantoms have been the mainstay of quantitative-CT bone densitometry since its inception, but they complicate the logistics of scan acquisition. The results of this

study — obtained across a wide range of CT scanners and patients — demonstrate that patient-specific phantomless calibration can reliably provide measurements of both BMD and bone strength that are equivalent to those provided by traditional phantom calibration. That is to say, for analysis of a single scan, the measurements from phantomless calibration are as precise as those from traditional phantom-based calibration, and are equivalent statistically in terms of absolute values. Despite concerns over extrapolation errors if using calibration reference materials that have attenuation values far beyond the range of bone [18], the current results demonstrate that the combination of just external air and either blood tissue (at the spine) or adipose tissue (at the hip) can be used to calibrate CT scans equivalently to traditional phantom-based calibration, the gold standard for quantitative CT. Consistent with these findings, we previously found that hip areal BMD measurements, obtained by applying the currently reported method of phantomless calibration to routine clinical CT colonography and enterography scans, were equivalent to those as measured directly by DXA [2, 3]. The current results complement those findings by extending them to using traditional phantom-based quantitative CT as the reference standard.

There are advantages and trade-offs to consider when using phantomless calibration in any quantitative CT analysis compared to using a traditional phantom calibration. One important advantage of this implementation of phantomless calibration is the use of reference tissues immediately adjacent to the target bone, which should not only capture the beam-hardening effects in those reference tissues but also any similar effects in the adjacent bone. By contrast, an external phantom alone cannot reliably measure those effects because it is placed outside the body [9]. External phantoms can also be susceptible to subtle artifacts arising from any air gaps between the patient and the external phantom. These artifacts can influence the measured attenuation from the phantom chambers, and in turn can diminish inter-operator precision if different analysts correct for such artifacts in slightly different ways; these artifacts can also diminish longitudinal precision if they vary over time. In the same way, artifacts internal to the body (arising from air in the lungs or in the bowels, aortic calcifications, disease in adipose tissue, etc.) can affect the phantomless calibration tissues, and thus in both methods an analyst should always review scans for such subtle artifacts and make manual adjustments if necessary. Clearly, the main advantage of phantom calibration is the physical uniformity of the phantom chambers, in space and over time. In developing our phantomless calibration technique, we have found that aortic blood at the spine and visceral adipose tissue at the hip appear relatively homogeneous across patients and do not vary much across body size and habitus, which explains in part why the use of these tissues as reference materials provided good agreement against the phantom-based values. With ease of use and precision in mind, the other reference material was the external air, which our current results demonstrate can be used for calibration even though air is already used to calibrate Hounsfield Unit values of attenuation. It remains for future work to demonstrate that these blood and adipose tissues are sufficiently stable over time to provide highly precise measurements of *change* in BMD and strength. In the meanwhile, our current results indicate that any inhomogeneity of these tissues across patients has little practical effect on the ability of our phantomless calibration to provide one-time measurements that are equivalent to those from traditional phantom calibration.

In performing any precision study, one key issue is the potential trade-off between automation — which improves reanalysis precision — and accuracy. By comparing the absolute values between the two calibration methods, we were able to confirm that the automated elements of our process did not degrade accuracy. We note also that the lower precision for phantomless calibration at the hip compared to phantom calibration does not imply that precision for phantomess calibration is better than for phantom-based calibration. Since the hip adipose tissue is relatively homogenous, it can be processed in a highly automated fashion, with minimal or no user input. Meanwhile, for our implementation of the phantom-based calibration, the analyst must choose the range over which the phantom is sampled, and that range can vary across analyses. If our implementation of phantom calibration had been completely automated, then we would expect no difference in reanalysis precision between the two methods.

Beyond the method of calibration, highly automated image processing algorithms are also integral to achieving good precision [40]. In this regard, VirtuOst employs a unique combination of segmentation and registration techniques that typically require little user input beyond a small number of seed points, and then readjustment only if necessary. In addition, the BMD regions of interest are initially placed automatically based on anatomic features, and can also be adjusted if necessary. These types of contemporary methods have been shown to improve precision compared to earlier generation image processing methods [40]. As regards our hip areal BMD measurements, they were derived from analysis of the segmented image in a standardized registered orientation that is the same for all femurs. Processing the segmented image for these measurements eliminates many of the soft-tissue artifacts that are inherent to DXA measurements [41]. In that sense, these CT-based areal BMD measurements are technically more robust than those provided by DXA. Unfortunately, we could not measure short-term precision since we did not have pairs of independent scans for each patient, although others have shown excellent short-term precision for CT-based measurements of hip areal BMD using traditional phantom calibration [42].

Regarding the finite element analysis-derived measurements, two important factors that can influence their precision, in addition to segmentation and calibration, are the orientation of the bone and application of the boundary conditions. To minimize such errors, VirtuOst uses an optimization scheme to rigidly transform the vertebral body and proximal femur to a standardized orientation for virtual loading, uniformly truncates the model boundaries using automatically detected anatomic landmarks, and resamples all CT images to uniform resolutions for the creation of the same-sized voxel finite elements (1.0 mm for the spine, 1.5 mm for the hip) regardless of the scan resolution. These types of techniques should improve not only inter-operator but also short-term precision. Beyond another (brief) inter-operator reanalysis precision study utilizing an earlier version of the VirtuOst software as applied to 39 patients who had a CT colonography exam [3], we are aware of only three reports of repeatability precision for finite element-derived results, all of which were short-term precision [5, 9, 43] (Table 5). In two studies, patients were scanned twice on the same CT machine, with repositioning between. The first of these studies [5] utilized large elements and a relatively simple implementation of the finite element analysis technique as applied to the vertebral body and had poor short-term precision (12.1%); by contrast, the

second of these studies [43] utilized a more contemporary technique as applied to the proximal femur and achieved good short-term precision (1.9%), consistent with our findings. In the third study [9], patients were scanned on two different CT scanners on the same day, and the resulting inter-scanner (short-term) precision error for finite element strength ranged from 5.9 to 19.6%, depending on how the cross-calibration was performed and the type of loading condition (stance or fall). The poor inter-scanner precision, which also occurred for the various types of BMD measurements in that study, underscores the potential challenges of cross-calibrating across different CT machines, an issue that was not addressed in this current study. In that third study, the authors also report reanalysis (inter-operator) precision errors for a stance strength (1.6%) and a fall strength (6.4%) from previously unpublished data; it is not clear why the reanalysis precision error was so high for the fall strength, although it may be related to the failure criterion used to estimate a strength value from a linear analysis. Although alternate loading conditions are possible with VirtuOst, such as a stance orientation of the proximal femur, we constrained the outcomes in this study to those that are used clinically. However, we would expect those alternate strength outcomes to have similar precision to what has been reported in the current study since the methods are identical with the exception of the bone orientation and boundary conditions, both of which are automated processes.

Our results have implications for both clinical practice and research studies. Clinically, the low inter-operator reanalysis precision errors, the equivalence of the phantom and phantomless measurements, and the previous validation against DXA [2, 3, 24], together imply that a patient-specific phantomless analysis of pre-existing routine CT scans can provide reliable one-time diagnostic-quality measurements. For clinical monitoring, since repeat CT scans might not be feasible due to cost or radiation-exposure issues, patients with an existing CT scan covering the spine or hip could instead be monitored by a follow-up DXA exam. One would expect the precision error of such a comparison to be equivalent to that for comparing BMD T-scores from two different DXA scanners. In research studies, if the magnitude of the short-term precision error is similar for strength and areal BMD — as it is for the reanalysis precision error — then one would expect a shorter monitoring time interval for strength since typical annual changes are larger for strength than for BMD [2, 3, 24]. Consistent with this concept, the greater statistical power of the femoral strength measurement for detecting treatment effects has been demonstrated [44]. Eliminating the need for a calibration phantom would also reduce the expense and logistical challenges of research studies. However, while we have demonstrated equivalence of cross-sectional measurements of phantomless and phantom calibration, further work is required to demonstrate that *longitudinal changes* are also equivalent.

This study has a number of limitations. Most importantly, we did not measure short-term precision error. Due to the constraints of minimizing radiation exposure in research studies, short-term precision studies — in which the subject is scanned twice the very same way but with repositioning between the scans — are rare for CT. Another limitation is that, although we sampled CT scans from nine different CT scanner models and from 24 different scanners, the scans all had similar acquisition and reconstruction settings, the most important being 120 kVp and a standard reconstruction kernel; and most scanners were from a single manufacturer (GE Healthcare). While it is unlikely that the inter-operator reanalysis

precision error would be sensitive to the manufacturer or the precise value of equivalent-BMD assigned to our reference materials, the absolute measurements are likely sensitive to the acquisition settings, and we did not investigate in this study how longitudinal or manufacturer-related differences in acquisition or reconstruction settings would alter the measurements, either for phantom or phantomless calibration. These issues remain topics of ongoing research [9]. Related, the scans in this study were acquired originally in clinical drug studies, albeit in many different imaging centers, and may therefore not reflect typical clinical practice. However, a prior precision study using VirtuOst performed on ordinary CT colonography scans, showed similar precision errors to those reported here (Table 5), suggesting that the current data are indicative of what would be expected clinically, at least for specific classes of CT scans. Further, the scans in this study, although acquired in various facilities, were processed by trained staff in a centralized laboratory (O.N. Diagnostics) and not by hospital personnel. Finally, it is possible that the internal reference tissues may be altered for certain medical conditions – for example abdominal visceral adipose tissue in patients after gastric bypass surgery [45] or in women with anorexia nervosa [46]. Such changes would likely influence our phantomless measurements, although the size of any possible effects is not currently known. In the meanwhile, the strong correlation observed in this study between phantom versus phantomless measurements, and between phantomless-derived hip BMD against DXA-derived hip BMD [2, 3], suggests that the blood and adipose tissues utilized in this study, when paired with the external air, are sufficiently uniform across individuals to provide a patient-specific phantomless calibration of diagnostic quality.

Acknowledgments

Analyses were performed by Kwang Lee and Paul Hoffmann.

Scan selections were analyzed with permission from: Eli Lilly and Company, Merck & Co. Inc., Pfizer Inc., and University of Pennsylvania.

This study was partially supported by NIH AR057616.

References

1. Kopperdahl DL, Aspelund T, Hoffmann PF, Sigurdsson S, Siggeirsdottir K, Harris TB, Gudnason V, Keaveny TM. Assessment of incident spine and hip fractures in women and men using finite element analysis of CT scans. J Bone Miner Res. 2014; 29(3):570–80. [PubMed: 23956027]
2. Weber NK, Fidler JL, Keaveny TM, Clarke BL, Khosla S, Fletcher JG, Lee DC, Pardi DS, Loftus EV Jr, Kane SV Jr, Barlow JM Jr, Murthy NS Jr, Becker BD Jr, Bruining DH Jr. Validation of a CT-derived method for osteoporosis screening in IBD patients undergoing contrast-enhanced CT enterography. Am J Gastroenterol. 2014; 109(3):401–8. [PubMed: 24445572]
3. Fidler JL, Murthy NS, Khosla S, Clarke BL, Bruining DH, Kopperdahl DL, Lee DC, Keaveny TM. Comprehensive assessment of osteoporosis and bone fragility with CT colonography. Radiology. 2016; 278(1):172–80. [PubMed: 26200602]
4. Keyak JH, Rossi SA, Jones KA, Skinner HB. Prediction of femoral fracture load using automated finite element modeling. J Biomech. 1998; 31(2):125–33. [PubMed: 9593205]
5. Faulkner KG, Cann CE, Hasegawa BH. Effect of bone distribution on vertebral strength: assessment with patient-specific nonlinear finite element analysis. Radiology. 1991; 179(3):669–74. [PubMed: 2027972]
6. Crawford RP, Cann CE, Keaveny TM. Finite element models predict in vitro vertebral body compressive strength better than quantitative computed tomography. Bone. 2003; 33(4):744–50. [PubMed: 14555280]

7. Homminga J, Weinans H, Gowin W, Felsenberg D, Huiskes R. Osteoporosis changes the amount of vertebral trabecular bone at risk of fracture but not the vertebral load distribution. Spine. 2001; 26(14):1555–61. [PubMed: 11462085]
8. Cann CE, Genant HK. Precise measurement of vertebral mineral content using computed tomography. J Comput Assist Tomogr. 1980; 4(4):493–500. [PubMed: 7391292]
9. Carpenter RD, Saeed I, Bonaretti S, Schreck C, Keyak JH, Streeper T, Harris TB, Lang TF. Inter-scanner differences in in vivo QCT measurements of the density and strength of the proximal femur remain after correction with anthropomorphic standardization phantoms. Med Eng Phys. 2014; 36(10):1225–32. [PubMed: 25001172]
10. Cann CE. Quantitative CT for determination of bone mineral density: a review. Radiology. 1988; 166:509–522. [PubMed: 3275985]
11. Pickhardt PJ, Pooler BD, Lauder T, del Rio AM, Bruce RJ, Binkley N. Opportunistic screening for osteoporosis using abdominal computed tomography scans obtained for other indications. Annals of internal medicine. 2013; 158(8):588–95. [PubMed: 23588747]
12. Garner HW, Paturzo MM, Gaudier G, Pickhardt PJ, Wessell DE. Variation in attenuation in L1 trabecular bone at different tube voltages: caution is warranted when screening for osteoporosis with the use of opportunistic CT. AJR Am J Roentgenol. 2017; 208(1):165–170. [PubMed: 27762605]
13. Budoff MJ, Malpeso JM, Zeb I, Gao YL, Li D, Choi TY, Dailing CA, Mao SS. Measurement of phantomless thoracic bone mineral density on coronary artery calcium CT scans acquired with various CT scanner models. Radiology. 2013; 267(3):830–6. [PubMed: 23440323]
14. Brown JK, Timm W, Bodeen G, Chason A, Perry M, Vernacchia F, DeJournett R. Asynchronously calibrated quantitative bone densitometry. J Clin Densitom. 2016
15. Pickhardt PJ, Bodeen G, Brett A, Brown JK, Binkley N. Comparison of femoral neck BMD evaluation obtained using Lunar DXA and QCT with asynchronous calibration from CT colonography. J Clin Densitom. 2015; 18(1):5–12. [PubMed: 24880495]
16. Summers RM, Baecher N, Yao J, Liu J, Pickhardt PJ, Choi JR, Hill S. Feasibility of simultaneous computed tomographic colonography and fully automated bone mineral densitometry in a single examination. J Comput Assist Tomogr. 2011; 35(2):212–6. [PubMed: 21412092]
17. Pickhardt PJ, Lee LJ, del Rio AM, Lauder T, Bruce RJ, Summers RM, Pooler BD, Binkley N. Simultaneous screening for osteoporosis at CT colonography: bone mineral density assessment using MDCT attenuation techniques compared with the DXA reference standard. J Bone Miner Res. 2011; 26(9):2194–203. [PubMed: 21590738]
18. Gudmundsdottir H, Jonsdottir B, Kristinsson S, Johannesson A, Goodenough D, Sigurdsson G. Vertebral bone density in Icelandic women using quantitative computed tomography without an external reference phantom. Osteoporos Int. 1993; 3(2):84–9. [PubMed: 8453195]
19. Goodenough, DJ., Stockham, C. Quantitative computed tomography system. Columbia Scientific Inc; USA: 1991.
20. Bonnick SL, Johnston CC Jr, Kleerekoper M, Lindsay R, Miller P, Sherwood L, Siris E. Importance of precision in bone density measurements. J Clin Densitom. 2001; 4(2):105–10. [PubMed: 11477303]
21. Gluer CC, Blake G, Lu Y, Blunt BA, Jergas M, Genant HK. Accurate assessment of precision errors: how to measure the reproducibility of bone densitometry techniques. Osteoporosis Int. 1995; 5(4):262–70.
22. Mindways. Bone mineral densitometry software CT calibration phantom. Mindways Software, Inc; San Francisco, CA: 2011.
23. Mindways. Model 3 QA phantom certificate of calibration. San Francisco, CA: 2005. p. 1
24. Keaveny TM, Kopperdahl DL, Melton LJ 3rd, Hoffmann PF, Amin S, Riggs BL, Khosla S. Age-dependence of femoral strength in white women and men. J Bone Miner Res. 2010; 25(5):994–1001. [PubMed: 19874201]
25. Wang X, Sanyal A, Cawthon PM, Palermo L, Jekir M, Christensen J, Ensrud KE, Cummings SR, Orwoll E, Black DM, Keaveny TM. Prediction of new clinical vertebral fractures in elderly men using finite element analysis of CT scans. J Bone Miner Res. 2012; 27(4):808–16. [PubMed: 22190331]

26. Melton LJ, Riggs BL, Keaveny TM, Achenbach SJ, Hoffmann PF, Camp JJ, Rouleau PA, Bouxsein ML, Amin S, Atkinson EJ, Robb RA, Khosla S. Structural determinants of vertebral fracture risk. J Bone Miner Res. 2007; 22(12):1885–1892. [PubMed: 17680721]
27. Melton LJ 3rd, Riggs BL, Keaveny TM, Achenbach SJ, Kopperdahl D, Camp JJ, Rouleau PA, Amin S, Atkinson EJ, Robb RA, Therneau TM, Khosla S. Relation of vertebral deformities to bone density, structure, and strength. J Bone Miner Res. 2010; 25(9):1922–30. [PubMed: 20533526]
28. Orwoll ES, Marshall LM, Nielson CM, Cummings SR, Lapidus J, Cauley JA, Ensrud K, Lane N, Hoffmann PR, Kopperdahl DL, Keaveny TM. Finite element analysis of the proximal femur and hip fracture risk in older men. J Bone Miner Res. 2009; 24(3):475–83. [PubMed: 19049327]
29. Amin S, Kopperdahl DL, Melton LJ 3rd, Achenbach SJ, Therneau TM, Riggs BL, Keaveny TM, Khosla S. Association of hip strength estimates by finite-element analysis with fractures in women and men. J Bone Miner Res. 2011; 26(7):1593–600. [PubMed: 21305605]
30. Anderson DE, Demissie S, Allaire BT, Bruno AG, Kopperdahl DL, Keaveny TM, Kiel DP, Bouxsein ML. The associations between QCT-based vertebral bone measurements and prevalent vertebral fractures depend on the spinal locations of both bone measurement and fracture. Osteoporos Int. 2014; 25(2):559–66. [PubMed: 23925651]
31. Burch S, Feldstein M, Hoffmann PF, Keaveny TM. Prevalence of bone quality in women undergoing spinal fusion using biomechanical-CT analysis. Spine (Phila Pa 1976). 2016; 41(3): 246–52. [PubMed: 26352741]
32. Zysset P, Qin L, Lang T, Khosla S, Leslie WD, Shepherd JA, Schousboe JT, Engelke K. Clinical use of quantitative computed tomography-based finite element analysis of the hip and spine in the management of osteoporosis in adults: the 2015 ISCD official positions-part II. J Clin Densitom. 2015; 18(3):359–92. [PubMed: 26277852]
33. Genant HK, Faulkner KG, Gluer CC, Engelke K. Bone densitometry: current assessment. Osteoporos Int. 1993; 3(Suppl 1):91–7. [PubMed: 8461589]
34. Morgan EF, Keaveny TM. Dependence of yield strain of human trabecular bone on anatomic site. J Biomech. 2001; 34(5):569–577. [PubMed: 11311697]
35. Morgan EF, Bayraktar HH, Keaveny TM. Trabecular bone modulus-density relationships depend on anatomic site. J Biomech. 2003; 36(7):897–904. [PubMed: 12757797]
36. Bayraktar HH, Morgan EF, Niebur GL, Morris GE, Wong EK, Keaveny TM. Comparison of the elastic and yield properties of human femoral trabecular and cortical bone tissue. J Biomech. 2004; 37(1):27–35. [PubMed: 14672565]
37. Ludbrook J. Statistical techniques for comparing measurers and methods of measurement: a critical review. Clin Exp Pharmacol Physiol. 2002; 29(7):527–36. [PubMed: 12060093]
38. Linnet K. Evaluation of regression procedures for methods comparison studies. Clinical chemistry. 1993; 39(3):424–32. [PubMed: 8448852]
39. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986; 1(8476):307–10. [PubMed: 2868172]
40. Engelke K, Mastmeyer A, Bousson V, Fuerst T, Laredo JD, Kalender WA. Reanalysis precision of 3D quantitative computed tomography (QCT) of the spine. Bone. 2009; 44(4):566–72. [PubMed: 19070691]
41. Bolotin HH. DXA in vivo BMD methodology: An erroneous and misleading research and clinical gauge of bone mineral status, bone fragility, and bone remodelling. Bone. 2007; 41(1):138–54. [PubMed: 17481978]
42. Khoo BC, Brown K, Cann C, Zhu K, Henzell S, Low V, Gustafsson S, Price RI, Prince RL. Comparison of QCT-derived and DXA-derived areal bone mineral density and T scores. Osteoporos Int. 2009; 20(9):1539–45. [PubMed: 19107384]
43. Cody DD, Hou FJ, Divine GW, Fyhrie DP. Short term in vivo precision of proximal femoral finite element modeling. Annals of Biomedical Engineering. 2000; 28(4):408–14. [PubMed: 10870897]
44. Keaveny TM. Biomechanical computed tomography-noninvasive bone strength analysis using clinical computed tomography scans. Ann N Y Acad Sci. 2010; 1192:57–65. [PubMed: 20392218]

45. Torriani M, Oliveira AL, Azevedo DC, Bredella MA, Yu EW. Effects of Roux-en-Y gastric bypass surgery on visceral and subcutaneous fat density by computed tomography. Obesity surgery. 2015; 25(2):381–5. [PubMed: 25381117]

46. Gill CM, Torriani M, Murphy R, Harris TB, Miller KK, Klibanski A, Bredella MA. Fat attenuation at CT in anorexia nervosa. Radiology. 2016; 279(1):151–7. [PubMed: 26509295]

47. Steiger P, Block JE, Steiger S, Heuck AF, Friedlander A, Ettinger B, Harris ST, Gluer CC, Genant HK. Spinal bone mineral density measured with quantitative CT: effect of region of interest, vertebral level, and technique. Radiology. 1990; 175:537–543. [PubMed: 2326479]

48. Lang TF, Li J, Harris ST, Genant HK. Assessment of vertebral bone mineral density using volumetric quantitative CT. J Comput Assist Tomogr. 1999; 23(1):130–7. [PubMed: 10050823]

HIGHLIGHTS

- Phantomless calibration of CT scans has excellent inter-operator precision.
- Precision for finite element-derived bone strength is comparable to precision for BMD.
- Accuracy of phantomless calibration was confirmed vs. phantom calibration.

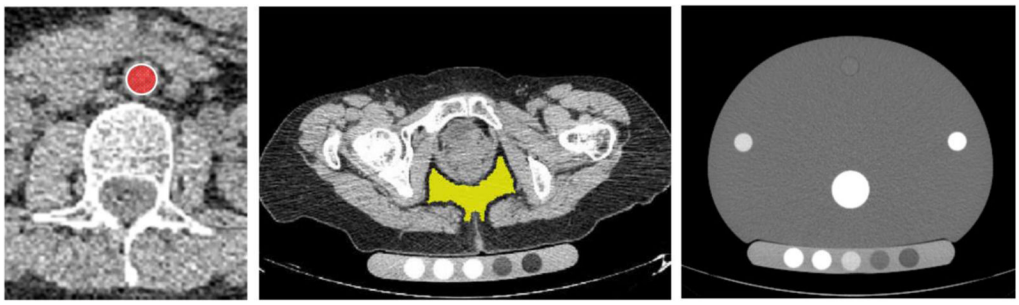

Figure 1.

Transverse images from a typical CT exam in this study for the spine (**left**), hip (**center**), and quality-assurance phantom (**right**), showing typical image quality and the internal reference tissues used for phantomless calibration – aortic blood tissue from the spine scan is shown in red (with white outline added); pelvic visceral adipose tissue (ischioanal fossa) from the hip scan is shown in yellow.

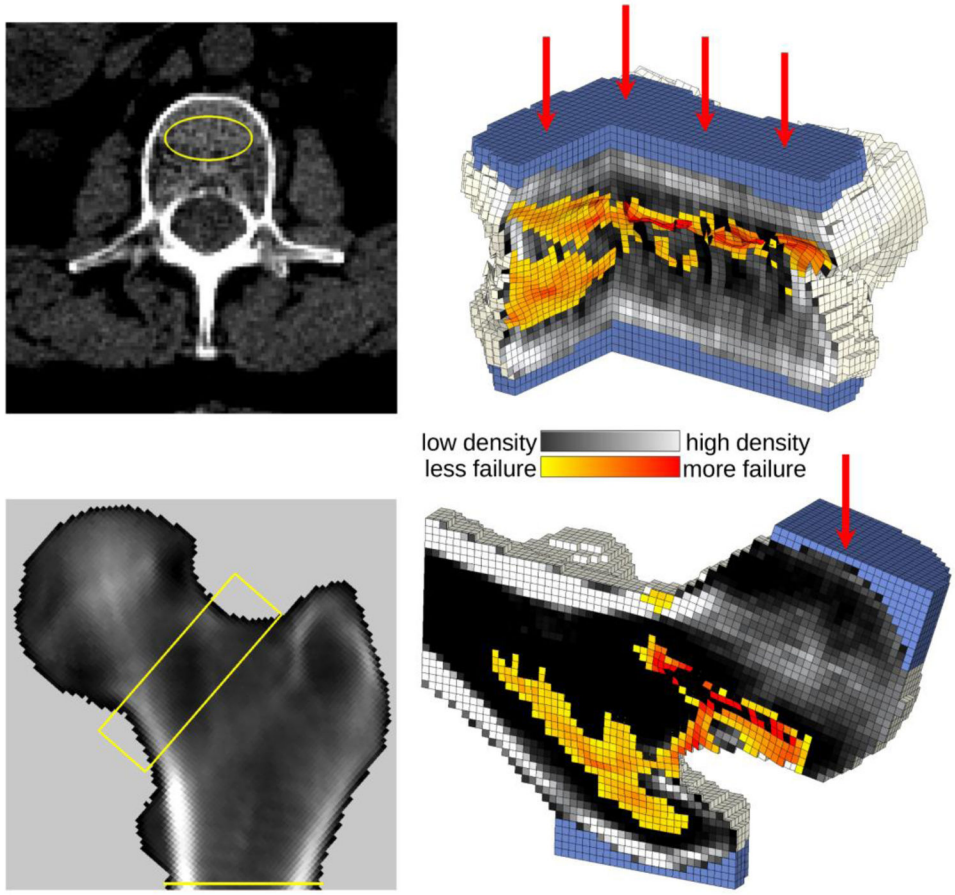

Figure 2.

Analysis of the spine and hip from a 70-year-old woman. **Left:** Automatically-placed regions of interest for the BMD analysis of the mid-vertebral trabecular bone (yellow ellipse) and DXA-equivalent areal BMD analysis of the hip, showing the femoral neck (yellow box) and total hip regions of interest (all bone above the yellow line, up to and including the femoral neck). **Right:** Virtual deformation patterns (magnified for viewing purposes) by finite element analysis for the spine and hip, showing regions of simulated bone tissue failure (colored). The gray colors denote the volumetric BMD values throughout the models. Boundary conditions were applied via virtual layers of plastic (light blue) to evenly distribute the applied loads over the bone surfaces, and simulated a compressive overload of the vertebral body and a sideways fall for the proximal femur.

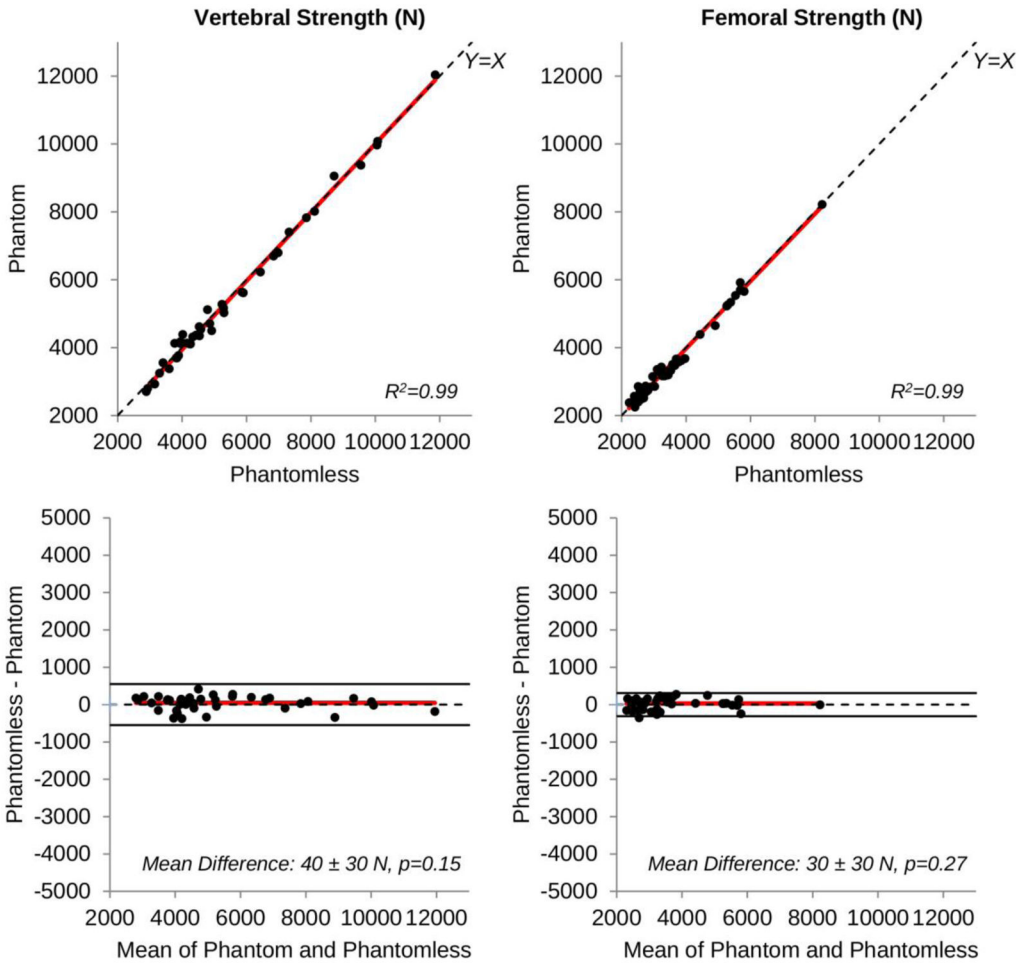

Figure 3.

Orthogonal regression analyses (**top**) and the corresponding Bland-Altman plots (**bottom**) comparing phantom- and phantomless-calibrated vertebral strength (**left**) and femoral strength (**right**). N=40 points in each plot; the same scales are used to enhance direct visual comparisons. For the regression plots: red line = orthogonal regression best fit; for reference, the *Y=X* line of unity is shown. For the Bland-Altman plots: red line = mean difference; dashed black line = zero difference; solid black lines = 95% limits of agreement; mean difference ± standard deviation, p-value versus zero mean.

Table 1

Forty scans each for the spine and hip were used in this analysis, acquired from nine different CT scanner models, from 24 different scanners.

CT Scanner Model*	Number of Scanners	Number of Scans†
GE BrightSpeed	3	4
GE LightSpeed Ultra	1	1
GE LightSpeed VCT	4	8
GE LightSpeed 16	5	8
Philips Brilliance 16	2	2
Philips Brilliance 64	3	7
Siemens Sensation 40	1	1
Siemens Sensation 64	1	2
Toshiba Aquilion	4	7
Total	*24*	*40*

*All scans were acquired at 120 kVp and reconstructed with a standard reconstruction kernel with a slice thickness of up to 3 mm.

†Number of paired spine and hip scans (n=40 total for each type of scan).

Table 2

Descriptive statistics for the analyzed cohort.

	Women	Men	Pooled
Number of Subjects	25	15	40
Age (yrs)	64 ± 9 (41–80)	72 ± 6 (65–86)	67 ± 9 (41–86)
Height (cm)	157 ± 6 (141–166)	177 ± 6 (167–185)	164 ± 12 (141–185)
Weight (kg)	58 ± 9 (39–73)	100 ± 18 (66–124)	74 ± 24 (39–124)
BMI (kg/m^2)	23.7 ± 3.9 (18.2–31.5)	31.7 ± 4.0 (23.3–36.7)	26.7 ± 5.5 (18.2–36.7)

BMI — Body Mass Index

Values are mean ± standard deviation (range in parentheses).

Table 3

Inter-operator reanalysis precision error (root-mean-square) for repeated analysis of scans at the spine and hip for n=40 subjects, for phantomless and phantom calibrations. Phantom calibration was performed with and without a field-uniformity correction (FUC) to assess the effect of measuring the FUC ratio on precision error. Results are expressed both as a coefficient of variation (CV_{RMS}, in %) and as a standard deviation (SD_{RMS}, in absolute units of the measured parameter).

	CV_{RMS} (%)			SD_{RMS} (absolute)		
	Phantomless	Phantom		Phantomless	Phantom	
		w/ FUC	w/o FUC		w/ FUC	w/o FUC
Clinical Measures:						
Vertebral Strength (N)	0.5	0.6	0.6	30	30	30
Mid-Vertebral Trabecular vBMD (mg/cm^3)	0.5	0.7	0.6	1	1	1
Femoral Strength (N)	0.4	0.8*	0.6	20	30	20
Femoral Neck aBMD (g/cm^2)	0.3	0.5*	0.4	0.002	0.003	0.002
Total Hip aBMD (g/cm^2)	0.1	0.4*	0.3*	0.001	0.004	0.002
Other vBMD Measures (mg/cm^3):						
Spine Integral	0.3	0.5*	0.3	<1	1	1
Spine Cortical	0.3	0.6*	0.4	1	1	1
Spine Trabecular	0.4	0.6	0.4	<1	1	1
Hip Integral	0.2	0.5*	0.3*	<1	1	1
Hip Cortical	0.3	0.6*	0.4	1	2	2
Hip Trabecular	0.1	0.4*	0.2*	<1	1	<1

BMD – bone mineral density; aBMD – areal BMD; vBMD – volumetric BMD

Absolute measurements reported to three significant digits (±10 N for strength, ± 1 mg/cm^3 for vBMD; ± 0.001 g/cm^2 for aBMD); percentages reported to one decimal place.

*Significantly different variances for phantom versus phantomless (non-overlapping 95% CI from a chi-square distribution).

Table 4

Comparison of the absolute measurements between phantom and phantomless calibrations: mean values (± standard deviation) for each measure, mean (paired) difference, and results from orthogonal regression analysis.

	Mean ± SD		Mean Difference [Percentage]*	Regression Analysis	
	Phantom	Phantomless		Slope (95% CI)	R^2
Clinical Measures:					
Vertebral Strength (N)	5400 ± 2240	5440 ± 2210	40 [0.7%]	1.01 (0.98–1.04)	0.99
Mid-Vertebral Trabecular vBMD (mg/cm^3)	92 ± 25	93 ± 25	1 [1.1%]	0.98 (0.94–1.03)	0.98
Femoral Strength (N)	3650 ± 1300	3680 ± 1310	30 [0.8%]	0.99 (0.95–1.03)	0.99
Femoral Neck aBMD (g/cm^2)	0.657 ± 0.137	0.661 ± 0.137	0.004 [0.6%]	1.00 (0.95–1.05)	0.98
Total Hip aBMD (g/cm^2)	0.723 ± 0.156	0.727 ± 0.156	0.004 [0.6%]	1.00 (0.95–1.05)	0.98
Other vBMD Measures (mg/cm^3):					
Spine Integral	158 ± 31	159 ± 31	1 [0.6%]	0.99 (0.95–1.04)	0.98
Spine Cortical	236 ± 43	237 ± 43	2 [0.8%]	0.99 (0.95–1.04)	0.98
Spine Trabecular	121 ± 30	122 ± 30	1 [0.8%]	1.00 (0.96–1.04)	0.99
Hip Integral	223 ± 40	225 ± 42	2 [0.9%]	0.96 (0.91–1.02)	0.97
Hip Cortical	383 ± 52	387 ± 55	4 [1.0%]†	0.95 (0.88–1.02)	0.95
Hip Trabecular	151 ± 35	151 ± 30	<1 [0.7%]	0.98 (0.93–1.04)	0.98

BMD – bone mineral density; aBMD – areal BMD; vBMD – volumetric BMD

Absolute measurements reported to three significant digits (±10 N for strength, ± 1 mg/cm^3 for vBMD; ± 0.001 g/cm^2 for aBMD).

*Percentage difference calculated as 100 x (Mean Difference)/(Average of Phantom and Phantomless measurements).

†Different than zero, $p<0.05$ paired T-test (phantom versus phantomless).

Bone. Author manuscript; available in PMC 2018 October 01.

Table 5

Precision errors for measures of hip areal BMD, mid-vertebral trabecular vBMD, and bone strength published since 1990, derived from CT scans. The coefficient of variation (%) is reported for both inter-operator (IP) and short-term (ST) precision errors. Studies using phantomless calibration are bolded.

	%CV		Demographic		
	IP	**ST**	**Women**	**Men**	**Age (yrs)**
Vertebral Strength					
Faulkner [5]		12.1	10		(premenopausal)
Fidler [3]	**0.7**		**29**	**10**	**59 ± 6**
(current study)	**0.5**		**25**	**15**	**67 ± 9**
Mid-Vertebral Trabecular vBMD					
Steiger [47]		1.5	9*		21–73
Gudmundsdottir [18]		**1.9**	**22**		**35–64**
Lang [48]	1.0	1.7	10		(osteoporotic)
Engelke [40]	0.6		29		(postmenopausal)
Fidler [3]	**0.4**		**29**	**10**	**59 ± 6**
(current study)	**0.5**		**25**	**15**	**67 ± 9**
Femoral Strength					
Cody [43] *(stance)*†		1.9	8	2	51–62
Carpenter [9] *(stance)*	1.6‡	5.9–7.4§	20		64 ± 3
Carpenter [9] *(fall)*	6.4‡	7.6–19.6§	20		64 ± 3
Fidler [3] ***(fall)***	**0.6**		**29**	**10**	**59 ± 6**
(current study, fall)	**0.4**		**25**	**15**	**67 ± 9**
Femoral Neck aBMD					
Khoo [42]		1.2‖	15	9	61 ± 11
Fidler [3]	**0.5**		**29**	**10**	**59 ± 6**
(current study)	**0.3**		**25**	**15**	**67 ± 9**
Total Hip aBMD					
Khoo [42]		1.0‖	15	9	61 ± 11
(current study)	**0.1**		**25**	**15**	**67 ± 9**

BMD – bone mineral density; aBMD – areal BMD; vBMD – volumetric BMD

Age is reported as either a mean (± standard deviation) or a range; if age was not available, a general description of the cohort was provided.

[*] Sex was not reported.

[†] The finite element analysis outcome was a stiffness value, which was converted via linear regression into a strength value.

[‡] Based on unpublished data [9].

[§] A range of %CV values were reported based on using six different types of anthropomorphic standardization phantoms for cross-calibration between the two CT scanners.

[//] %CV was calculated by dividing the reported RMSE (root-mean-square error) by the reported mean value.