Vertebral Bone Mineral Density Measured by Quantitative Computed Tomography With and Without a Calibration Phantom: A Comparison Between 2 Different Software Solutions

Josephine Therkildsen[1], Jesper Thygesen[2], Simon Winther[3], My Svensson[4], Ellen-Margrethe Hauge[5], Morten Böttcher[1], Per Ivarsen[6], and Hanne Skou Jørgensen[6]*

[1]Department of Internal Medicine, Hospital UnitWest, Herning, Denmark;

[2]Department of Clinical Engineering, Aarhus University Hospital, Aarhus, Denmark;

[3]Department of Cardiology, Aarhus University Hospital, Aarhus, Denmark;

[4]Department of Nephrology, Division of Medicine, Akershus University Hospital, Oslo, Norway;

[5]Department of Rheumatology, Aarhus University Hospital, and Department of Clinical Medicine, Aarhus University, Denmark; and

[6]Department of Nephrology, Aarhus University Hospital, Aarhus, Denmark

*Corresponding author:

Hanne Skou Jørgensen,

Department of Nephrology, Aarhus University Hospital, Aarhus, Denmark.

E-mail: hsjorgensen@clin.au.dk

Key Words: Bone density; chronic renal insufficiency; computed tomography; osteoporosis.

Introduction

Quantitative computed tomography (QCT) is an accepted method for assessing bone mineral density (BMD) to detect osteoporosis and monitor BMD changes over time (1).BMD can be measured using routine whole-body computed tomography (CT) scans, which is a widely used diagnostic modality (1). Any CT scanner-produced image can be used for BMD analyses as long as a suitable calibration technique is utilized (2).

Two calibration methods exist: phantom-based and phantom-less calibration. Phantom-based calibration uses reference values from a scanned phantom, which contains regions of known concentrations of calcium hydroxyapatite. The phantom can be scanned simultaneously with the patient (synchronous) or separately as a series of phantom scans (asynchronous), which the software will then use to generate calibration data. In contrast, internal calibration uses reference values from conversion factors calculated based on the patient's own tissues. The literature on phantom-less internal calibration software is limited, and the 2015 ISCD official position statement is that: "There is insufficient evidence to judge the feasibility of internal calibration" (3).

There are many different commercially available software solutions for BMD analysis, but there is a lack of knowledge as to the interchangeability of these methods. Mueller et al found that precision of the phantom-less method was inferior to the phantom-based, but still concluded that the phantom-less method is robust in measuring BMD (4).

This study aimed to determine agreement and precision between BMD measurements by the phantom-less internal tissue calibration method by Extended Brilliance Workspace (Philips Healthcare, Cleveland, OH) and the asynchronous phantom-based calibration by QCT Pro (Mindways Software Inc., Austin, TX) in a group of patients with chronic kidney disease (CKD).

Materials and Methods

Patients

Patients with CKD, referred for cardiovascular assessment before kidney transplantation, were included from 9 hospitals. Criteria of inclusion and exclusion have been described in a prior publication (5). Written informed consent was obtained from all patients. The study followed the principles of the declaration of Helsinki and approval was given by the Central Denmark Regional Committee on Health Research Ethics and the Danish Data Protection Agency. The study was registered at clinicaltrials.gov (NCT01344434).

Images Acquisition and Analysis Methods

The scan protocol for the CT angiography has been described in detail in a previous paper (6). In brief, CT scans were performed using a dual-source scanner (SOMATOM definition Flash; Siemens Healthcare, Erlangen, Germany) with contrast-enhancement. A contrast dose of 95 mL ioversol (Optiray 350 mg/mL; Mallinckrodt, Hennef, Germany) was given intravenously. A mean delay time of 30 ± 5 s was present between contrast administration and the imaging procedure. CT scans were high-pitch flash scans with a gantry rotation of 0.28 s and a pitch of 3.4.The detector collimation was 2 × 64 × 0.6 mm, and images were reconstructed to 3-mm thickness.A standard soft tissue kernel was used throughout (syngo.via; Siemens Healthcare, Erlangen, Germany).

Bone Density Analyses and Fracture

Mean BMD value in units of milligram per cubic centimeter (mg/cm3), as well as BMD for each separate vertebra was derived from both software solutions. Fractured vertebrae were excluded from BMD analyses. Deformed vertebrae were identified on 2-dimensional sagittal reconstruction images (7) by 1 investigator (HSJ), and reviewed by a radiologist and signed a final fracture grade (8) based on the classification described by Genant et al (9).

Phantom-Based BMD

The phantom-based BMD measurements were performed using the software solution QCT Pro (Mindways Software Inc.,Austin, TX) and the quality assurance calibration phantom Mindways Solid (Mindways Software Inc.; Fig. 1). BMD measurements were performed on 3 consecutive vertebrae from T12 to L4 by a single investigator (HSJ) blinded to study data. L1 to L3 were preferred, although T12 and/or L4 were allowed in cases of observed deformity, pathology, or fractures of L1, L2, or L3. An elliptical volume of interest (VOI) was automatically placed in the anterior part of the vertebral body (Fig. 2A) and manually adjusted when necessary.We aimed at placing the largest possible VOI, avoiding the posterior venous plexus and any focal pathology, such as bone islands and calcified herniated disks.

Phantom-Less BMD

The phantom-less measurements were performed using the Extended Brilliance Workspace (Philips Healthcare, Cleveland, OH).BMD analysis was performed by a single investigator (JT) blinded to study data and to the phantombased results.The analyses were anatomically matched to the phantom-based measurements. Elliptical VOIs were manually placed in the anterior part of the mid-vertebral body (Fig. 2B) by a procedure identical to the one detailed for the phantom-based measurement.VOIs of muscle and fat were placed in approximately the same position on each image, in the posterior subcutaneous fat on the right side and the paraspinal muscle group on the left side, throughout. VOI size and shape for muscle and fat were adjusted for optimal fit. The main goal was to achieve a normal distribution on the Hounsfield unit (HU) histogram with the predominant component of the Gaussian fit within the specified values provided by the software.Thus, we aimed to achieve fat HUs between −150 and 0, muscle HUs above 40, and trabecular HUs above 15.BMD values were not evaluated during VOI placement, and HUs of muscle and fat were not matched for consecutive measurements on the same patient, as this was not stated in the Philips protocol. A reanalysis of 5 outliers was performed weeks after the original analysis to evaluate if any technical mistakes were present.

Statistical Analysis

Statistical analyses were done using software package STATA/IC 13.1 (StataCrop LP, College Station, TX). All variables were visually assessed for normal distribution using QQ-plots. Normally distributed variables are given as mean with standard deviation (SD), whereas skewed variables are given as median with interquartile range [IQR]. Comparative data are given as absolute mean difference with 95% confidence interval (CI) and 95% prediction interval (limits of agreement).XY- and Bland-Altman-plots were used to assess systematic bias between methods.A paired Student’s t test was used to test the differences in BMD between methods, with a 2-sided p-value <0.05 considered statistically significant. Univariate correlation, displayed as Pearson’s correlation coefficients r, was used to explore possible associations between patient-related factors and the between-methods difference in BMD.

Reproducibility: Inter- and Intraoperator Variability

For the analysis of intraoperator variability, a reanalysis of 53 patients (36%) using both software solutions was performed by the 2 investigators (JT and HSJ). Furthermore, JT performed reanalysis of 30 patients (20%) using the phantom-based method and HSJ performed reanalysis of 29 patients (19%) using the phantom-less method to investigate interoperator variability. Intraoperator and interoperator variability was assed using coefficient of variation (CV %) as described by Glüer et al (10) and the intraclass correlation coefficient (ICC). ICC statistic was applied using the 2-way mixed effect analysis of variance model. Furthermore, kappa-statistic based on categorical variables and Bland-Altman plots were made (displayed in the supplementary data).

Results

Of the 157 patients included in the study, we excluded 6 patients because of lack of an intravenous access for contrast media administration and 2 patients based on technical failure. Thus, the final analysis included 149 patients. Demographic data of participants are displayed in Table 1. Median age was 54 yr (range 23–72), 69% were men, and 33% were diabetics. Patients not yet on dialysis (n = 91, 61%) had a median estimated glomerular filtration rate of 11 [9, 14] mL/min/m3.

XY- and Bland-Altman plots are shown in Fig. 3. BMD measured by the phantom-based method was systematically higher than BMD measured by the phantom-less method. At increasing levels of BMD, this shifted toward higher values by the phantom-less method, which was caused by a few high measurements. In a subsequent reanalysis excluding outliers with a mean difference between methods of ±50 mg/cm3, no systematic bias could be visualized (Fig. 3, bottom half). The Bland-Altman plot similarly showed an increase in variance at higher BMD values. When the outliers were excluded, an equal variance was observed across the range of BMD values. A reanalysis of the 5 outliers was performed (Table 2), but no apparent reason for the BMD deviation could be found. Details of the original 5 outliers identified are displayed in the supplementary data.

The calculated mean difference between the 2 software solutions at different levels of the lumbar spine is shown in Table 3.The average difference was slight (3.3 mg/ cm3), and not statistically significant ($p = 0.07$), but the variation was high, with large standard deviations and wide prediction intervals. Looking at the relative difference, the phantom-based measurements were 5.1% higher than the phantom-less (CI: 2.2%–8.1%, $p < 0.001$). If the 5 outliers were excluded, the mean difference increased to 5.4 mg/cm3 (CI: 3.0–7.9 mg/cm3, $p < 0.001$), with limits of agreement −24 to +35 mg/cm3. The relative difference without outliers was 6.1% (CI: 3.6%–8.6%, $p < 0.001$).

We performed univariate correlation analyses to investigate if patient-related factors were related to the between method difference in BMD. No significant correlations were found with age ($r = 0.15$, $p = 0.06$), BMI ($r = −0.12$, $p = 0.15$), weight ($r = −0.12$, $p = 0.15$), or abdominal circumference ($r = −0.04$, $p = 0.69$). As suspected from the XY- and Bland- Altman plots, the phantom-less BMD measurement was positively correlated to the between-method difference ($r = 0.55$, $p < 0.001$).The same was not true for the phantom based measurement ($r = 0.10$, $p = 0.23$).

The intraoperator variability was higher for the phantomless method with a CV of 5.8% and ICC of 0.965 (CI: 0.936–0.980), compared to corresponding values of 0.8% and 0.999 (CI: 0.998–0.999) by the phantom-based solution. Similarly, the interoperator analysis yielded a CV of 5.8% and an ICC of 0.979 (CI: 0.956–0.990) for the phantom-less method, and a CV of 1.8% and an ICC of 0.996 (CI: 0.992–0.998) for the phantom-based method. No systematic bias was found investigating

intra- and interoperator variability. The phantom-based method had higher kappa values compared to the phantom-less method.

Discussion

In this study we examined the differences between 2 commercially accessible software solutions for measuring BMD of the spine, 1 using phantom-based calibration, and the other, phantom-less internal calibration. Although there was no significant difference in the average BMD values between the 2 software solutions, we found a wide prediction interval ranging from −47 to +40 mg/cm3, illustrating large interindividual differences. BMD values by the phantom-based approach were systematically higher than the phantom-less measurements. The variance between methods was relatively constant, although at high BMD values, the between-method difference increased, which was mainly due to few very high measurements by the phantomless method. Mueller et al reported similar findings with a mean difference between methods of 0.9 mg/cm3, with a narrower prediction interval at −15 to +14 mg/cm3 (4). Similar to our findings, they found increasing variation of between-methods difference at higher BMD values, but outliers were based on high measurements by the phantombased method (4) and not the phantom-less method as we saw in our dataset.

The phantom-less software solution is designed to eliminate the need for a calibration phantom, which should ease the implementation into clinical practice. However, we found that BMD measurement by this approach was highly sensitive to small changes in VOI location, despite apparently correct placement in the desired tissue. We identified 5 outliers with a between-method difference greater than ±50 mg/cm. These outliers were re-evaluated, but we could not identify any obvious mistakes in the internal calibration technique. Although we strove to achieve HU values in the specified interval, this was not always possible, as can be seen in Table 2.This could lead to calibration failure resulting in high BMD values. One outlier had very sparse muscle tissue and 1 low fat tissue, which could have contributed to the high BMD values measured. We note that although the phantom-based approach overall produced higher BMD values, 4 of the 5 outliers had very high values by the phantom-less software, with high-normal values by the phantom-based method. Thus, the phantom-less approach may be less accurate at higher levels of BMD.

In contrast to the phantom-less approach, the phantom-based method has a highly structured and detailed protocol with an internal error finding procedure; it specified that if BMD deviated more than 20 mg/cm3 between vertebrae, the analysis should be checked for errors. The phantom-less protocol did not state that the analyses should be adjusted depending on BMD values, and so, this was not performed. Upon subsequent review, we noted large differences in the phantom-less BMD between vertebrae of the same patient. A similar quality check of the calibration data when using the phantom-less approach may therefore be recommended.

None of the investigators performing the BMD analysis were radiologists. This could have contributed to variability in the BMD measurements, especially concerning the phantom-less method, which is highly dependent on correct placement of the VOIs in the desired tissue. The phantom-less method demand critical judgment when placing the VOI.

Patients with CKD suffer from reduced muscle size (11), which could have affected the results, as the phantomless method is highly dependent on the presence of calibration tissue. It is required to check whether the HU of the VOI are correct to avoid calibration errors. Theoretically, actors such

as general muscle atrophy, increased fat infiltration, increased marrow fat, hydration, and impaired hemodynamics could affect the BMD measurements, but there is a lack of data on these issues. All our patients suffered from severe CKD and we expect these patients to have some changes in normal body composition. Using a non-CKD population could provide less variation in BMD measurements for the phantom-less method. Our interoperator variability for phantom-less analyses was 5.8% which is higher, but comparable to what was reported by Mueller et al with a reanalysis precision CV of 4.0% (4). In their interoperator analysis they matched HU of fat and muscle VOIs when reanalyzing, as it was argued that this would improve precision. We did not follow this approach. Their conclusion was that phantom-based methods ave a precision up to 3 times better than the phantom-less method on noncontrast CT scans.

Our interoperator analysis or the phantom-based method revealed a CV of 1.8%. This is comparable to previous studies reporting a reanalysis precision of 1.7% when analyzing the spine in 29 postmenopausal women (12). Our inter- and intraoperator variability analyses are reanalysis of the same single data set of a single scan. Interscan variability have previously been reported to be 2.8% using thoracic QCT (13).

We found the phantom-based software user-friendly and robust. However, the inherent drawback of this solution is the use of a phantom. Disadvantages to external phantoms include air gap artifacts, dependency on patient size, and other patient-moderated artifacts, resulting in repositioning errors affecting the precision (14). AutomaticVOI placement for spine BMD analysis results in lower mean intrasubject variation than manual placement and could improve precision (15). This could contribute to the difference observed between the phantom-based and phantomless method, as the phantom-less method offered manual placement of the VOI only, whereas the phantom-based method offered automatic placement of the VOI with the possibility to adjust.

Our results suggest agreement between these 2 methods is insufficient to allow them to be used interchangeably in patients with CKD.

A major strength of this study is the relatively high number of patients and that we evaluated both intra- and interoperator variation for both methods in question. Limitations include the heterogeneity of our study cohort in age, cause of CKD, duration of disease, and degree of comorbidity. Further, the use of contrast administration does affect BMD measurements (3) and could have had an impact on the agreement between our methods. Data indicate that internal tissue and trabecular bone are unequally affected by contrast medium (16).As the external phantom is not affected by contrast media, this could have affected the between-method differences in BMD. Lastly, our scan protocol was not standardized to bone measurements, and phantoms were scanned asynchronously, which presumes a high stability of the scanner, which may not always be present (3).

In conclusion, we found a high correlation between the BMD values by 2 different software solutions using different calibration techniques, but with large interindividual differences. The phantom-less solution had a higher variability than the phantom-based software. The agreement between methods is insufficient and we would not recommend the 2 methods to be used interchangeably in patients with CKD.We strongly encourage future studies to include patients without CKD, to see if this is also the case for other patient groups.

Acknowledgments
The authors are grateful to all study participants and wish to thank Shadman Neghabat (Department of Radiology, Aarhus University Hospital) for diagnosing vertebral fractures. We also acknowledge the work done by the staff at Departments of Nephrology at Aalborg, Viborg, and Holstebro hospitals.

References

1. Mao SS, Li D, Luo Y, et al. 2015 Application of quantitative computed tomography for assessment of trabecular bone mineral density, microarchitecture and mechanical property. Clin Imaging 40(2):330–338.

2. Link TM, Lang TF. 2014 Axial QCT: clinical applications and new developments. J Clin Densitom 17(4):438–448.

3. Engelke K, Lang T, Khosla S, et al. 2015 Clinical use of quantitative computed tomography-based advanced techniques in he management of osteoporosis in adults: the 2015 ISCD official positions—part III. J Clin Densitom 18(3):393–407.

4. Mueller DK,Kutscherenko A, Bartel H, et al. 2011 Phantomless QCT BMD system as screening tool for osteoporosis without additional radiation. Eur J Radiol 79(3):375–381.

5. Winther S, Svensson M, Jorgensen HS, et al. 2015 Diagnostic performance of coronary CT angiography and myocardial perfusion imaging in kidney transplantation candidates. JACC Cardiovasc Imaging. 8(5):553–562.

6. Jorgensen HS, Winther S, Bottcher M, et al. 2016 Effect of intravenous contrast on volumetric bone mineral density in patients with chronic kidney disease. J Clin Densitom 19(4):423–429.

7. Bauer JS, Muller D, Ambekar A, et al. 2006 Detection of osteoporotic vertebral fractures using multidetector CT. Osteoporos Int 17(4):608–615.

8. Grados F, Fechtenbaum J, Flipon E, et al. 2009 Radiographic methods for evaluating osteoporotic vertebral fractures. Joint Bone Spine 76(3):241–247.

9. Genant HK, Jergas M, Palermo L, et al. 1996 Comparison of semiquantitative visual and quantitative morphometric assessment of prevalent and incident vertebral fractures in osteoporosis.The Study of Osteoporotic Fractures Research Group. J Bone Miner Res 11(7):984–996.

10. Gluer CC, Blake G, Lu Y, et al. 1995 Accurate assessment of precision errors: how to measure the reproducibility of bone densitometry techniques. Osteoporos Int 5(4):262–270.

11. Segura-Orti E, Gordon PL, Doyle JW, Johansen KL. 2017 Correlates of physical functioning and performance across the spectrum of kidney function. Clin Nurs Res 1054773816689282.

12. Engelke K, Mastmeyer A, Bousson V, et al. 2009 Reanalysis precision of 3D quantitative computed tomography (QCT) of the spine. Bone 44(4):566–572.

13. Budoff MJ, Hamirani YS, Gao YL, et al. 2010 Measurement of thoracic bone mineral density with quantitative CT. Radiology 257(2):434–440.

14. Boden SD, Goodenough DJ, Stockham CD, et al. 1989 Precise measurement of vertebral bone density using computed tomography without the use of an external reference phantom. J Digit Imaging 2(1):31–38.

15. Louis O, Luypaert R, Kalender W, Osteaux M. 1988 Reproducibility of CT bone densitometry: operator versus automated ROI definition. Eur J Radiol 8(2):82–84.

16. Kaesmacher J, Liebl H, Baum T, Kirschke JS. 2016 Bone mineral density estimations from routine multidetector computed omography: a comparative study of contrast and calibration effects. J Comput Assist Tomogr 41(2):217–223.

Tables

Table 1 Characteristics of participating patients with chronic kidney disease

Characteristics	All (n = 149)
Age , yr	54 [45, 64]
Gender, M:F	103:46
Body mass index, kg/cm²	25.8 (4.3)
Abdominal circumference, cm	98 (13)
Body weight, kg	78 (15)
Cause of chronic kidney disease:	
Diabetes mellitus type 1 or 2	40 (27%)
Hypertension or glomerulosclerosis	38 (26%)
Glomerulonephritis or vasculitis	34 (23%)
Adult polycystic kidney disease	20 (13%)
Other/unknown	17 (11%)
Maintenance dialysis therapy	58 (39%)
Previous fragility fracture	26 (17%)
Prevalent vertebral fracture	16 (11%)
Bisphosphonate-therapy	1 (1%)
Calcimimetics	2 (1%)
Phosphate binders	109 (73%)
Calcium-containing phosphate binders	80 (54%)
Vitamin D supplements	37 (25%)
Vitamin D receptor activators	105 (70%)

Data are mean (SD), median [IQR] or n (%)

Table 2 Reanalysis of the 5 outliers with details from the phantom-less internal calibration

		1	**2**	**3**	**4**	**5**	Sample (n = 41, 28 %)	
Details of BMD analysis								
Between-method difference	mg/cm³	-39	-80	-92	51	-91	-2	(-26, 30)
Phantom-based BMD	mg/cm³	178	238	158	123	117	129	(54, 218)
Phantom-less BMD	mg/cm³	217	318	250	72	208	127	(48, 223)
Details of internal calibration								
Vertebrae VOI	HU [>15]	282	373	224	111	155	182	(76, 307)
Reference tissues, fat	HU [-150 to 0]	-115	-83	-44	-103	-28	-103	(-118, -78)
Reference tissues, muscle	HU [>40]	34	46	47	50	42	49	(35, 68)

Data are mean values with range for sample group. Abbr.: BMD = bone mineral density, HU = Hounsfield Units, VOI = volume of interest

Table 3 Difference in lumbar spine bone mineral density between phantom-less and phantom-based software solutions

Site	*N*	Phantom-less BMD (mg/cm³)	Phantom-based BMD (mg/cm³)	Mean difference (Phantom-based – phantom-less)	Min to max	Prediction interval
L1	(n=140)	125 (51)	126 (40)	1.0 (26)	-47 to 135	[-52 to 50]
L2	(n=148)	120 (47)	125 (40)	4.6 (24)	-59 to 117	[-52 to 43]
L3	(n=144)	114 (47)	121 (39)	6.2 (22)	-88 to 106	[-50 to 37]
Average	N/a	120 (47)	123 (39)	3.3 (22)	-63 to 118	[-47 to 40]

Data are mean (SD) and 95% prediction interval. Abbr.: BMD = bone mineral density

Figures

Fig. 1. Illustration of volume of interest placement in the quality assurance phantom, Mindways Solid.

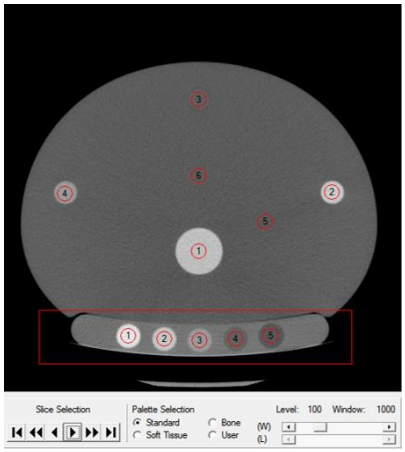

Fig. 2. (A) Illustration of the QCT Pro Mindways phantom-based software method with volume of interest placed in the vertebrae. (B) Illustration of the Philips phantom-less software method. Volumes of interest placed in vertebrae, muscle, and fat tissue. Hounsfield unit histograms of the respective tissues are displayed at the bottom. QCT, quantitative computed tomography.

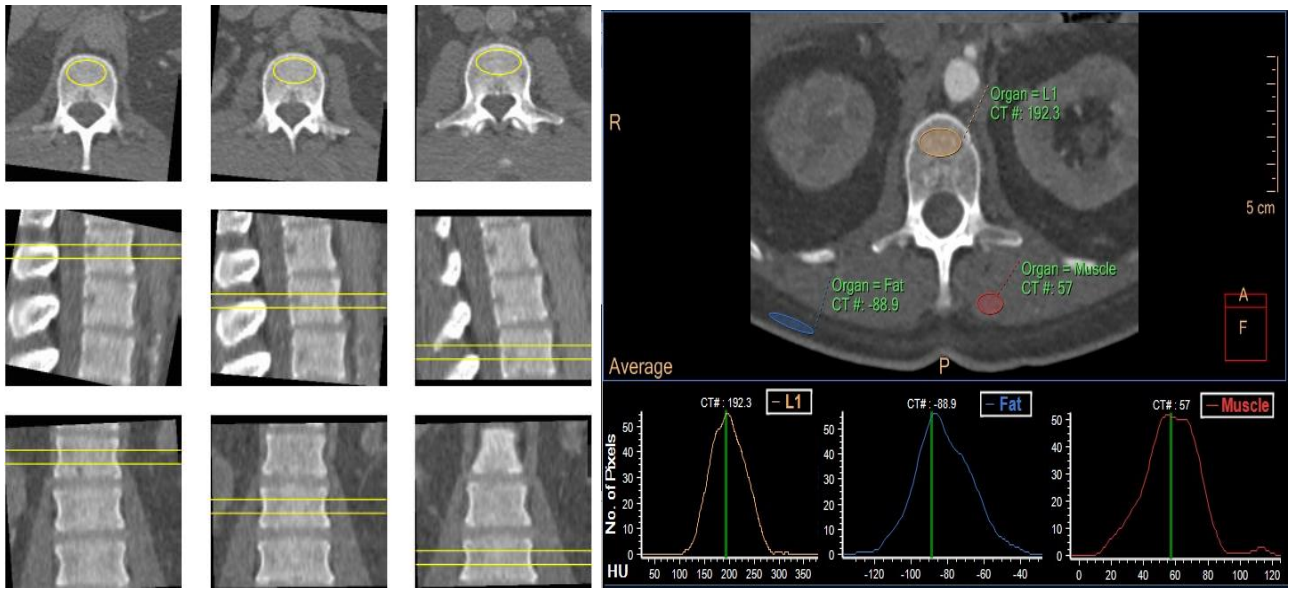

Fig. 3. Comparison of spinal bone mineral density by the phantom-based and the phantom-less approach. (Top, left): XY plot with fitted function (dotted line). (Top, right): Bland-Altman plot with mean difference (full line) and 95% prediction interval (dotted lines). (Bottom): XY- and Bland-Altman plots after exclusion of 5 outliners with a difference of ±50 mg/cm3.