

# Inter-observer and inter-examination variability of manual vertebral bone attenuation measurements on computed tomography

Esther Pompe<sup>1</sup> · Pim A. de Jong<sup>2</sup> · Werner U. de Jong<sup>2</sup> · Richard A. P. Takx<sup>2</sup> · Anouk L. M. Eikendal<sup>2</sup> · Martin J. Willemink<sup>2</sup> · Matthijs Oudkerk<sup>3</sup> · Ricardo P. J. Budde<sup>4</sup> · Jan-Willem J. Lammers<sup>1</sup> · Firdaus A. A. Mohamed Hoesin<sup>2</sup>

Received: 4 May 2015 / Revised: 25 August 2015 / Accepted: 30 November 2015 / Published online: 22 January 2016  
© The Author(s) 2016. This article is published with open access at Springerlink.com

## Abstract

**Objective** To determine inter-observer and inter-examination variability of manual attenuation measurements of the vertebrae in low-dose unenhanced chest computed tomography (CT).

**Methods** Three hundred and sixty-seven lung cancer screening trial participants who underwent baseline and repeat unenhanced low-dose CT after 3 months because of an indeterminate lung nodule were included. The CT attenuation value of the first lumbar vertebrae (L1) was measured in all CTs by one observer to obtain inter-examination reliability. Six observers performed measurements in 100 randomly selected CTs to determine agreement with limits of agreement and Bland-Altman plots and reliability with intraclass correlation coefficients (ICCs). Reclassification analyses were performed using a threshold of 110 HU to define osteoporosis.

**Results** Inter-examination reliability was excellent with an ICC of 0.92 ( $p < 0.001$ ). Inter-examination limits of agreement ranged from -26 to 28 HU with a mean difference of  $1 \pm 14$  HU. Inter-

observer reliability ICCs ranged from 0.70 to 0.91. Inter-examination variability led to 11.2 % reclassification of participants and inter-observer variability led to 22.1 % reclassification. **Conclusions** Vertebral attenuation values can be manually quantified with good to excellent inter-examination and inter-observer reliability on unenhanced low-dose chest CT. This information is valuable for early detection of osteoporosis on low-dose chest CT.

## Key Points

- Vertebral attenuation values can be manually quantified on low-dose unenhanced CT reliably.
- Vertebral attenuation measurements may be helpful in detecting subclinical low bone density.
- This could become of importance in the detection of osteoporosis.

**Keywords** Bone mineral density · Computed tomography · Osteoporosis · Inter-observer variability · Intra-observer variability

**Electronic supplementary material** The online version of this article (doi:10.1007/s00330-015-4145-x) contains supplementary material, which is available to authorized users.

✉ Esther Pompe  
e.pompe@umcutrecht.nl

<sup>1</sup> Department of Pulmonology, University Medical Center Utrecht, P.O. Box 85500, 3508 GA Utrecht E.03.511, The Netherlands

<sup>2</sup> Department of Radiology, University Medical Center Utrecht, Utrecht, The Netherlands

<sup>3</sup> Department of Radiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

<sup>4</sup> Department of Radiology, Erasmus Medical Center, Rotterdam, The Netherlands

## Abbreviations

BMD	Bone mineral density
BMI	Body mass index
CT	Computed tomography
DXA	Dual-energy x-ray absorptiometry
HU	Hounsfield unit
ICC	Intraclass correlation coefficient
L1	First lumbar vertebral body
QCT	Quantitative computed tomography
ROI	Region of interest
SD	Standard deviation
T12	12th thoracic vertebral body

## Introduction

Osteoporosis is a worldwide health problem contributing considerably to health care costs, morbidity and mortality [1]. Due to the increasing age of the population, this problem will become even more substantial over time. Osteoporosis is a good candidate for screening, because it has a long preclinical phase and cost-effective therapeutic possibilities are readily available [2]. Therefore, early detection may be a reasonable strategy to prevent osteoporosis related fractures [3].

Bone mineral density (BMD) can be derived from quantitative computed tomography (QCT), ultrasound, and dual energy x-ray absorptiometry (DXA), but the World Health Organization (WHO) definition for osteoporosis only includes the DXA T-score [4]. DXA is a strong predictor of bone deterioration and fracture risk [5]. Nevertheless, osteoporosis remains underdiagnosed in the general population [6]. Recent studies proposed the use of regular clinical computed tomography (CT)-scans for bone mineral density assessment as an opportunistic screening method [7–14]. In this way, BMD surrogates could be derived from CT without additional radiation dose. Deterioration of BMD could be determined in the early stages of disease, which could prompt suitable medical treatment of osteoporosis, and thus prevent fractures in patients at risk. Additionally, scans performed in screening programs such as lung cancer screening or coronary artery disease evaluation may also be used, since BMD measurements on CT have been shown to predict all-cause mortality in lung cancer screening participants [15].

Bone density measurements on CT are mostly performed by manually placing a region of interest (ROI) in a lumbar or thoracic vertebra. Although in the future ROI measurements may possibly be executed automatically using software, the placement of a ROI is currently performed manually, and thus may vary in size, shape and location [16]. This could introduce variability in measurements and may depend on the experience of the observer. Furthermore, the inter-examination reliability is currently unknown. High inter-examination reliability would enable monitoring changes over time; for example, after an intervention. A few studies addressed the reproducibility of bone density assessment of the vertebrae [9, 17], but data is missing for unenhanced low-dose thoracic CT. Therefore, the aim of this study was to assess the inter-observer and inter-examination reliability and agreement in attenuation measurements of the vertebrae on low-dose unenhanced CT.

## Materials and methods

### Subjects

The study population was derived from the Dutch and Belgian Lung Cancer Screening Trial (NELSON). For

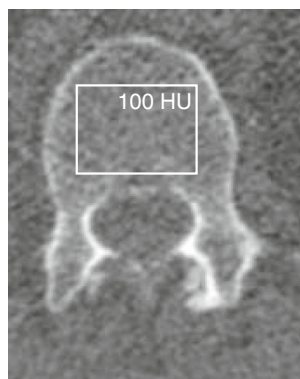
this study, subjects from the University Medical Center Utrecht were included. A detailed description of study rationale and design is published elsewhere [18, 19]. In short, current and former smokers aged 50–75 years were included. Included participants had a smoking history of >15 cigarettes/day during >25 years or >10 cigarettes/day during >30 years. The NELSON trial was approved by the Dutch Ministry of Health and the local institutional ethical review board. Written informed consent was obtained for all participants. Participants who underwent a short-term follow-up CT after 3 months because of an intermediate risk lung nodule were included. Participants were excluded if the scan protocol, especially the kilovoltage, differed between baseline and follow-up, or when the interval between the CTs was longer than 100 days.

### Image acquisition

All participants underwent a low-dose inspiratory volumetric CT scan using a 16-slice scanner (Brilliance 16P or MX8000 IDT; Philips Healthcare, Best, The Netherlands). The same scanning protocols were used at baseline and follow-up in the subjects included in this study. CT data were obtained using  $16 \times 0.75$  mm collimation (pitch = 1.3) [20]. No intravenous contrast injection was applied. Participants weighing less than 80 kg were scanned with 120 kVp at 30 mAs. Participants weighing 80 kg or more were scanned with 140 kVp at 30 mAs. Slice thickness was 1.0 mm and axial images were reconstructed at 0.7 mm increment, using a smooth reconstruction filter (B-filter; Philips Healthcare, Best, The Netherlands). The scanners were calibrated according to the manufacturers recommendations and screening scans were obtained within 24 hours after calibration.

### Assessment of BMD – ROI placement

Before measurements were made, all observers attended an interactive training session to become skilful in the measurement technique. First, observers were instructed to evaluate the first lumbar vertebral body (L1) for the presence of fractures and focal heterogeneities. Secondly, the upper part of L1 was identified by defining the area between the endplate and the entrance of vessels at the midportion. Thirdly, a rectangular ROI was as large as possible and positioned in a homogeneous area of trabecular bone, without inclusion of the cortex and inhomogeneous areas. All measurements were performed in axial view. The mean CT attenuation was measured in Hounsfield Units (HU). An example is given in Fig. 1. If L1 was fractured, inhomogeneous or not visualized, the first intact vertebra from L1 upwards was measured. For instance, if L1 was fractured, the twelfth thoracic vertebra (T12) was used.



**Fig. 1** Example of placement of region of interest (ROI) in the first lumbar vertebra (L1) of a participant. Axial image of L1 in bone window (C: 300; W: 1600). A ROI is placed in the upper part of L1 between the endplate and the entrance of vessels at the midportion. The ROI was as large as possible and positioned in a homogenous area without inclusion of the cortex

### Image analysis

Six observers with various levels of experience in CT reading participated in the current study: one board certified chest radiologist with 10 years of experience (P.J.), one board certified radiologist with 8 years of experience (R.B.), three research physicians with 2 years of experience (R.T., M.W. and E.P.) in CT reading and one medical student (W.J.).

To obtain inter-observer agreement, all six observers performed ROI measurements in 100 randomly selected CT scans. Measurements were conducted independently by all observers without knowledge of the outcomes of the other observers. For the inter-examination reliability and agreement, one observer (E.P.), who showed good agreement with the experienced observers, examined all baseline and follow-up CTs. Measurements at follow-up scans were made in the same session as the baseline scans. All scans were measured in a random order and the reader was blinded for acquisition date. Measurements were made at the same vertebral level for baseline and follow-up. If measurements in L1 were not feasible because of inhomogeneous areas or was not included on CT, the first intact vertebra from L1 upwards was measured both at baseline and at follow-up. Observers performed measurements in bone window (C: 300; W: 1600) and were blinded for subject characteristics.

### Statistical analysis

Inter-examination reliability was assessed using the single measures intraclass correlation coefficient (ICC). ICCs were compared using a student t-test. The limits of inter-examination agreement were defined as the mean difference  $\pm 1.96 \times$  the standard deviation (SD) and was plotted using the Bland-Altman method [21].

Inter-observer reliability was estimated for all measurements made in 100 participants, using the single measures ICC. Inter-observer agreement was calculated using the mean difference  $\pm$  SD and was assessed by a graphical method proposed by Jones et al. [22] This method is based on the Bland-Altman graphical method for the assessment of agreement between two observers, and is modified to allow for agreement between multiple observers. The limits of agreement from the mean represent how different a measurement of an individual observer can be compared with the mean measurement of all observers.

In addition, the effect of inter-examination variability or inter-observer variability on reclassification of patients as osteoporotic versus non-osteoporotic based on previously suggested cutoffs was calculated. To define osteoporosis a threshold of 110 HU was used, as derived from Pickhardt et al. [7] All analyses were performed using SPSS Version 20.0 (SPSS, Chicago, Illinois, USA). *P* values below 0.05 were considered statistically significant. Results are reported according to the GRRAS guidelines [23].

## Results

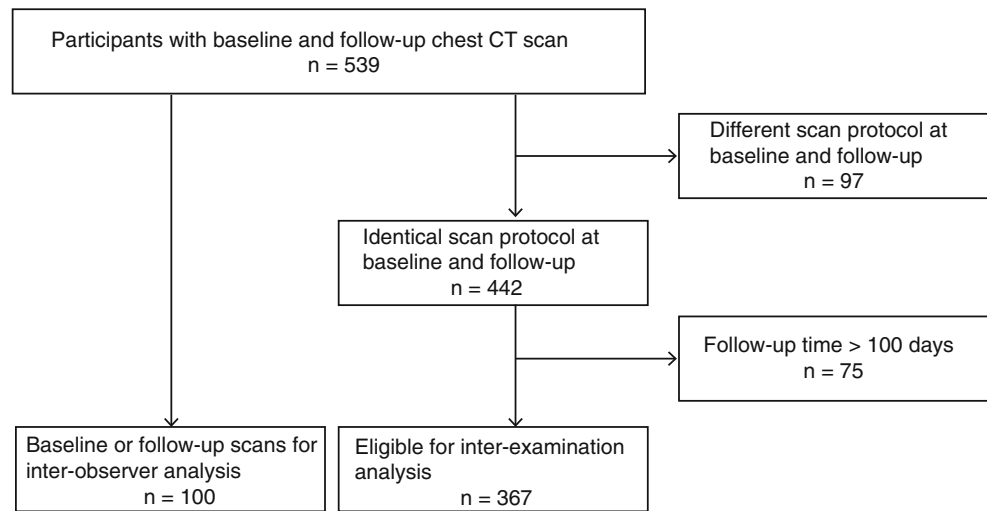
### Baseline characteristics

In total, 539 participants were rescanned after 3 months because of an indeterminate lung nodule. One hundred randomly selected scans were selected for inter-observer analysis. For inter-examination analysis, 97 participants were excluded because of the use of a different kVp at baseline and follow-up. Another 75 participants were excluded because of a follow-up time of more than 100 days; this resulted in 367 eligible participants for inter-examination analysis. A flowchart of in- and exclusion of participants is shown in Fig. 2. Mean  $\pm$  SD age was  $60.6 \pm 5.9$  years. Median time between the baseline and follow-up CT was 91 (P25 – P75: 91 – 91) days. Additional baseline characteristics are presented in Table 1.

### Inter-examination variability

287 (78.2 %) measurements were made at level L1, 69 (18.8 %) at level T12, 10 (2.7 %) at level T11 and 1 (0.3 %) at level T10. Mean  $\pm$  SD bone density was  $108 \pm 35$  HU at baseline and  $107 \pm 35$  HU at follow-up. The inter-examination agreement for ROI measurements of 367 participants was excellent with an ICC of 0.92 (0.90 – 0.94,  $p < 0.01$ ) and did not differ significantly between men and women (0.92 (0.91–0.94) vs. 0.94 (0.90–0.96)). Inter-examination ICCs did not differ significantly between vertebral levels with an ICC for L1 of 0.92 (0.90–0.94), an ICC for T12 of 0.91 (0.86–0.95) and an ICC for T11 of 0.92 (0.72–0.98). Mean  $\pm$  SD difference between baseline and follow-up was  $1 \pm 14$  HU. The inter-

**Fig. 2** Flow-chart of in- and exclusion of participants



examination agreement is plotted in Fig. 3. Limits of agreement were -26 and 28 HU.

**Inter-observer variability**

Eleven (1.8 %) measurements were missing due to software problems. For this reason, multiple imputations were used to complete the data set (20 imputations). Of all measurements, the mean ± SD HU-value was 100 ± 28 HU. Mean HU-values for all observers ranged from 95 to 105 HU, with SD ranging from 28 to 33 HU. Overall inter-observer reliability for measuring HU-attenuation of the vertebrae was excellent, with an ICC of 0.82 (*p* < 0.001). All ICCs compared between two observers are shown in Table 2 and varied from moderate (0.70) to excellent (0.91). The mean difference ± SD between all observers was 1 ± 6 HU. The inter-observer agreement of all 100 examinations is plotted in Fig. 4. Limits of agreement with the mean ranged from -12 to 12 HU.

**Table 1** Baseline characteristics (*n* = 367)

Characteristic	Value
Gender, n (%)	
Male	310 (84.5 %)
Female	57 (15.5 %)
Age (years), mean ± SD	60.6 ± 5.9
Smoking status, n (%)	
Current smoker	217 (59.1 %)
Former smoker	150 (40.9 %)
Pack years (years), Median (25th percentile – 75th percentile)	38.7 (29.7–53.2)

SD: standard deviation; HU: Hounsfield Unit

**Reclassification analysis**

For reclassification analyses, only measurements at L1 were used (*n* = 287). Based on a threshold of 110 HU, 159 (55.4 %) participants were classified as having osteoporosis at baseline and 163 (56.8 %) participants at follow-up. Fourteen (4.9 %) participants were classified as having osteoporosis at baseline, but not at follow-up, and 18 (6.3 %) participants were classified as having osteoporosis at follow-up, but not at baseline. Table 3 shows reclassification results due to inter-examination variation.

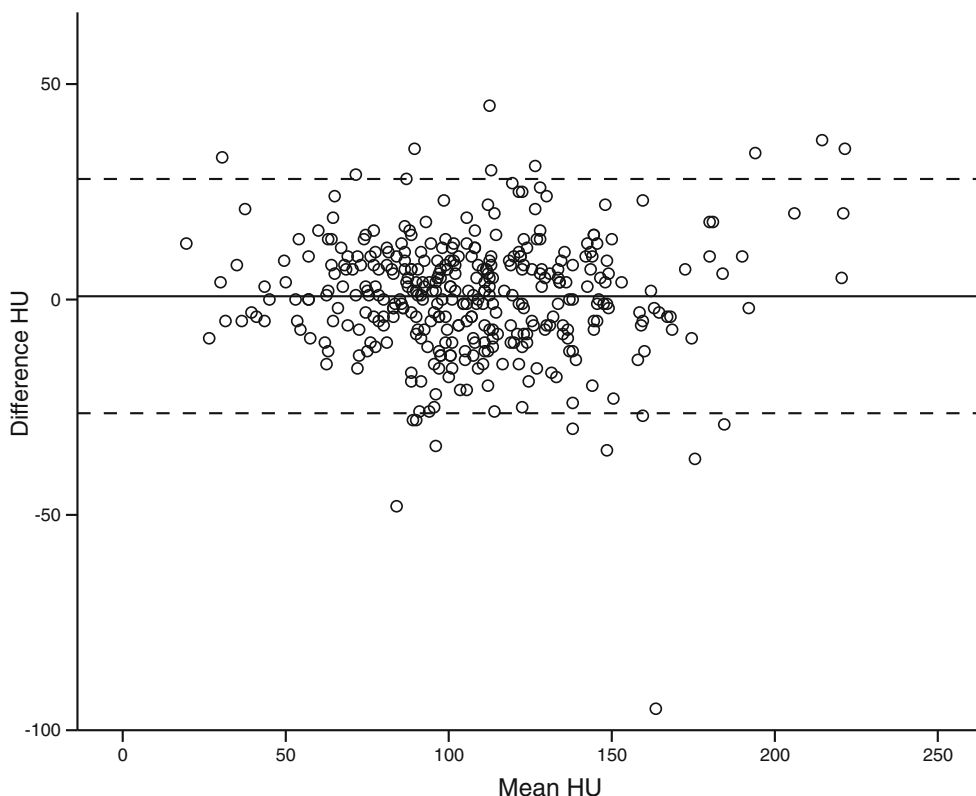
When using the same threshold of 110 HU to define osteoporosis in 77 participants with inter-observer measurements at L1, all six observers agreed about the diagnosis (osteoporosis yes/no) in 60 (77.9 %) participants and at least one observer did not agree in 17 (22.1 %) participants.

**Discussion**

In this study, we found excellent inter-examination reliability for manual bone density measurements of the vertebrae. Limits of agreement ranged from -26 to 28 HU, which means a change of at least 28 HU is needed in order to detect a real change in bone attenuation. Therefore, these results have to be taken into account when planning to use bone density measurements for longitudinal studies (e.g., for measuring therapeutic effects). Inter-observer reliability was good to excellent and limits of agreement with the mean ranged from -12 to 12 HU, which indicates that observers can be discordant with the mean estimated bone attenuation by 12 HU.

Our results imply that manual placement of a ROI in L1 is a reliable method for the quantification of vertebral attenuation. Therefore, in a lung cancer screening setting, low-dose chest CTs may be used to measure bone attenuation. Because these

**Fig. 3** Bland-Altman plot displays the inter-examination differences in vertebral Hounsfield unit (HU) measurements. Agreement is shown for 367 participants. The mean of differences (solid horizontal line) was 1 HU. The upper dashed line shows the upper 95 % limit of agreement (28 HU), and the bottom dashed line shows the lower 95 % limit of agreement (-26 HU)



measurements are performed manually, in theory, experience could influence the precision of the measurement. However, the present study shows that radiological experience has no major effect on attenuation measurements. Moreover, ICCs between more experienced observers were not better than between less experienced observers. Low-dose CT scans could therefore gain a role in early detection of osteoporosis.

With the recent recommendation on the implementation of lung cancer screening [24], a large number of subjects will receive a low-dose chest CT. Next to screening for lung cancer, this can provide an opportunity for the assessment of other abnormalities, such as chronic obstructive pulmonary disease and coronary artery calcifications [25]. Because smoking is associated with lower bone density [26], this could be an

opportunity for the detection of osteoporosis in this smoking population. By diagnosing low bone density as well, this could improve the yield and cost-effectiveness of lung cancer screening.

To our knowledge, this is the first study to describe the inter-examination agreement and reliability of attenuation measurements of the vertebrae on unenhanced low-dose CT in a large population. In addition, we extensively studied inter-observer agreement and reliability. Although several studies used attenuation measurements in the search for an appropriate screening tool for osteoporosis, studies on the agreement and reliability are lacking.

Ohara et al. [9] studied the correlation between pulmonary emphysema and reduced bone density. For this purpose, they used manual vertebral bone measurements. They validated their measurements by calculating correlation coefficients of two observers. This resulted in ICCs of 0.995, 0.993, 0.950 and 0.996 for T4, T7, T10 and L1, respectively. Their strength was the evaluation of multiple vertebral levels, but they concluded that the average bone density of three thoracic vertebral bones was highly correlated with bone density in L1 alone ( $r=0.914, p<0.001$ ). Pickhardt et al. [7] elaborated on this and found that measurements at L1 are as or more accurate than the results at other levels, including multilevel assessment. Also, Romme et al. [27] showed no added value of using three thoracic vertebral levels to assess bone density compared to one measurement at L1. Although L1 seems to provide the

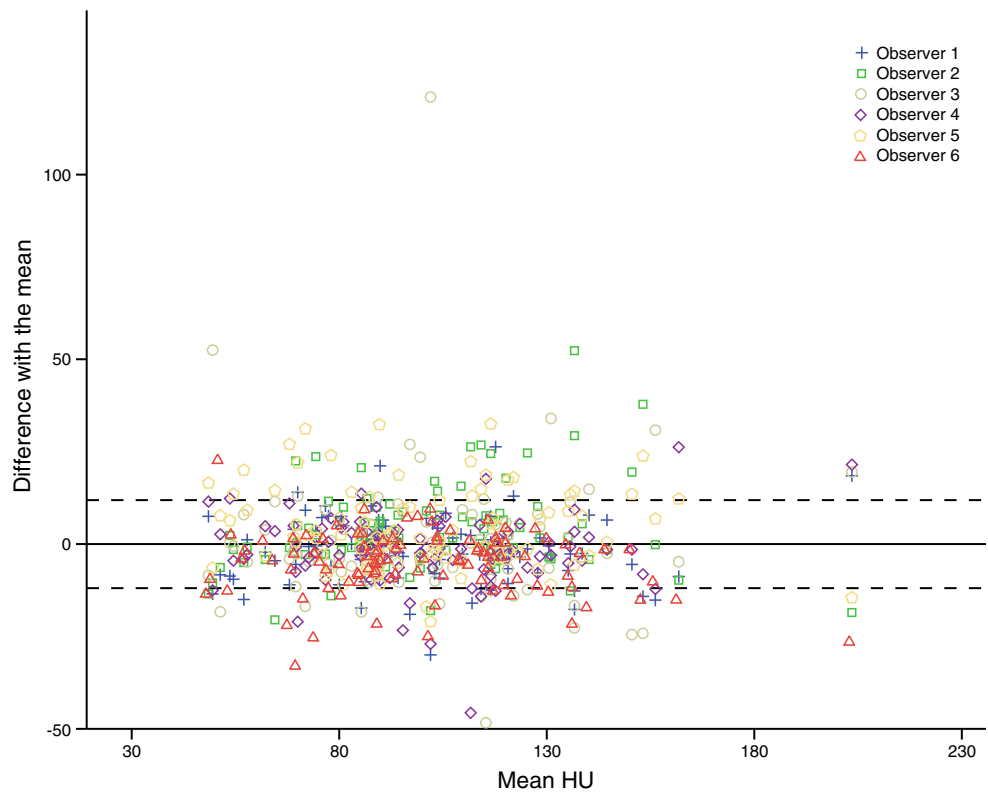
**Table 2** Inter-observer agreement for ROI measurements

	Observer 1				Observer 2	
Observer 2	0.87					
Observer 3	0.76	0.70				Observer 3
Observer 4	0.91	0.87	0.73			Observer 4
Observer 5	0.86	0.86	0.70	0.87		Observer 5
Observer 6	0.89	0.88	0.75	0.89	0.88	

Data given are intraclass correlation coefficients. All correlation coefficients were significantly correlated ( $p<0.001$ ).

ROI: region of interest

**Fig. 4** Inter-observer agreement plot between six observers for vertebral attenuation measurements. Observers represent different symbols. The differences of all measurements with the mean (y-axis) are plotted against the mean Hounsfield unit (HU) values for all participants (x-axis). The horizontal dashed lines indicate the limits of agreement with the mean of the six observers and ranged from -12 to 12 HU. This could indicate that the observers can be discordant with the mean BMD by as much as 12 HU



most accurate results in terms of attenuation measurements, this vertebral level is not always included on thoracic CT. In our population of 376 participants, 89 (23.7 %) measurements were made at a vertebral level different from L1.

Both Pickhardt et al. [17] and Romme et al. [27] studied inter-observer agreement and found limits of agreement between two observers of -6 HU to 16 HU for T12-L5 and -9 to 5 HU for T4-T7-T10, respectively. We complemented to this by using six observers with different experience levels and showed limits of agreement with the mean ranging from -12 to 12 HU. The intra-observer limits of agreement from Romme et al. ranged from -9 to 5 HU in 20 participants. Our limits of agreement were substantially wider, ranging from -26 to 28 HU, but could be more realistic as a result of a larger study cohort.

Next to presenting positive results in terms of agreement and reliability, it is important to estimate the impact of these results on clinical practice. In order to perform reclassification analyses, we used a threshold of 110 HU to define osteoporosis, which was derived from Pickhardt et al [7]. This threshold

was proposed for a routine care population with lower osteoporosis risk because of its high specificity. Buckens et al. [28] validated this threshold as being the most optimal as compared to DXA. By using this threshold, 159 (55.4 %) participants were classified as having osteoporosis. This high prevalence is in line with some findings of osteoporosis prevalence in a high-risk chronic obstructive pulmonary disease (COPD) cohort [29]. With this heavy smoking population being at risk for osteoporosis as well, these prevalence numbers could be appropriate. Another explanation for the high prevalence could be that the HU in the vertebra was systematically lower compared to the study by Pickhardt and Buckens.

Our reclassification analysis showed that inter-examination variability can lead to a different diagnosis in 11.2 % of included participants. Moreover, variability of inter-observer measurements can lead up to 22.1 % misclassified participants. As a consequence, when measuring bone density that is close to a threshold that defines disease, the effect of variability within a patient and between different observers could

**Table 3** Reclassification analysis according to inter-examination variability

		<110HU at follow-up		
		No (n)	Yes (n)	Total
<110HU at baseline	No (n)	110 (38.3 %)	18 (6.3 %)	128 (44.6 %)
	Yes (n)	14 (4.9 %)	145 (50.5 %)	159 (55.4 %)
Total		124 (43.2 %)	163 (56.8 %)	287 (100 %)

Reclassification of participants with a threshold of 110 HU according to inter-examination variability

be substantial. Considering the development of guidelines for osteoporosis screening, variability consequently has to be taken into account.

Our study has limitations. First, the follow-up scans for the assessment of inter-examination variability were performed three months after baseline. In this period, CT attenuation values could have altered. However, we think the impact will be limited because decline in bone density progresses slowly. Still, a follow-up CT examination directly after baseline would be more ideal to eliminate changes over time. Second, we only used measurements of one vertebra in our evaluation and did not include more vertebral levels. Nevertheless, one may assume that, even if bone attenuation may vary at each vertebral level, inter-observer and inter-examination agreement may be similar [9]. Thereby, former studies have shown that ROI placement in multiple vertebrae does not add value compared to one measurement at L1 [7, 27]. Third, as a consequence from the study design of the lung cancer screening trial, only a small amount of our cohort consisted of women. But, in this cohort, no difference was seen in inter-examination differences in HU between men and women. Lastly, although our scanners were calibrated weekly, we did not use a calibration phantom in this study as is done in QCT of the spine. We were therefore unable to provide BMD as milligrams hydroxyapatite per cubic centimetre and our method has lower precision compared to QCT [30, 31]. However, previous studies have shown that, although precision was lower compared to QCT, BMD estimation techniques without phantom calibration were nevertheless promising for assessing fracture risk [11].

In conclusion, this study shows that bone attenuation can be measured by manual ROI placement on unenhanced low-dose chest CT examinations with good reliability. However, when developing guidelines for early detection of osteoporosis, variability still has to be taken into account. While the distinctive character of this technique is excellent, diagnostic studies are needed to confirm these results, to evaluate its accuracy and ultimately its cost-effectiveness.

**Acknowledgements** The scientific guarantor of this publication is Dr. P.A. de Jong. The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article. Prof. dr. Lammers reports grants from TiPharma and grants from EU during the conduct of the study. Drs. Pompe, Dr. P. de Jong, Prof.dr. Oudkerk, Drs. Takx, Drs. Eikendal, Drs. Willemink, Dr. Budde, W. de Jong and Dr. Mohamed Hoesein have nothing to disclose. No complex statistical methods were necessary for this paper. Institutional Review Board approval was obtained. Written informed consent was obtained from all subjects (patients) in this study. Some study subjects or cohorts have been previously reported; a list of publications with overlap in study participants can be found in Supplementary Material. Methodology: retrospective, observational, performed at one institution.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

1. Sánchez-Riera L, Carnahan E, Vos T et al (2014) The global burden attributable to low bone mineral density. *Ann Rheum Dis* 73:1635–1645
2. Wilson JM, Jungner YG (1968) Principles and practice of mass screening for disease. World Health Organization, Geneva
3. Altkom D, Cifu AS (2015) Screening for osteoporosis. *JAMA* 313:1467–1468
4. Lewiecki EM, Gordon CM, Baim S et al (2008) International society for clinical densitometry 2007 adult and pediatric official positions. *Bone* 43:1115–1121
5. Stone KL, Seeley DG, Lui L et al (2003) BMD at multiple sites and risk of fracture of multiple types: long-term results from the study of osteoporotic fractures. *J Bone Miner Res* 18:1947–1954
6. Siris ES, Miller PD, Barrett-Connor E et al (2001) Identification and fracture outcomes of undiagnosed low bone mineral density in postmenopausal women: results from the National Osteoporosis Risk Assessment. *JAMA* 286:2815–2822
7. Pickhardt P, Pooler B, Lauder T, Muñoz del Rio A, Bruce RJ, Binkley N (2013) Opportunistic screening for osteoporosis using abdominal computed tomography scans obtained for other indications. *Ann Intern Med* 158:588–595
8. Romme EAPM, Murchison JT, Edwards LD et al (2013) CT-measured bone attenuation in patients with chronic obstructive pulmonary disease: relation to clinical features and outcomes. *J Bone Miner Res* 28:1369–1377
9. Ohara T, Hirai T, Muro S et al (2008) Relationship between pulmonary emphysema and osteoporosis assessed by CT in patients with COPD. *Chest* 134:1244–1249
10. Tay W-L, Chui C-K, Ong S-H, Ng AC-M (2012) Osteoporosis screening using areal bone mineral density estimation from diagnostic CT images. *Acad Radiol* 19:1273–1282
11. Miyabara Y, Holmes D, Camp J, Miller VM, Kearns AE (2012) Comparison of calibrated and uncalibrated bone mineral density by CT to DEXA in menopausal women. *Climacteric* 15:374–381
12. Schreiber JJ, Anderson PA, Rosas HG, Buchholz AL, Au AG (2011) Hounsfield units for assessing bone mineral density and strength: a tool for osteoporosis management. *J Bone Joint Surg Am* 93:1057–1063
13. Pompe E, Willemink MJ, Dijkhuis GR et al (2014) Intravenous contrast injection significantly affects bone mineral density measured on CT. *Eur Radiol* 25:283–289
14. Budoff MJ, Hamirani YS, Gao YL et al (2010) Measurement of thoracic bone mineral density with quantitative CT. *Radiology* 257:434–440
15. Buckens CF, van der Graaf Y, Verkooijen HM et al (2015) Osteoporosis markers on low-dose lung cancer screening chest computed tomography scans predict all-cause mortality. *Eur Radiol* 25:132–139
16. Steiger P, Block JE, Steiger S et al (1990) Spinal bone mineral density measured with quantitative CT: effect of region of interest, vertebral level, and technique. *Radiology* 175:537–543
17. Pickhardt PJ, Lee LJ, Muñoz del Rio A et al (2011) Simultaneous screening for osteoporosis at CT colonography: bone mineral density assessment using MDCT attenuation techniques compared

- against the DXA reference standard. *J Bone Miner Res* 26:2194–2203
18. Mets OM, Buckens CFM, Zanen P et al (2011) Identification of chronic obstructive pulmonary disease in lung cancer screening computed tomographic scans. *JAMA* 306:1775–1781
  19. Van Iersel CA, de Koning HJ, Draisma G et al (2007) Risk-based selection from the general population in a screening trial: selection criteria, recruitment and power for the Dutch-Belgian randomised lung cancer multi-slice CT screening trial (NELSON). *Int J Cancer* 120:868–874
  20. Mohamed Hoesein FAA, de Hoop B, Zanen P et al (2011) CT-quantified emphysema in male heavy smokers: association with lung function decline. *Thorax* 66:782–787
  21. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1:307–310
  22. Jones M, Dobson A, O'Brian S (2011) A graphical method for assessing agreement with the mean between multiple observers using continuous measures. *Int J Epidemiol* 40:1308–1313
  23. Kottner J, Audigé L, Brorson S et al (2011) Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *J Clin Epidemiol* 64:96–106
  24. Humphrey LL, Deffebach M, Pappas M et al (2013) Screening for lung cancer with low-dose computed tomography: a systematic review to update the US Preventive services task force recommendation. *Ann Intern Med* 159:411–420
  25. Mets OM, de Jong PA, Prokop M (2012) Computed tomographic screening for lung cancer: an opportunity to evaluate other diseases. *JAMA* 308:1433–1434
  26. De Jong WU, de Jong PA, Vliegenthart R et al (2014) Association of COPD and smoking status with bone density and vertebral fractures in male lung cancer screening participants. *J Bone Miner Res* 29:2224–2229
  27. Romme EAPM, Murchison JT, Phang KF et al (2012) Bone attenuation on routine chest CT correlates with bone mineral density on DXA in patients with COPD. *J Bone Miner Res* 27:2338–2343
  28. Buckens CF, Dijkhuis G, de Keizer B, Verhaar HJ, de Jong PA (2015) Opportunistic screening for osteoporosis on routine computed tomography? An external validation study. *Eur Radiol* 25:2074–2079
  29. Graat-Verboom L, van den Borne BEEM, Smeenk FWJM, Spruit MA, Wouters EFM (2011) Osteoporosis in COPD outpatients based on bone mineral density and vertebral fractures. *JBMR* 26: 561–568
  30. Link TM (2012) Osteoporosis imaging: state of the art and advanced imaging. *Radiology* 263:3–17
  31. Issever AS, Kentenich M, Köhlitz T, Diederichs G, Zimmermann E (2013) Osteoporosis and atherosclerosis: a post-mortem MDCT study of an elderly cohort. *Eur Radiol* 23:2823–2829