Intra- and inter-observer agreement and reliability of bone mineral density measurements around acetabular cup
a porcine ex-vivo study using single- and dual-energy computed tomography

Mussmann, Bo Redder; Overgaard, Soren; Torfing, Trine; Bøgehøj, Morten Foged; Gerke, Oke; Andersen, Poul Erik

Published in:
Acta Radiologica Open

DOI:
10.1177/2058460117719746

Publication date:
2017

Document version
Publisher's PDF, also known as Version of record

Document license
CC BY-NC

Citation for published version (APA):
Intra- and inter-observer agreement and reliability of bone mineral density measurements around acetabular cup: a porcine ex-vivo study using single- and dual-energy computed tomography

Bo Mussmann¹,², Søren Overgaard²,³, Trine Torfing¹,², Morten Bøgehøj²,³, Oke Gerke⁴,⁵ and Poul Erik Andersen¹,²

Abstract

Background: Periprosthetic bone loss is considered to be a potentially contributing factor in aseptic loosening of acetabular hip components, but no studies have shown this association. The lack of association might be caused by insufficient image quality because of metal artifacts and challenges in measuring bone density (BMD) in complex anatomic structures which might be overcome using dual-energy computed tomography (DECT).

Purpose: To test inter- and intra-observer agreement and reliability of in-house segmentation software measuring BMD adjacent to acetabular cup and to compare measurements performed with single-energy CT (SECT) and DECT in cemented and cementless cups.

Material and Methods: Twenty-four acetabular cups inserted in porcine hip specimens were scanned with SECT and DECT. Bone density was measured in a three-dimensional volume adjacent to the cup. Double measurements were performed.

Results: BMD derived from SECT was approximately four times higher than that of DECT. In both scan modes, intraclass correlation coefficient (ICC) was >0.90 with no differences between repeated measurements, except for uncemented cups where a statistically significant difference of 11 mg/cm³ was found with DECT. DECT showed narrower limits of agreement than SECT. Inter-observer analysis showed small differences.

Conclusion: BMD can be estimated with high intra- and inter-observer reliability with SECT and DECT around acetabular cups using custom software. The intra- and inter-observer agreement of DECT is superior to that of SECT and better in the cementless concept. Good intra- and inter-observer reliability can be obtained in both cemented and cementless cups using the segmentation software. SECT and DECT cannot be used interchangeably.

Keywords

Segmentation, computed tomography (CT), dual-energy CT, bone mineral density, bone loss, hip arthroplasty

Date received: 3 April 2017; accepted: 17 June 2017
Introduction

Aseptic loosening of orthopedic implants is the most common cause of failure after total hip arthroplasty (THA) (1,2) and periprosthetic bone loss is considered a predictor of aseptic loosening (3,4). To our knowledge, no studies have shown this association, but a recent study found osteoporotic patients with low systemic bone mineral density (BMD) to be at higher risk of cup migration compared with patients with normal BMD (5). Bone loss may occur around THA because of stress shielding and may be located around both the cup and stem (4,6). Most attention has been paid to the femoral side (4), but cementless press-fitted cups loading the acetabular rim may also cause stress shielding of the central iliac bone (7). Bone loss occurs in all compartments adjacent to the cup (8,9), but in cementless THA it occurs most often in the supra-acetabular ileum (10) and often also in the medial wall in both cemented and cementless cups (11). In cemented cup systems, bone loss seems to be less substantial (12). Bone loss must be considerable, i.e. 30–50%, before it can be detected on radiographs (13) and furthermore the sensitivity of radiographs is low which might explain the lack of correlation between bone loss and loosening of the prosthesis.

Dual X-ray absorptiometry (DXA) is widely used for bone density measurements in osteoporosis, but it is of limited value in the presence of metal because it acquires two-dimensional (2D) images only and it cannot obtain reliable measurements in the presence of bone cement (14). Bone density changes adjacent to acetabular hip implants have been studied with computed tomography (CT) (12,15–18), but CT is challenged in visualizing the interface between bone and implant due to metal artifacts (14), thus none of the studies have focused solely on the interface. Dual-energy CT (DECT) can theoretically remove or reduce beam hardening artifacts (19–21) and has shown promising results in bone density evaluation adjacent to spinal implants (22,23) and better delineation of the acetabular bone-implant interface in THA (24). In general, measuring BMD using CT in the acetabular region is also difficult because of the complex three-dimensional (3D) anatomic structure of the bone.

The primary objective of the present study was to test the intra- and inter-observer agreement and reliability of an in-house 3D segmentation software solution that measures BMD in close proximity to the acetabular cup with the use of volumetric CT images. The secondary objective was to compare BMD measurements performed with single-energy CT (SECT) and DECT in cemented and cementless prosthetic acetabular cups. If measurements can be performed with sufficiently high intra- and inter-observer reliability, the method might contribute to further research in bone loss and potentially impact the choice of acetabular cup in future patients.

Material and Methods

Twenty-four fresh frozen female porcine hemi-pelvic specimens were included. The animals were skeletally mature sows with closed triradiate cartilage. The exact age of the animals was unknown, but usually they are slaughtered when they reach 2–4 years. One orthopedic surgeon with more than ten years of experience in hip replacement surgery inserted 12 cementless and 12 cemented cups in the specimens. The specimens were stored at −25°C between surgery and the imaging procedure.

Surgical procedure

The specimens were partly thawed at room temperature for approximately 3 h prior to surgery to ensure that surgery was practically feasible yet avoiding the possibility of decomposition of the specimens. Surgery was performed with Stryker components and cement (Stryker Corp., Kalamazoo, MI, USA). In the cementless concept 48-mm Trident® hemispherical acetabular shells were used in combination with Triden® X3® polyethylene liners. The acetabulum was reamed line-to-line or underreamed according to the surgeon’s judgment. In the cemented concept, we used 48-mm Contemporary® Hooded cups. The acetabulum was over-reamed by 2 mm according to the manufacturer’s recommendations. Antibiotic Simplex® Bone Cement was used. In both concepts, an Exeter stem was placed in the cup for SECT and DECT examination. In the cementless concept, a 32-mm steel head was used and in the cemented concept a 28-mm head was used. The divergence was due to the relatively small porcine acetabulum, i.e. approximately 50 mm. As the cementless concept demands over-reaming by 2 mm, 48-mm cups were used. However, this cup size can only contain a 28-mm head. The stem was not fitted in the femoral bone because the femoral side was beyond the scope of the current study.

CT scanning

All scans were performed with a GE Discovery CT750 HD 64-channel scanner (GE Healthcare, Milwaukee, WI, USA). DECT and SECT were performed with the parameters shown in Table 1, and synthetic monochromatic DECT images were reconstructed at 130 keV. The keV level was chosen because energy levels above 110 keV provide stable density measurements (23) with an optimum at 130 keV in presence of...
orthopedic metal devices (25,26). Double measure-
ments were achieved by dismantling and rebuilding
the experimental setup and repeating the scans. Each
specimen was positioned on a vacuum bag and a
calibration phantom (MINDWAYS®/C213
QCT Pro,
Mindways Software Inc., Austin, TX, USA) was
placed beneath the specimen. The phantom contains
five calibration rods with known densities. The mean
pixel value in each rod was measured and BMD in any
region of interest (ROI) could be derived (Fig. 1).

Quantitative image analysis
We used a custom in-house 3D Fiji plugin to perform
the image analysis. Fiji is a platform for biological
image analysis based on ImageJ (27). Using the
plugin we could draw free-hand ROIs slice by slice
and position circular ROIs in the phantom calibration
rods (Fig. 1). The plugin stacked the ROIs and seg-
mented a cup-shaped 3D ROI based on a user-defined
diameter and thickness, and the ROI was divided into
two quadrants (Fig. 2). In this study, a combination of
50-mm diameter and 5-mm thickness was used. The
plugin created a comma-separated file with the pixel
values of the ROIs. A Python script (Python Software
Foundation, Beaverton, OR, US) was then used to cal-
culate the mean pixel value of each ROI using a lower
pixel value threshold of –300 Hounsfield units to
exclude air deposits. Image analysis was performed by
a radiographer with more than 20 years of experience in
CT and image analysis (BM). In the cementless con-
cept, the analysis was repeated by a second observer
(KK) with no radiologic experience in order to assess
inter-observer agreement and reliability.

Table 1. Acquisition parameters for SECT and DECT.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Single-energy CT</th>
<th>Dual-energy CT</th>
</tr>
</thead>
<tbody>
<tr>
<td>kVp</td>
<td>120</td>
<td>80/140 dual</td>
</tr>
<tr>
<td>Tube current (mA)</td>
<td>300</td>
<td>375</td>
</tr>
<tr>
<td>Scan time (s)</td>
<td>0.5</td>
<td>0.6</td>
</tr>
<tr>
<td>Pitch</td>
<td>0.984:1</td>
<td>0.984:1</td>
</tr>
<tr>
<td>Collimation (mm)</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>Image acquisition (mm)</td>
<td>$64 \times 0.625$</td>
<td>$64 \times 0.625$</td>
</tr>
<tr>
<td>Scan field of view</td>
<td>Body large (50 cm)</td>
<td>Body large (50 cm)</td>
</tr>
<tr>
<td>Kernel</td>
<td>“Bone”</td>
<td>“Detail”</td>
</tr>
<tr>
<td>ASIR</td>
<td>30%</td>
<td>Not available</td>
</tr>
<tr>
<td>CTDI$_{vol}$ (mGy)</td>
<td>11.35</td>
<td>12.92</td>
</tr>
</tbody>
</table>

ASIR, adaptive statistical iterative reconstruction; CTDI$_{vol}$, volumetric CT dose index.

Fig. 1. Example of an axial DECT slice with the hemispherical
volume defined by the pixels included in the intersection
between the circles (red) and the borders of the free-hand drawn
area (blue). Underneath the specimen, five ROIs (yellow) are
positioned in the calibration phantom rods.

Statistical analysis
According to Kress et al. (16), we used a minimal rele-
vant difference of 5%. Based on a sample size calcula-
tion using an estimated SD of 15, a difference of 15 mg/
$\text{cm}^3$ (i.e. 5%) could be detected by the inclusion of ten
specimens in each group. All variables were continuous
and summarized by mean, number of observations, and
95% confidence intervals (CI). The differences between
repeated BMD measurements were estimated with mixed effects regression modeling, using measurement
as fixed effect. The absolute agreement between the
methods and between the observers was analyzed with
Bland–Altman plots (28), including mean difference
and limits of agreement. The intra- and inter-observer
reliability was assessed by intraclass correlation coeffi-
cients (ICC) based on two-way random effects models
(29,30). The statistical significance of the difference
between SECT- and DECT-based BMD measure-
ments was assessed by comparing the 95% CIs. Repeatability coefficients (RC) were calculated for the
repeated measurements according to Bartlett and Frost
(30). RC is an estimate below which the absolute dif-
ference would be expected to lie with 95% certainty
(30–32). To compare RC and the minimal clinically
relevant difference, the RCs were converted to percent-
ages by dividing RC with the mean BMD of measure-
ment 1. All analyses were performed using STATA/SE
14.0 (StataCorp. LP, College Station, TX, USA). The
study is reported in accordance with the Guidelines
for Reporting Reliability and Agreement Studies
(GRRAS) (33).
Fig. 2. Example of a ROI divided into quadrants superimposed onto axial and coronal slices in SECT (left) and DECT (right). Beam hardening artifacts are present in both scan modes, but appear slightly more prominent in SECT (arrow).

Table 2. Intra-observer analysis. Differences between repeated BMD measurements derived from SECT and DECT (units of mg K$_2$HPO$_4$/cm$^3$).

<table>
<thead>
<tr>
<th>Scan mode</th>
<th>Volume of interest</th>
<th>Mean BMD measurement 1</th>
<th>Mean BMD measurement 2</th>
<th>Difference</th>
<th>95% CI</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-energy CT</td>
<td>Cementless cups, n = 10</td>
<td>1188</td>
<td>1180</td>
<td>−8</td>
<td>−40 to 24</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>Superior quadrant</td>
<td>1351</td>
<td>1349</td>
<td>−2</td>
<td>−53 to 49</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>Inferior quadrant</td>
<td>1187</td>
<td>1211</td>
<td>24</td>
<td>−43 to 92</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>Anterior quadrant</td>
<td>1270</td>
<td>1220</td>
<td>−50</td>
<td>−93 to −8</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>Posterior quadrant</td>
<td>598</td>
<td>657</td>
<td>59</td>
<td>1−116</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>Cemented cups, n = 12</td>
<td>1282</td>
<td>1249</td>
<td>−33</td>
<td>−67 to 1</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>Superior quadrant</td>
<td>1334</td>
<td>1304</td>
<td>−30</td>
<td>−66 to 6</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>Inferior quadrant</td>
<td>1382</td>
<td>1300</td>
<td>−82</td>
<td>−160 to −5</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>Anterior quadrant</td>
<td>1476</td>
<td>1442</td>
<td>−34</td>
<td>−88 to 21</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>Posterior quadrant</td>
<td>962</td>
<td>921</td>
<td>−41</td>
<td>−106 to 24</td>
<td>0.21</td>
</tr>
<tr>
<td>Dual-energy CT</td>
<td>Cementless cups, n = 10</td>
<td>295</td>
<td>297</td>
<td>2</td>
<td>−6 to 10</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>Superior quadrant</td>
<td>328</td>
<td>328</td>
<td>0</td>
<td>−7 to 8</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>Inferior quadrant</td>
<td>276</td>
<td>285</td>
<td>9</td>
<td>−11 to 30</td>
<td>0.38</td>
</tr>
</tbody>
</table>
Results

Two specimens with cementless cups were excluded because of fractures which had occurred during the surgical procedure, but remained undiscovered until imaging had been performed. The repeated BMD measurements in the “all quadrants” ROI showed no statistically significant differences in the cementless cups. In the cemented concept, we found no significant difference in SECT, while DECT showed a difference of...
11 mg/cm³ ($P = 0.014$) between the measurements (Table 2). The Bland–Altman plots (Figs. 3 and 4) showed narrower limits of agreement in DECT (–21 to 44) compared with SECT (–86 to 153) and the RCs (Table 3) were significantly smaller than those of SECT. ICC and the corresponding 95% CI in SECT was 0.90 (95% CI = 0.67–0.98) and 0.91 (95% CI = 0.70–0.97) for the cementless and cemented concept, respectively, while ICC in DECT was 0.91 (95% CI = 0.69–0.98) for the cementless and 0.93 (95% CI = 0.68–0.98) for the cemented concept. The results of the inter-observer analysis are listed in Table 4 and plotted in Fig. 5, with corresponding Bland–Altman plots given in Fig. 6. The inter-observer ICC was 0.91 (95% CI = 0.47–0.98) and 0.93 (95% CI = 0.31–0.99) for SECT and DECT, respectively.

**Discussion**

In this experimental ex-vivo study with porcine specimens, we found that BMD estimates derived from SECT were approximately four times higher than those of DECT using segmentation software. The difference is supposedly caused by differences in the estimation of BMD, i.e. the pixel values in the low-density rods of the calibration phantom were comparable between SECT and DECT, while the pixel values at higher densities differed. Because the BMD estimation is basically a regression maneuver, the difference will cause SECT to result in higher BMD than DECT (34).

In the cemented concept, we found a statistically significant difference between the repeated measurements in DECT and a borderline significant difference in SECT. The difference is probably caused by the relatively low number of specimens. In all cases, the reliability expressed by ICC was good and no statistically significant difference was found between SECT and DECT.

In both scan modes, the RCs were larger than the minimal clinically relevant difference of 5%, i.e. 8% in the cementless concept and 11% in the cemented concept.

Measurements in the subdivided acetabular ROIs returned small differences between the repeated measurements and followed the same tendency between the scan modes. However, the results of this analysis must be considered explorative, because such sub-group analyses would supposedly require more specimens as reflected in the variability of the $P$ values.

The study has some limitations. The statistically significant difference between the two observers may be caused by the difference in experience between the observers potentially leading to a consensus issue. The choice of observers was made to create a worst-case scenario in order not to overestimate the reliability and agreement of the method. Even in this worst-case

---

**Table 3.** Repeatability coefficients and corresponding 95% CIs in units of mg K$_2$PO$_4$/cm$^3$ for repeated BMD measurements performed with SECT and DECT. The absolute difference between the measurements would be expected to lie below RC with 95% certainty.

<table>
<thead>
<tr>
<th>Volume of interest</th>
<th>$R_{SECT}$</th>
<th>$R_{DECT}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cementless cups, n = 10</td>
<td>101 (64–160)</td>
<td>25 (16–40)</td>
</tr>
<tr>
<td>Cemented cups, n = 12</td>
<td>117 (77–178)</td>
<td>32 (21–48)</td>
</tr>
</tbody>
</table>

---

**Table 4.** Inter-observer analysis. Differences between average BMD measurements in cementless cups for observers 1 and 2 in SECT and DECT (units of mg K$_2$HPO$_4$/cm$^3$). $n = 10$.

<table>
<thead>
<tr>
<th>Volume of interest</th>
<th>Mean BMD observer 1</th>
<th>Mean BMD observer 2</th>
<th>Difference</th>
<th>95% CI</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Single-energy CT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All quadrants</td>
<td>1188</td>
<td>1146</td>
<td>–42</td>
<td>–70 to –13</td>
<td>0.005</td>
</tr>
<tr>
<td>Superior quadrant</td>
<td>1351</td>
<td>1339</td>
<td>–12</td>
<td>–56 to 32</td>
<td>0.59</td>
</tr>
<tr>
<td>Inferior quadrant</td>
<td>1187</td>
<td>1047</td>
<td>–140</td>
<td>–193 to –86</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Anterior quadrant</td>
<td>1270</td>
<td>1235</td>
<td>–35</td>
<td>–62 to –9</td>
<td>0.009</td>
</tr>
<tr>
<td>Posterior quadrant</td>
<td>598</td>
<td>620</td>
<td>22</td>
<td>–66 to 110</td>
<td>0.63</td>
</tr>
<tr>
<td><strong>Dual-energy CT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All quadrants</td>
<td>295</td>
<td>285</td>
<td>–10</td>
<td>–14 to –5</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Superior quadrant</td>
<td>328</td>
<td>324</td>
<td>–4</td>
<td>–10 to 1</td>
<td>0.14</td>
</tr>
<tr>
<td>Inferior quadrant</td>
<td>276</td>
<td>244</td>
<td>–32</td>
<td>–41 to –22</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Anterior quadrant</td>
<td>336</td>
<td>331</td>
<td>–5</td>
<td>–11 to 1</td>
<td>0.09</td>
</tr>
<tr>
<td>Posterior quadrant</td>
<td>150</td>
<td>141</td>
<td>–9</td>
<td>–22 to 4</td>
<td>0.19</td>
</tr>
</tbody>
</table>
scenario the absolute differences were small and most likely not clinically relevant, but experienced observers should perform all analyses in clinical studies. Further steps should be taken to ensure consensus if the method is to be used with multiple observers.

Because the study solely focused on agreement between repeated measurements and not accuracy of the BMD estimates, the external validity of the study is limited. The implants were inserted in dead porcine specimens. Thus, no biochemical reactions from the implants could occur and the cement did not infiltrate the bone tissue as much as it would do in living subjects. Furthermore, the animals were young compared with hip patients who are most often elderly with degenerative changes and lower BMD. The anatomy of porcine hips differs somewhat from human anatomy and even though the specimens were kept at low temperature there were small air deposits in the cancellous bone. We also considered positioning the specimens in a water tank to mimic the absorption of human tissue, but it was not possible to obtain stable positioning without movement during the scan procedure. Thus, more image noise would be expected in a patient compared to the experimental setup. However, the image quality of both scan modes would supposedly decrease similarly from this. Finally, the cemented cups could not contain stem heads larger than 28 mm compared with the commonly used 32-mm heads. This may in principle have caused fewer artifacts in the cemented concept.

In conclusion, the present study suggests that the intra-observer agreement of acetabular BMD measurements performed with DECT is better than that of SECT, while the intra-observer reliability is equally high for both scan modes and cup types. The inter-observer agreement was better in DECT while the inter-observer reliability was equally high in SECT and DECT. However, due to the difference in experience between the observers the results on inter-observer reliability and agreement must be interpreted with caution. Thus, DECT may be beneficial as a research tool for longitudinal studies of bone loss around acetabular cups. Further experimental in vivo studies are needed before the software can be used in longitudinal clinical studies of bone loss.

Acknowledgements
We would like to thank specialist radiographer Peter Traise, Bathurst, Australia who worked at our department for his assistance during the scans and data acquisition procedure.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Fig. 5. Box plot with medians, quartiles, and ranges comparing BMD in units of mg K2PO4/cm3 for two observers (BM and KK) in the cementless concept. n = 10. Dots are outliers.

Fig. 6. Bland–Altman plots of density measurements from observers 1 and 2 derived from SECT (left) and DECT (right). Horizontal lines indicate limits of agreement and the mean difference between the observers. n = 10.
Funding
The author(s) received no financial support for the research, authorship, and/or publication of this article.

References


