



**TOWARDS  
CLINICAL IMPLEMENTATION  
OF FINITE ELEMENT BASED  
FRACTURE RISK PREDICTION IN  
FEMORAL BONE METASTASES**



Florieke Eggermont

# **Towards clinical implementation of finite element based fracture risk prediction in femoral bone metastases**

**Florieke Eggermont**

The work presented in this thesis was performed at the Orthopaedic Research Laboratory of the Radboud university medical center, Nijmegen and was carried out within the Radboud Institute for Health Sciences.

**ISBN:**

978-94-028-1784-3

**Cover design:**

Florieke Eggermont

**Layout:**

Bregje Jaspers, ProefschriftOntwerp.nl

**Printing:**

Ipskamp Printing, Enschede

**© Florieke Eggermont. Nijmegen 2019.**

All rights reserved. No part of this thesis may be reproduced in any form without written permission of the author.

# **Towards clinical implementation of finite element based fracture risk prediction in femoral bone metastases**

## **Proefschrift**

ter verkrijging van de graad van doctor  
aan de Radboud Universiteit Nijmegen  
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,  
volgens besluit van het college van decanen  
in het openbaar te verdedigen op woensdag 18 december 2019  
om 14:30 uur precies

door

**Florieke Eugenie Eggermont**

geboren op 5 februari 1989  
te Nijmegen



**Promotor**

Prof. dr. ir. N. J. J. Verdonschot

**Copromotoren**

Dr. E.J.M. Tanck

Dr. Y.M. van der Linden (Leiden Universitair Medisch Centrum)

**Manuscriptcommissie**

Prof. dr. M. Verheij

Prof. dr. H.M. Verkooijen (Universitair Medisch Centrum Utrecht)

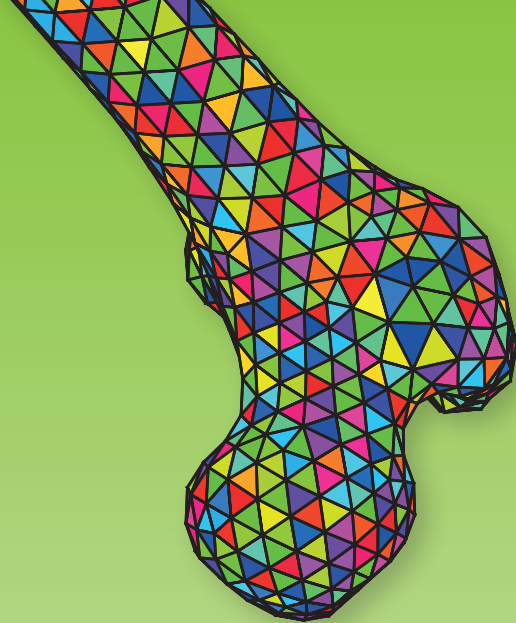
Prof. dr. ir. R.M. Verdaasdonk (Universiteit Twente)





# CONTENTS

<b>Chapter 1</b>	General introduction and outline	9
<b>Chapter 2</b>	Axial cortical involvement of metastatic lesions to identify impending femoral fractures; a clinical validation study	23
<b>Chapter 3</b>	Limited short-term effect of palliative radiation therapy on quantitative computed tomography-derived bone mineral density in femora with metastases	39
<b>Chapter 4</b>	Can patient-specific finite element models better predict fractures in metastatic bone disease than experienced clinicians? Towards introducing computational modelling into daily clinical practice	61
<b>Chapter 5</b>	The effect of different CT scanners, scan parameters and scanning setup on Hounsfield units and calibrated bone density: a phantom study	81
<b>Chapter 6</b>	Effect of different CT scanners and settings on femoral failure loads calculated by finite element models	107
<b>Chapter 7</b>	Calibration with or without phantom for fracture risk prediction in cancer patients with femoral bone metastases using CT-based finite element models	129
<b>Chapter 8</b>	Patient-specific finite element computer models improve fracture risk predictions in cancer patients with femoral bone metastases compared to clinical guidelines	147
<b>Chapter 9</b>	General discussion and future perspectives	165
<b>Chapter 10</b>	Summary	185
<b>Chapter 11</b>	Samenvatting	193
<b>Chapter 12</b>	Dankwoord	203
	Curriculum Vitae	207
	List of publications	209
	Data management	213
	PhD Portfolio	215
<b>Appendix:</b>	Flowchart of patients in all studies	218



# CHAPTER 1

## GENERAL INTRODUCTION AND OUTLINE



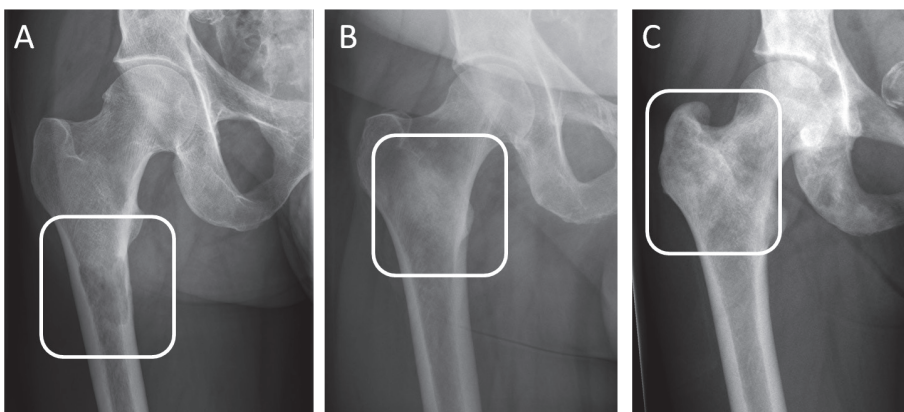


## Cancer

In 2017, more than 100.000 patients were diagnosed with cancer in the Netherlands and the incidence is still increasing each year.<sup>1</sup> Worldwide, the same trend is seen, with an estimation of 14.1 million new cases of cancer in 2012 and a predicted incidence of 21.6 million for 2030 and even 24.0 million for 2035.<sup>2</sup> Prostate and breast cancer are the most common types of cancer in the Netherlands, but also skin, colorectal, and lung cancer are often seen.<sup>1</sup> Since treatment options become better, patients tend to live longer with cancer. As a downside, more complications such as the development of metastatic disease are seen.<sup>3</sup>

## Bone metastases

Bone is, after lung and liver, the third most common tissue affected with metastases.<sup>4</sup> Many tumours, including breast, prostate, lung, kidney and thyroid, favour the skeleton for dissemination.<sup>5-7</sup> Normally, healthy bone undergoes a dynamic remodelling process existing of bone resorption by osteoclasts followed by bone formation by osteoblasts. However, in the case of metastases, this remodelling process is distorted. For bone metastases to arise, primary tumour cells must invade their surrounding healthy tissue. They then enter the small blood vessels and travel to the distant bone marrow cavity after which they generate their own blood supply to travel to the endosteal bone surface.<sup>7</sup> From there, the metastatic cancer cells excrete cytokines that can stimulate osteoclasts and osteoblasts. Pathways that stimulate bone resorption by osteoclasts lead to (osteo)lytic lesions, whereas (osteo)blastic or sclerotic lesions are caused by osteoblast proliferation, differentiation and bone formation (Figure 1). Additionally, metastases often are present in a mixed form, existing of both lytic and blastic processes. Exact pathways of the metastatic lesion development are different for each primary tumour type.<sup>7,8</sup> Furthermore, the incidence of bone metastases differs between the primary tumour types, but they are for example occurring in up to 70% of the advanced breast and prostate cancer patients.<sup>8</sup>



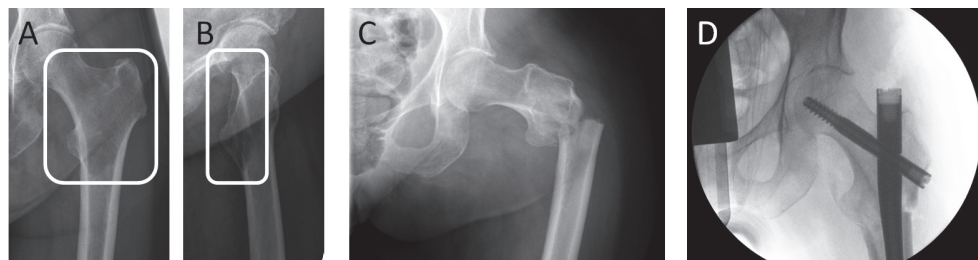
**Figure 1:** Lytic (A), blastic (B) and mixed (C) femoral bone metastases.



Bone metastases are often seen in the most vascularised parts of the skeleton, predominantly in the axial skeleton and proximal ends of the long bones, the skull, the ribs and the vertebrae.<sup>4-6</sup> Such bone metastases are often very painful and can cause complications such as impaired mobility, pathological fractures, hypercalcaemia, and in case of vertebral metastases, spinal cord compression.<sup>4-6</sup> Although in general survival of cancer patients improves due to better systemic treatment options, about half of the patients with bone metastases will be deceased within seven months after diagnosis, depending on their primary tumour.<sup>9</sup>

Approximately ten percent of the metastatic lesions occur in the femur.<sup>10</sup> About five to ten percent of all patients with femoral bone metastases will sustain an impending or actual fracture during the course of their disease.<sup>11,12</sup> Femoral pathologic fractures (Figure 2) usually occur during everyday activities, such as walking, turning in bed, rising from a chair, or stair climbing. In case of a pathological fracture in the femur, the patient's mobility and self-care is severely hampered, which obviously significantly affects the quality of life. Also, having a sudden pathological fracture evidently causes anxiety and stress to patients.

Surgical treatment of such fractures is complex, and survival of patients suffering from pathological fractures is decreased,<sup>13,14</sup> possibly due to an increased risk of deep vein thrombosis and further dissemination of the tumour as a result of a damaged microcirculation.<sup>14</sup> In addition, if patients are not fit enough, or unwilling to undergo surgery, they are at risk to become bedridden, and a cascade of negative physical consequences follow, such as lung and bladder infections and skin decubitus, leading to earlier death. Hence, it is important to prevent femoral fractures in patients with bone metastases.



**Figure 2:** Anteroposterior (A) and lateral (B) radiographs of a patient with lytic bone metastases around the lesser trochanter. The patient suffered from a pathological fracture (C), which was treated with a gamma nail (D).

Treatment of femoral bone metastases is based on the fracture risk: patients with a high fracture risk are considered for preventive surgical stabilization, whereas patients with a low fracture risk, or with insufficient clinical condition to undergo surgery, will be treated conservatively, for example with radiotherapy.

Prophylactic stabilizing surgery of impending lesions, i.e. before the pathological fracture occurs, results in lower morbidity and mortality compared to surgery after a fracture.<sup>13,14</sup> Additionally, preventive stabilization is associated with shorter hospital stays and lower costs than post-fracture surgery.<sup>15</sup> Nonetheless, prophylactic surgery remains an invasive procedure that comes with a certain risk of complications, especially in patients with disseminated cancer that often have a poor general clinical condition and a limited life expectancy, and should therefore be carefully weighted. Consequently, if a patient is not surgically treated, conservative treatment may follow with the aim to relieve pain, for example with radiotherapy.

Overall, about 60% of the patients treated with radiotherapy will have some relief of pain, with approximately 25% having complete pain relief.<sup>16</sup> Usually, patients with smaller uncomplicated lesions and a low risk of fracture will be treated with a single dose of 8 Gy. Patients with more complicated lesions and a high fracture risk who are considered for surgical stabilization but have deteriorating clinical condition will receive radiotherapy in multiple fractions, for example 5 x 4 Gy or 10 x 3 Gy, with the goal that after about 3 months, the higher total radiation dose results in remineralization, restoration of bone stability, and prevention a pathological fracture.<sup>17-19</sup>

## Fracture risk prediction in current clinical practice

Since treatment of patients with bone metastases is dependent on their fracture risk, fracture risk assessment should be accurate. Currently, several methods are used to estimate fracture risk, usually based on conventional radiographs. Many risk factors have been evaluated, including lesion size,<sup>17,20,21</sup> cortical bone involvement,<sup>17,22-26</sup> increasing pain,<sup>17,20-24</sup> and radiographic appearance,<sup>17,20,21,23</sup> but none of these parameters have been shown to be a genuine powerful predictor of fracture. Worldwide, physicians often use the Harrington's criterion, which states that lytic lesions involving more than 50% of the diameter of the bone, greater than 2.5 cm in diameter, or associated with persistent pain or radiographic progression after radiation, are at risk of fracture.<sup>27</sup> This guideline, however, has never been properly validated. Another widely used method is Mirels' scoring system. This score runs from 4 to 12 and results from a combination of ratings of pain, and lesion type, size and location. Generally, a patient should be considered for surgery if the Mirels' score is 9 or higher.<sup>23</sup> However, although almost all pathological fractures are circumvented when using the Mirels' score, the method appears to lead to large numbers of unnecessary surgeries, i.e. overtreatment.<sup>17,26,28,29</sup> Additionally, it has been shown that agreement between different observers when applying the Mirels' scoring system is limited.<sup>30</sup>

Van der Linden *et al.* searched for a simple radiographic parameter to identify lesions at risk of pathological fracture in patients who received palliative radiotherapy for pain.<sup>17,18</sup> During follow-up, 14 of the 102 patients sustained a pathological fracture. They found that an axial cortical involvement of more than 30 mm was predictive in 86% of the fractured femurs, although there was also a substantial number of lesions with a cortical involvement over 30 mm that did not fracture (42%).<sup>17,18</sup> The 30 mm threshold was compared with Mirels' scoring system, and it was shown that few fractures would be

missed when using the 30 mm threshold (sensitivity of 86% versus 100% for 30 mm threshold versus Mirels', respectively), whereas more patients would be correctly assessed as low risk (58% versus 13%), thus decreasing overtreatment.<sup>17</sup> Additionally, the number of patients that needs to be surgically stabilized to avoid one pathological fracture would decrease from seven to four when using the 30 mm threshold instead of Mirels' scoring system.<sup>17</sup> Based on the studies by Van der Linden *et al.*,<sup>17,18</sup> the Dutch national guideline states that lesions larger than 30 mm should be surgically stabilized.<sup>11</sup> Nevertheless, using the currently available guidelines, it remains challenging to accurately predict fracture risk. We would like to improve fracture risk predictions. Additionally, for a considerable number of metastases, mainly permeative or diffuse lesions, it appears to be difficult to indicate the margins, which makes it impossible to accurately measure the dimensions of the lesions.<sup>22,26,31</sup> The main limitation of the above-mentioned tools for fracture risk assessment is that they oversimplify the problem. Fracture risk prediction based on two dimensional radiographic representations fails to consider for example location and shape of the lesion, general bone quality, or shape of the bone.<sup>32</sup>

## Finite element models to predict fracture risk

One of the methods that does take aspects, such as shape of the lesion, general bone quality, as well as forces and stresses within the affected bone, into account is finite element (FE) modelling. FE models are computer models that were initially used in the fields of aerospace and civil engineering to solve complex elasticity and structural problems. Currently, application of FE models is investigated in other fields as well, for instance in the field of biomechanics to calculate bone strength.

FE modelling has shown to be promising as a tool to predict fracture risk for patients with osteoporosis by many research groups (for example Bessho *et al.*,<sup>33</sup> Keyak *et al.*,<sup>34</sup> Zysset *et al.*<sup>35</sup>). The application of FE models for fracture risk prediction of metastatic femurs has been studied to a much lesser extent.<sup>36-40</sup> The first ones that attempted to predict strength of femurs with metastases by the use of FE models were Cheal *et al.*<sup>36</sup> They compared the strength of experimentally loaded femurs with simulated metastases with femoral FE models with comparable lesions. However, the FE model they used was based on one femur that approximated the average of their database, and thus their FE models were not specific for the femurs they used in the experiments, which might also explain the considerable underestimation of the FE strength. In addition, Keyak and colleagues<sup>37,41-44</sup> have put effort in developing and validating an FE model. They developed a model to describe the post-failure material behaviour of bone by empirically determining the relationship between CT densities and ash densities, and mechanical material properties subsequently. They validated their material model using cadaveric femurs with and without metastatic lesions and found good agreement between the measured and predicted individual femoral strength under axial loading ( $R^2 = 0.83$ ).<sup>37</sup> Moreover, Keyak *et al.*<sup>38</sup> showed that FE models were able to predict fracture loads of cadaveric femoral shafts with and without metastatic lesions in a four point bending experiment ( $R^2$  between 0.92 and 0.98), even when bones with metastases were FE modelled using mechanical property relationships for bone without

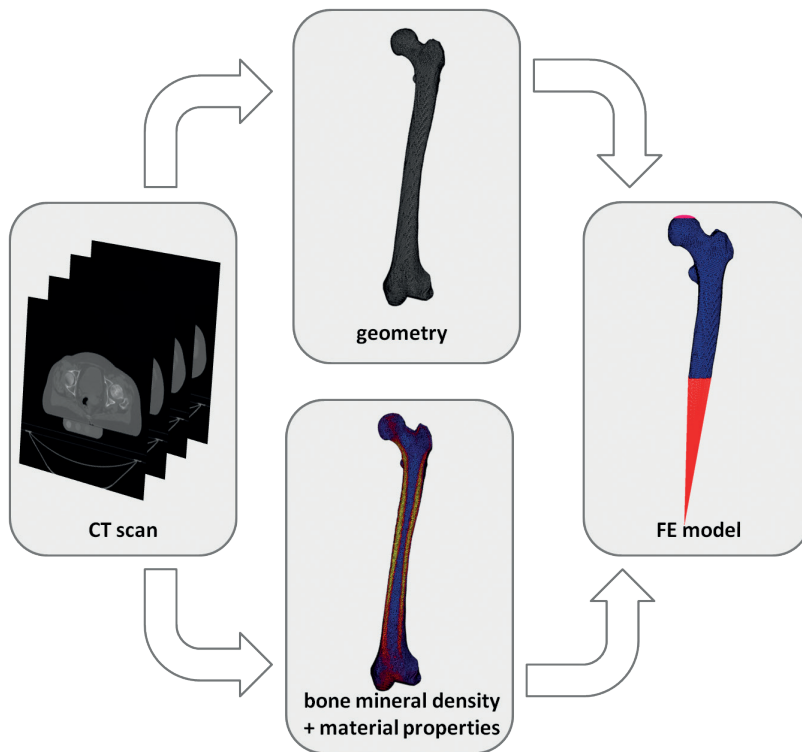
metastases. Another research group loaded femurs with simulated metastatic lesions until failure in a torsion experiment and replicated the experiments in FE models. They found that fracture locations were reasonably predicted by the FE model and the experimental moment of failure correlated with the stiffness predicted by the FE model ( $R^2 = 0.72$ ).<sup>39</sup> More recently, Yosibash *et al.*<sup>40</sup> used cadaveric femurs that were affected with actual metastases and axially loaded them until failure. They examined histopathologically whether the fractures occurred in a metastatic lesion. Additionally, they created FE models mimicking the mechanical experiments and found a good agreement between predicted and experimental yield load ( $R^2 = 0.78$ ).

## The ORL finite element model

During the past years, the development of a patient-specific FE model for fracture risk prediction in metastatic femurs has also been subject of research within the Orthopaedic Research Laboratory.<sup>45-48</sup> Computed Tomography (CT) functions as input to the FE models (Figure 3). From the CT scan, the geometry is obtained by segmentation, which is then converted into a 3D solid mesh consisting of tetrahedral elements. The femoral geometry is positioned in such a manner that it represents the body in a standing position: the hip joint centre is aligned with the knee joint centre.

To determine patient-specific material properties, the Hounsfield units (HU) in the CT scan need to be calibrated to calcium equivalent densities, which is a measure of bone mineral density (BMD). For this, the patient is scanned on top of a calibration phantom. This calibration phantom contains rods with known calcium equivalent densities which are used to obtain a calibration function by linearly fitting the known densities with the corresponding HU of the rods. Subsequently, the whole CT scan is calibrated to calcium equivalent densities by using the calibration function. These calcium equivalent densities are a measure of bone mineral densities (BMD) and are used to calculate ash densities and subsequent nonlinear isotropic mechanical bone properties specific for each tetrahedral element of the femoral geometry, based on equations empirically formulated by Keyak *et al.*<sup>37</sup> Keyak's material model describes post-failure behaviour by an initial perfectly plastic phase, followed by a strain softening phase and an indefinite perfectly plastic phase.<sup>37</sup> In short, the higher the calcium equivalent values, the higher the ash densities and the higher subsequent strength, stiffness and yield stress until plasticity is reached.

Next, the boundary conditions are applied to the FE model. The FE model is distally fixated at the knee joint centre by means of two bundles of high-stiffness springs. During the simulation, a load is applied to the femoral head via a cup that is incrementally displaced in axial direction. The main output of the FE simulation is the force-displacement curve, based on displacement of the cup and contact normal forces that are registered for each increment. From the force-displacement curve, we obtain the failure force, which is defined as the maximum total reaction force. Additionally, failure locations can be visually assessed by determining the elements that have plastically deformed at the moment of structural failure.



**Figure 3:** Process of patient-specific finite element (FE) model generation

Loes Derikx described the first steps of the development and validation of the FE model in her Doctoral thesis in 2015.<sup>48</sup> First of all, the FE models were experimentally validated.<sup>46,47</sup> For this, artificial metastatic lesions were created in cadaveric femurs by drilling defects. Femurs were placed in a water basin and CT scans were made, after which the femurs were axially loaded in a material testing machine until failure. FE models were constructed and failure loads were calculated. Subsequently, experimental failure loads were compared to simulated failure loads, and showed that the FE models accurately calculated failure load.<sup>46,47</sup> Additionally, fracture locations were predicted fairly well.<sup>46,47</sup> Moreover, experienced clinicians were asked to rank the femurs on strength, and it was shown that ranking of bone strength by the computer model outperformed the clinicians.<sup>46</sup> These results were very promising.

Nevertheless, the experimental test set up using cadaveric femurs is a highly simplified representation of the physiological circumstances. *In vivo*, metastases are not drilled artefacts with clear edges, but they consist of lytic or blastic tissue that possibly behave differently mechanically, and fractures do not limitedly occur under axial loading and muscle forces can play a role. Furthermore, in contrast to CT scans of femurs in water basins, *in vivo* scanned femurs are affected by surrounding material such

as bony structures and soft tissue that can cause for example X-ray scatter and beam hardening.<sup>49,50</sup> Therefore, the FE model had to be validated *in vivo* to verify its ability to assess fracture risk in patients. For this purpose, a multicenter patient cohort study (CT femur study, funded by the Dutch Science Foundation NWO-STW (NPG.06778)) was performed including 62 patients with painful femoral metastases from three radiotherapy institutes in the Netherlands. Patients underwent multiple CT scans at baseline and different time points after radiotherapy. Through their hospital records, the patients were prospectively followed for six months after inclusion to investigate which patients suffered from an unexpected pathological fracture. In Derikx's thesis,<sup>48</sup> preliminary results of a subgroup of 23 patients from one of the three institutes were published. Patient-specific FE models were created to calculate femoral failure load. The median failure load of the five fractured femurs was significantly lower compared to the median failure load of the non-fractured femurs. Additionally, the FE predictions were compared to fracture risk assessments by experienced clinicians, and it was shown that the FE model tended to be more accurate in identifying patients at risk of a fracture. These results showed the potential of the FE model, but the predictive power of the FE model should be confirmed in the complete patient study.

## Aim and outline of this thesis

The goal of this thesis was to further develop and validate the patient-specific FE model for prediction of fracture risk in patients with cancer and femoral bone metastases. Therefore, we validated the available finite element model with the use of real patient data. For this, we used two consecutive patient studies: the CT femur study (2006-2009) and the additional KWF femur study (2015-2017, funded by the Dutch Cancer Society (KUN 2012-5591)). Additionally, we addressed several issues which would hamper clinical implementation of the developed computer model. The following outline describes the focus of this thesis in more detail.

To underline the need for a better fracture risk prediction tool, in **Chapter 2** we initiated with the validation of the previously described guideline that states that femurs with a cortical involvement of more than 30 mm on diagnostic imaging are at risk of fracture and should be prophylactically stabilized. All available conventional radiographs of the patients in the CT femur study and KWF femur study maximally two months prior to radiotherapy treatment were obtained. Three clinicians assessed the fracture risk by measuring whether the axial cortical involvement was over 30 mm and the diagnostic accuracy of the clinical guideline was determined for this new patient group.

Chapter 3 and 4 describe results on remineralization and fracture risk prediction within the first prospective patient cohort study (CT femur study), including patients with painful femoral metastases from three radiotherapy institutes in the Netherlands. In **Chapter 3**, the short-term effect of radiotherapy on remineralization of femoral bone metastases is investigated, using the multiple CT scans that were acquired at baseline and the different time points after palliative radiotherapy. In **Chapter 4**, the preliminary results from Derikx's thesis were expanded by analyzing the whole patient

group, aiming to confirm the predictive power of the FE model over the current clinical guidelines. FE models were constructed and failure loads were calculated for each of the patients in the CT femur study. Subsequently, it was determined if the FE models were able to distinguish the fracture from the non-fracture patients. Additionally, the FE predictions were compared to clinical assessments on expected fracture risk using digital reconstructed radiographs of two radiation oncologists and an orthopaedic surgeon.

A problem that we encountered while generating the FE models in Chapter 4 was the differences between CT scanners of the different participating centres. Since we had started an additional patient study to include a larger patient dataset (KWF femur study), we liked to investigate the effects of different CT scanners and changes in CT protocols. **Chapter 5** describes the effects of different CT scanners and changes in CT protocols on HU and BMD in tissue characterizing phantoms. Additionally, we performed a supplementary study with cadaveric femurs that were placed in an anatomical body model to mimic the lower body of a patient, which is described in **Chapter 6**. The effect of different CT scanners and CT settings on cortical and trabecular HU and BMD were measured, as well as on failure loads calculated by the FE models.

One of the drawbacks of the FE model for widely clinical exploitation is the requirement of calibrated CT scans as input. As a result, clinical CT scans are only usable when a calibration phantom is scanned along with the patient. Since we liked to be less dependent on this calibration phantom, we developed two phantomless calibration methods: one based on air, fat and muscle tissue within the specific CT scan and one non-patient-specific calibration method. In **Chapter 7** we tested these calibration methods on their ability to produce similar FE failure loads compared to calibration with the use of the phantom.

In **Chapter 8**, we compared the fracture risk prediction ability of the FE models with the 30 mm threshold for axial cortical involvement measured on conventional radiographs as described in the clinical guideline. For this, we combined the clinical assessments from Chapter 2 with the FE database we built from Chapter 4 (CT femur study) and the KWF femur study, and compared diagnostic accuracy of both methods applied on the same patient group.

To conclude, **Chapter 9** reflects on the outcomes of this thesis. Additionally, future perspectives of FE modelling for fracture risk assessment in cancer patients with bone metastases are discussed and how we envision daily clinical utilization of this method on a national scale.

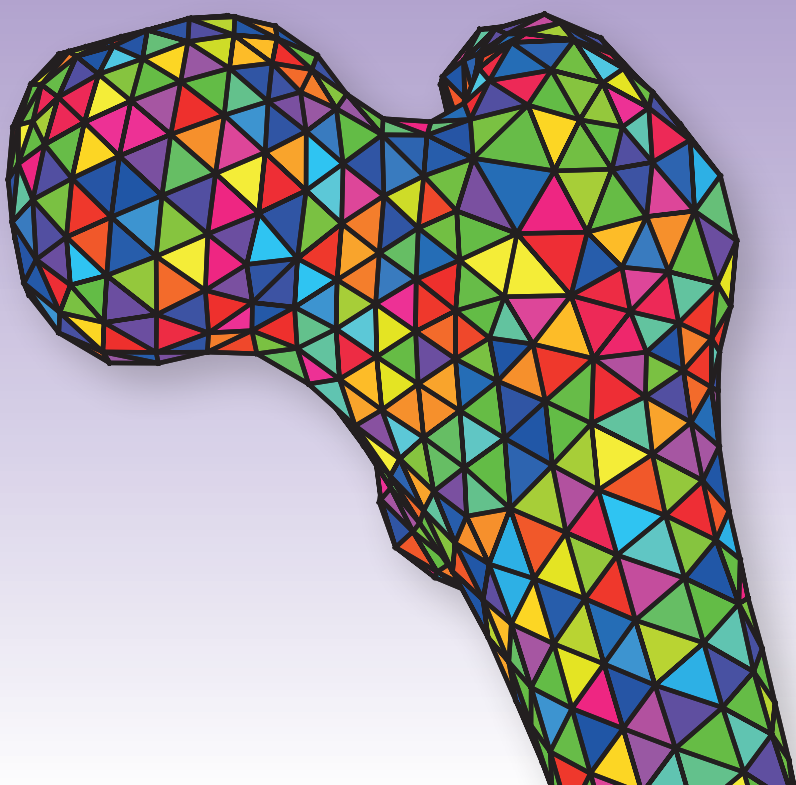
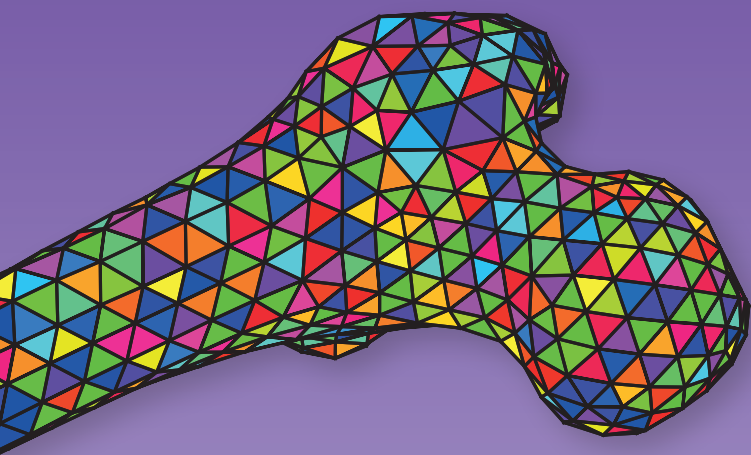
## References

1. IKNL. June 2018. Nederlandse Kankerregistratie, Integraal Kankercentrum Nederland. Available from: [www.cijfersoverkanker.nl](http://www.cijfersoverkanker.nl).
2. Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. 2013. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Available from: <http://globocan.iarc.fr>. Lyon, France: International Agency for Research on Cancer.
3. Schulman KL, Kohles J. 2007. Economic burden of metastatic bone disease in the U.S. *Cancer* 109:2334-2342.
4. Mundy GR. 1997. Mechanisms of bone metastasis. *Cancer* 80:1546-1556.
5. Coleman RE. 1997. Skeletal complications of malignancy. *Cancer* 80:1588-1594.
6. Coleman RE. 2006. Clinical features of metastatic bone disease and risk of skeletal morbidity. *Clin Cancer Res* 12:6243s-6249s.
7. Mundy GR. 2002. Metastasis to bone: causes, consequences and therapeutic opportunities. *Nature reviews Cancer* 2:584-593.
8. Guise T. 2010. Examining the metastatic niche: targeting the microenvironment. *Semin Oncol* 37 Suppl 2:S2-14.
9. Steenland E, Leer JW, van Houwelingen H, Post WJ, van den Hout WB, Kievit J, de Haes H, Martijn H, Oei B, Vonk E, van der Steen-Banasik E, Wiggenraad RG, Hoogenhout J, Warlam-Rodenhuis C, van Tienhoven G, Wanders R, Pomp J, van Reijn M, van Mierlo I, Rutten E. 1999. The effect of a single fraction compared to multiple fractions on painful bone metastases: a global analysis of the Dutch Bone Metastasis Study. *Radiother Oncol* 52:101-109.
10. van der Linden YM, Lok JJ, Steenland E, Martijn H, van Houwelingen H, Marijnen CA, Leer JW, Dutch Bone Metastasis Study G. 2004. Single fraction radiotherapy is efficacious: a further analysis of the Dutch Bone Metastasis Study controlling for the influence of retreatment. *Int J Radiat Oncol Biol Phys* 59:528-537.
11. IKNL. Botmetastasen. Landelijke richtlijn, Versie 1.0.
12. van de Sande MA, Bramer JA, Jutte PC, Schreuder HW, Dijkstra PD. 2010. [Diagnosis and treatment of bone metastasis]. *Ned Tijdschr Geneesk* 154:A2125.
13. Ratasvuori M, Wedin R, Keller J, Nottrott M, Zaikova O, Bergh P, Kalen A, Nilsson J, Jonsson H, Laitinen M. 2013. Insight opinion to surgically treated metastatic bone disease: Scandinavian Sarcoma Group Skeletal Metastasis Registry report of 1195 operated skeletal metastasis. *Surg Oncol* 22:132-138.
14. Mavrogenis AF, Pala E, Romagnoli C, Romantini M, Calabro T, Ruggieri P. 2012. Survival analysis of patients with femoral metastases. *J Surg Oncol* 105:135-141.
15. Blank AT, Lerman DM, Patel NM, Rapp TB. 2016. Is Prophylactic Intervention More Cost-effective Than the Treatment of Pathologic Fractures in Metastatic Bone Disease? *Clin Orthop Relat Res* 474:1563-1570.
16. Rich SE, Chow R, Raman S, Liang Zeng K, Lutz S, Lam H, Silva MF, Chow E. 2018. Update of the systematic review of palliative radiation therapy fractionation for bone metastases. *Radiother Oncol* 126:547-557.
17. Van der Linden YM, Dijkstra PD, Kroon HM, Lok JJ, Noordijk EM, Leer JW, Marijnen CA. 2004. Comparative analysis of risk factors for



- pathological fracture with femoral metastases. *J Bone Joint Surg Br* 86:566-573.
18. Van der Linden YM, Kroon HM, Dijkstra SP, Lok JJ, Noordijk EM, Leer JW, Marijnen CA. 2003. Simple radiographic parameter predicts fracturing in metastatic femoral bone lesions: results from a randomised trial. *Radiother Oncol* 69:21-31.
  19. Koswig S, Budach V. 1999. Remineralisation und Schmerzlinderung von Knochenmetastasen nach unterschiedlich fraktionierter Strahlentherapie (10 mal 3 Gy vs. 1 mal 8 Gy). Eine prospektive Studie. *Strahlenther Onkol* 175:500-508.
  20. Snell W, Beals RK. 1964. Femoral Metastases and Fractures from Breast Cancer. *Surg Gynecol Obstet* 119:22-24.
  21. Beals RK, Lawton GD, Snell WE. 1971. Prophylactic internal fixation of the femur in metastatic breast cancer. *Cancer* 28:1350-1354.
  22. Dijkstra PD, Oudkerk M, Wiggers T. 1997. Prediction of pathological subtrochanteric fractures due to metastatic lesions. *Arch Orthop Trauma Surg* 116:221-224.
  23. Mirels H. 1989. Metastatic disease in long bones. A proposed scoring system for diagnosing impending pathologic fractures. *Clin Orthop*:256-264.
  24. Parrish FF, Murray JA. 1970. Surgical treatment for secondary neoplastic fractures. A retrospective study of ninety-six patients. *J Bone Joint Surg Am* 52:665-686.
  25. Fidler M. 1973. Prophylactic internal fixation of secondary neoplastic deposits in long bones. *Br Med J* 1:341-343.
  26. Tatar Z, Soubrier M, Dillies AF, Verrelle P, Boisgard S, Lapeyre M. 2014. Assessment of the risk factors for impending fractures following radiotherapy for long bone metastases using CT scan-based virtual simulation: a retrospective study. *Radiat Oncol* 9.
  27. Harrington KD. 1982. New trends in the management of lower extremity metastases. *Clin Orthop Relat Res*:53-61.
  28. Damron TA, Morgan H, Prakash D, Grant W, Aronowitz J, Heiner J. 2003. Critical evaluation of Mirels' rating system for impending pathologic fractures. *Clin Orthop Relat Res*:S201-207.
  29. Shimoyama T, Katagiri H, Harada H, Murata H, Wasa J, Hosaka S, Suzuki T, Takahashi M, Asakura H, Nishimura T, Yamada H. 2017. Fracture after radiation therapy for femoral metastasis: incidence, timing and clinical features. *J Radiat Res* 58:661-668.
  30. El-Husseiny M, Coleman N. 2010. Inter- and intra-observer variation in classification systems for impending fractures of bone metastases. *Skeletal Radiol* 39:155-160.
  31. Keene JS, Sellinger DS, McBeath AA, Engber WD. 1986. Metastatic breast cancer in the femur. A search for the lesion at risk of fracture. *Clin Orthop Relat Res*:282-288.
  32. Benca E, Patsch JM, Mayr W, Pahr DH, Windhager R. 2016. The insufficiencies of risk analysis of impending pathological fractures in patients with femoral metastases: A literature review. *Bone reports* 5:51-56.
  33. Bessho M, Ohnishi I, Matsumoto T, Ohashi S, Matsuyama J, Tobita K, Kaneko M, Nakamura K. 2009. Prediction of proximal femur strength using a CT-based nonlinear finite element method: differences in predicted fracture load and site with changing load and boundary conditions. *Bone* 45:226-231.
  34. Keyak JH, Sigurdsson S, Karlsdottir GS, Oskarsdottir D, Sigmarsdottir A, Kornak J, Harris TB, Sigurdsson G, Jonsson BY, Siggeirsdottir K, Eiriksdottir G, Gudnason V, Lang TF. 2013. Effect of finite element model loading condition on

- fracture risk assessment in men and women: the AGES-Reykjavik study. *Bone* 57:18-29.
35. Zysset P, Pahr D, Engelke K, Genant HK, McClung MR, Kendler DL, Recknor C, Kinzl M, Schwiedrzik J, Museyko O, Wang A, Libanati C. 2015. Comparison of proximal femur and vertebral body strength improvements in the FREEDOM trial using an alternative finite element methodology. *Bone* 81:122-130.
  36. Cheal EJ, Hipp JA, Hayes WC. 1993. Evaluation of finite element analysis for prediction of the strength reduction due to metastatic lesions in the femoral neck. *J Biomech* 26:251-264.
  37. Keyak JH, Kaneko TS, Tehranzadeh J, Skinner HB. 2005. Predicting proximal femoral strength using structural engineering models. *Clin Orthop*:219-228.
  38. Keyak JH, Kaneko TS, Rossi SA, Pejic MR, Tehranzadeh J, Skinner HB. 2005. Predicting the strength of femoral shafts with and without metastatic lesions. *Clin Orthop* 439:161-170.
  39. Spruijt S, van der Linden JC, Dijkstra PDS, Wiggers T, Oudkerk M, Snijders CJ, van Keulen F, Verhaar JAN, Weinans H, Swierstra BA. 2006. Prediction of torsional failure in 22 cadaver femora with and without simulated subtrochanteric metastatic defects - A CT scan-based finite element analysis. *Acta Orthop* 77:474-481.
  40. Yosibash Z, Mayo RP, Dahan G, Trabelsi N, Amir G, Milgrom C. 2014. Predicting the stiffness and strength of human femurs with real metastatic tumors. *Bone* 69:180-190.
  41. Keyak JH, Lee IY, Nath DS, Skinner HB. 1996. Postfailure compressive behavior of tibial trabecular bone - Three anatomic directions. *J Biomed Mater Res* 31:373-378.
  42. Kaneko TS, Pejic MR, Tehranzadeh J, Keyak JH. 2003. Relationships between material properties and CT scan data of cortical bone with and without metastatic lesions. *Med Eng Phys* 25:445-454.
  43. Kaneko TS, Bell JS, Pejic MR, Tehranzadeh J, Keyak JH. 2004. Mechanical properties, density and quantitative CT scan data of trabecular bone with and without metastases. *J Biomech* 37:523-530.
  44. Keyak JH. 2001. Improved prediction of proximal femoral fracture load using nonlinear finite element models. *Med Eng Phys* 23:165-173.
  45. Derikx LC, Vis R, Meinders T, Verdonchot N, Tanck E. 2011. Implementation of asymmetric yielding in case-specific finite element models improves the prediction of femoral fractures. *Comput Methods Biomech Biomed Engin* 14:183-193.
  46. Derikx LC, van Aken JB, Janssen D, Snyers A, van der Linden YM, Verdonchot N, Tanck E. 2012. The assessment of the risk of fracture in femora with metastatic lesions: comparing case-specific finite element analyses with predictions by clinical experts. *J Bone Joint Surg Br* 94:1135-1142.
  47. Tanck E, van Aken JB, van der Linden YM, Schreuder HW, Binkowski M, Huizenga H, Verdonchot N. 2009. Pathological fracture prediction in patients with metastatic lesions can be improved with quantitative computed tomography based computer models. *Bone* 45:777-783.
  48. Derikx L. 2015. Femoral fracture risk prediction in metastatic bone disease. Doctoral Thesis, Radboud University Nijmegen, The Netherlands.
  49. Hsieh J, Molthen RC, Dawson CA, Johnson RH. 2000. An iterative approach to the beam hardening correction in cone beam CT. *Med Phys* 27:23-29.
  50. Siewerdsen JH, Daly MJ, Bakhtiar B, Moseley DJ, Richard S, Keller H, Jaffray DA. 2005. A simple, direct method for x-ray scatter estimation and correction in digital radiography and cone-beam CT. *Med Phys* 33:187-197.





# CHAPTER 2

## **AXIAL CORTICAL INVOLVEMENT OF METASTATIC LESIONS TO IDENTIFY IMPENDING FEMORAL FRACTURES; A CLINICAL VALIDATION STUDY**

---

Gerco van der Wal, Florieke Eggermont, Marta Fiocco, Herman Kroon, Onarisa Ayu, Annerie Slot,  
An Snyers, Tom Rozema, Nico Verdonchot, Sander Dijkstra, Esther Tanck, Yvette van der Linden

Radiotherapy & Oncology, in press



## Introduction

Due to increased life expectancy, higher cancer incidence and improved systemic treatments for cancer patients, incidence of symptomatic bone metastases increases.<sup>1</sup> Focussing on bone metastases in the femur, most patients experience bone pain, and, if left untreated, progressive lesions can eventually lead to pathological fractures, with high symptom burden, impaired mobility and the need for emergency surgeries. Choice of local treatment for femoral bone metastases depends on the expected fracture risk as appraised by the treating physician, besides the patient's estimated survival and preferences. Assessing femoral fracture risk is challenging but of key importance.

In general, for low risk lesions, a single dose of radiotherapy (RT) is the treatment of choice, with about 60% of patients experiencing pain relief.<sup>2</sup> For high risk lesions, surgical options consist of intramedullary nailing, plate osteosynthesis, or (endo) prosthetic reconstruction.<sup>3</sup> Alternatively, if patients are too ill to undergo surgery or refuse surgery, higher doses RT are used to induce remineralisation of the bone defect.<sup>4</sup>

Studies concerning femoral fracture prediction reported several radiological risk factors.<sup>5-11</sup> Most were based on lesion characteristics in patients who were subsequently operated on, and therefore, the natural course of lesions without fixation during follow-up was not taken into account. In 1989, Mirels described a scoring system composed of four factors: site, aspect and size of the metastatic lesion and patient reported pain, with a maximum of 12 points.<sup>12</sup> Although only one in seven patients actually experienced a fracture, he advised that all patients who scored nine points or higher should be prophylactically stabilised. In 2004, we compared the predictive value of all frequently-cited risk factors in 102 patients with femoral metastases who were treated with palliative radiotherapy within a phase 3 trial.<sup>10</sup> We concluded that most risk factors were unreliable in assessing fracture risk and led to overtreatment of patients. The use of Mirels' scoring system highly overestimated fracture risk with a positive predictive value (PPV) of only 14%, indicating that approximately 6 out of 7 patients would have undergone unnecessary surgery, and thus exposing them to possible complications and the burden of time for rehabilitation.<sup>10</sup>

Other studies proposed to measure axial cortical destruction on conventional radiographs.<sup>5,8</sup> Based on the same abovementioned 102 patients, we showed in 2003 that lesions with more than 30 mm axial cortical involvement were at risk for fracturing, with a PPV and negative predicting value (NPV) of 23% and 97%, respectively, illustrating that approximately one in four patients with a lesion  $\geq 30$  mm would encounter a fracture.<sup>11</sup>

The use of axial cortical involvement was not validated in other studies yet. Therefore, the aim of our present study was to validate the clinical reliability of the 30 mm threshold for axial cortical involvement and reevaluate other published risk factors measured on conventional radiographs for fracture risk assessment in patients with femoral metastases.

## Methods and patients

For the current study, we used a patient cohort originally aimed to design and improve a patient-specific and CT-based finite element (FE) computer simulation model to predict fracture risk of metastatic femoral bone lesions.<sup>13</sup> This cohort comprised 156 patients included in two prospective multicentre studies, conducted in four radiotherapy departments in the Netherlands, between 2006 and 2009<sup>13,14</sup> and 2015 and 2017. In these studies, advanced cancer patients with painful femoral metastases who were referred for radiotherapy were asked to participate. At study entry, the following baseline characteristics were obtained: primary tumour, sex, age, height, weight, pain score (0 to 10), Karnofsky Performance Score (KPS, 0 to 100) and bisphosphonate use. Treatment guidelines in both studies stated that patients with lesions < 30 mm should receive a single fraction (SF) of 8 Gy to treat pain. For larger lesions, patients were preferably operated on to prevent pathological fracturing, but if a patient refused surgery, or was in poor clinical condition, multiple (5 or 6) fractions (MF) of 4 Gy were given to induce remineralisation. If patients were too ill to travel to the radiotherapy department for multiple times, the use of SF was allowed. Both studies received local medical ethical review board approvals and all patients signed informed consent. Specifics concerning inclusion and exclusion criteria have been discussed elsewhere.<sup>13</sup>

Adequate follow-up was established through multiple telephone interviews or questionnaires in the first six months after treatment, or until death or fracture occurred. In addition, the follow-up was updated until November 2017 via electronic patient records.

For the current study, all conventional anteroposterior (AP) and/or lateral radiographs available within a two months period before radiotherapy were collected. After excluding patients lacking conventional radiographs within the two-month time frame ( $n = 56$ ), we included 100 patients in this study.

All conventional radiographs were individually reviewed by three experienced observers: a radiation oncologist, an orthopaedic surgeon and a musculoskeletal radiologist. Identical reviewing methods in comparison with the original study in 2003 were used.<sup>11</sup> Firstly, observers were asked to indicate based on their clinical experience whether or not the affected bone was at high risk for fracturing and if the patient should be discussed with an orthopaedic surgeon to consider prophylactic surgery. Hereafter, the appearance of the bone lesion (lytic, blastic or mixed), largest axial cortical involvement (in mm) and circumference of the lesion using a three-tiered approach ( $\leq 1/3$ ,  $1/3 - 2/3$  and  $> 2/3$ ) was assessed. When multiple lesions were at risk, they were scored separately if they were more than 50mm apart. In case of disagreement between the observers on any of the variables, consensus was reached through discussion. Additionally, we applied the Mirels' scoring system<sup>12</sup> identically to the previous study.<sup>10</sup>

## Statistical analyses

To assess any possible bias between included and excluded patients based on availability of radiographs, and between patients with and without fractures during follow-up, baseline characteristics were assessed for normality with the Shapiro-Wilk test and compared using independent t-tests and Mann-Witney U tests for continuous and categorical variables, respectively. Reverse Kaplan Meier was used to assess median follow-up.<sup>15</sup>

Competing risk models<sup>16,17</sup> with two competing events fracture and death, (Supplementary Data 1,2 and 3) were used to estimate the cumulative incidence function of death and fracture for cortical involvement ( $\geq 30$  mm or  $< 30$  mm) and for radiotherapy treatment schedule (SF or MF). The competing risk model has been estimated since patients may die without developing a pathological fracture. Fine and Gray's test<sup>18</sup> was used to assess the differences in the estimated cumulative incidence between groups. To study the effect of risk factors on fracture, univariate (UV) Cox proportional hazards regression models were estimated. Cause specific hazard ratios ( $HR_{cs}$ ) along with their 95% confidence interval (CI) were estimated. This estimated hazard ratio represents the relative increase of hazard of experiencing the event of interest among those patients who have not experienced any event yet, i.e. they are still event free and in follow-up.<sup>16</sup>

Interobserver variability was assessed using Spearman's rank correlation. Sensitivity (SE, percentage of high risk patients in the fracture group), specificity (SP, percentage of low risk patients in the non-fracture group), PPV (percentage of patients with a fracture in the high risk group) and NPV (percentage of patients without a fracture in the low risk group) were calculated.

Competing risk analysis was performed by using the mstate library in R.<sup>19,20</sup> All other analyses were performed using SPSS 24.0 (Armonk, NY, USA). Statistical significance was set at  $p < 0.05$ .

## Results

Baseline characteristics were not different between included and excluded patients and between patients with or without fractures (Table 1). Median follow-up of the 100 included patients was 23.0 months (95%CI: 10.6 – 35.5). Thirteen patients (13%) developed a pathological fracture during follow-up. One patient fractured both femurs. At time of database closure, 73 patients (73%) had died; 62 without a fracture and 11 patients died after they had sustained a fracture. The cumulative incidences of death for  $< 30$  mm versus  $\geq 30$  mm cortical involvement and for single versus multiple fractions RT are shown in Figure 1.



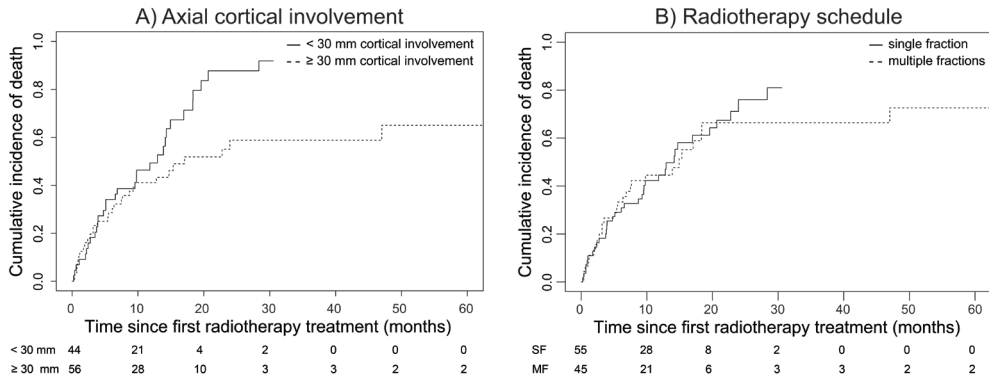
**Table 1:** Baseline characteristics

	Included patients					
	Included n = 100	Excluded* n = 56	p-value	Fracture n = 13	No fracture n = 87	p-value
<b>Sex</b>			0.45			0.28
male	60 (60.0%)	37 (66.1%)		6 (46.2%)	54 (62.1%)	
<b>Age in years<sup>§</sup></b>			0.58			0.99
mean (SD)	67.69 (10.6)	68.77 (11.8)		67.7 (8.1)	67.7 (11.0)	
<b>Primary tumour</b>			0.16			0.73
Breast	18 (18.0%)	14 (25.0%)		4 (30.8%)	14 (16.1%)	
Prostate	32 (32.0%)	25 (44.6%)		2 (15.4%)	30 (34.5%)	
Lung	25 (25.0%)	3 (5.4%)		3 (23.1%)	22 (25.3%)	
Multiple myeloma	13 (13.0%)	4 (7.1%)		4 (30.8%)	9 (10.3%)	
Other	12 (12.0%)	10 (17.9%)		-	12 (13.8%)	
<b>Pain score<sup>¶</sup></b>			0.58			0.64
mean (SD)	5.46 (2.7)	5.2 (2.6)		5.9 (3.2)	5.4 (2.6)	
<b>KPS</b>			0.57			0.88
≥ 80	59 (59.0%)	36 (64.3%)		8 (61.5%)	51 (58.6%)	
≤ 70	40 (40.0%)	20 (35.7%)		5 (38.5%)	35 (40.2%)	
Missing	1 (1.0%)	-		-	1 (1.1%)	
<b>Radiotherapy schedule<sup>#</sup></b>			0.61			0.93
SF	55 (55.0%)	33 (58.9%)		7 (53.8)	48 (55.2)	
MF	45 (45.0%)	22 (39.3%)		6 (46.2)	39 (44.8)	
<b>Bisphosphonate use</b>			0.89			0.54
Yes	24 (24.0%)	14 (25.0%)		4 (30.8%)	20 (23.0%)	
No	76 (76.0%)	42 (75.0%)		9 (69.2%)	67 (77.0%)	
<b>Body Mass Index<sup>&amp;</sup></b>			0.56			0.55
mean (SD)	25.1 (4.2)	26.47 (3.95)		25.6 (3.1)	25.0 (4.4)	

\*Due to missing (eligible) conventional radiographs, <sup>§</sup>Age is missing for one excluded patient, <sup>¶</sup>Pain score is missing for 3 included patients and 2 excluded patients, <sup>#</sup>One excluded patient was not treated with radiotherapy, <sup>&</sup>Body Mass Index is missing for 16 included patients and 5 excluded patients

KPS, Karnofsky Performance Score, SF = single fraction radiotherapy, MF = multiple fraction radiotherapy, Body Mass Index, (weight (kg) / (height (m)<sup>2</sup>)

On the conventional radiographs, 110 lesions were identified in the 100 patients (Table 2). Appearance was considered lytic in 73 lesions (66%), mixed in 25 (23%) and blastic in 12 (11%). Most lesions were located subtrochanteric (n = 22, 20%), followed by the shaft (n = 17, 15%) and intertrochanteric (n = 16, 15%), head or neck (n = 14, 13%) and distal (n = 1, 1%). Forty lesions (36%) were visible on bone scintigraphy or PET/CT, but were not clearly visible on the conventional radiographs, and therefore the exact location or size could not be assessed. Fractures occurred in 11 osteolytic and three mixed lesions, of which ten were located in the proximal femur and two in the shaft. Two fractures occurred in lesions that were not clearly visible.



**Figure 1:** Estimated cumulative incidence of death from a competing risk model with two competing events: fracture and death; (A) axial cortical involvement  $\geq 30$  mm ( $n=56$ ) versus  $< 30$  mm ( $n=44$ ) ( $p=0.037$ ) (B) radiotherapy schedule single ( $n=61$ ) versus multiple fractions ( $n=45$ ) ( $p=0.647$ )

Spearman's rank correlations for interobserver agreement were 0.34 – 0.54 for judgement of impending fracture based on clinical experience, 0.44 – 0.57 for axial cortical involvement, 0.43 – 0.68 for Mirels' scoring system, 0.29 – 0.56 for lesion appearance, and 0.39 – 0.55 for circumferential cortical involvement.

Twenty-seven lesions, of which six fractured (22%), were deemed as high risk based on clinical experience, i.e. that observers advised consultation with an orthopaedic surgeon to consider prophylactic surgery. Of the 83 lesions that were indicated as low risk, three fractured (10%). When judging impending fracture based on clinical experience, SE, SP, PPV and NPV were 43%, 78%, 22% and 90%, respectively (Table 3), although  $HR_{cs}$  was non-significant (2.3; 95%CI: 0.8-6.5,  $p = 0.13$ ).

In the 50 lesions with an axial cortical involvement  $< 30$  mm, two fractures (4%) occurred. In the 60 lesions  $\geq 30$  mm, 12 fractures (20%) occurred. Using axial cortical involvement for predicting fractures resulted in a SE, SP, PPV and NPV of 86%, 50%, 20% and 96%, respectively (Table 3). Lesions  $\geq 30$  mm had a  $HR_{cs}$  of 5.3 (95%CI: 1.2 – 23.9,  $p = 0.03$ ), indicating a 5.3 times higher risk of fracture than  $< 30$  mm lesions (Table 2). Figure 2A shows the estimated cumulative fracture incidence for  $< 30$  mm versus  $\geq 30$  mm cortical involvement.

Mirels' scoring system resulted in a SE, SP, PPV and NPV of 77%, 45%, 17% and 93%, respectively (Table 3). When impending fracture was based on clinical experience, these diagnostic values were 43%, 78%, 22% and 90%, respectively. Mirels' scoring system, as well as circumferential axial involvement (except  $> 2/3$ ) were not found to be statistically associated with the occurrence of a fracture (Table 2). Radiation schedule was not associated with fracture risk ( $HR_{cs}$ : 0.8, 95%CI: 0.3 – 2.2,  $p = 0.62$ ). The estimated cumulative fracture incidence for single versus multiple fractions RT was not significantly different ( $p = 0.69$ , Figure 2B).

**Table 2:** Differences in risk factors between patients that did and did not develop a fracture, and the univariate cox regression analysis of those risk factors

	Fracture Yes n = 14	No n = 96	p-value	UVA HR <sub>cs</sub> (95% CI)	p-value
<b>Judgement of impending fracture based on clinical experience</b>			0.09		
Yes	6 (42.9%)	21 (21.8%)		2.3 (0.8 – 6.5)	0.13
No	8 (57.1%)	75 (78.1%)		ref	
<b>Axial cortical involvement<sup>11</sup></b>			0.01		
≥ 30 mm	12 (85.7%)	48 (50.0%)		5.3 (1.2 – 23.9)	0.03
< 30 mm	2 (14.3%)	48 (50.0%)		ref	
<b>Circumference</b>			0.01		
> 2/3	5 (35.7%)	12 (12.5%)		4.4 (1.2 – 16.2)	0.03
1/3 – 2/3	5 (35.7%)	26 (20.8%)		2.5 (0.7 – 9.3)	0.18
≤ 1/3	4 (28.6%)	58 (60.4%)		ref	
<b>Mirels scoring system<sup>12</sup></b>			0.14		
9 or higher	10 (71.4%)	50 (52.1%)		2.9 (0.8 – 10.7)	0.10
8 or lower	3 (21.4%)	41 (42.7%)		ref	
Missing <sup>§</sup>	1 (7.1%)	5 (5.2%)			

<sup>§</sup>Due to missing data regarding pain

UVA= Univariate analyses

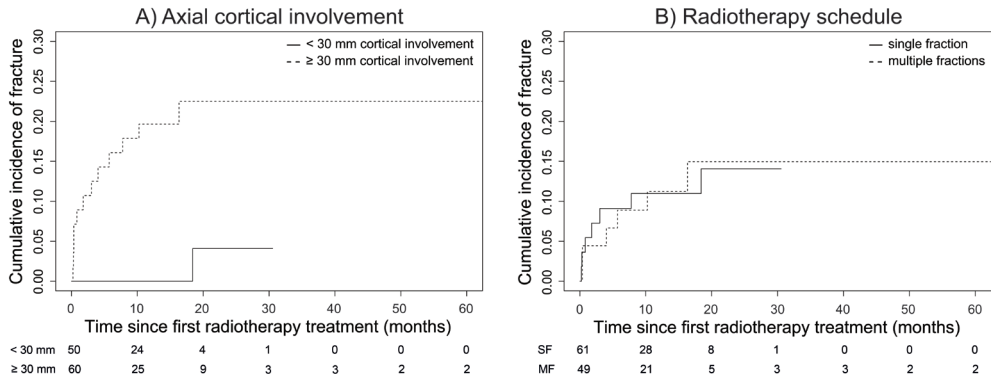
HR<sub>cs</sub> = Cause Specific Hazard Ratio

**Table 3:** Sensitivity, specificity, positive and negative predicting values of risk factors for impending pathological fracture

	Fracture n = 14 (%)	No fracture n = 96 (%)	SE	SP	PPV	NPV
<b>Judgement of impending fracture based on clinical experience*</b>			43	78	22	90
Yes, 26 (24%)	6 (40.0)	21 (21.1)				
No, 84 (76%)	8 (60.0)	75 (78.9)				
<b>Axial cortical involvement<sup>11</sup></b>			86	50	20	96
≥ 30 mm, 60 (55%)	12 (86.7)	48 (50.0)				
< 30 mm, 50 (45%)	2 (13.3)	48 (50.0)				
<b>Mirels' scoring system<sup>12</sup></b>			77	45	17	93
9 or higher, 60 (55%)	10 (71.4)	50 (76.8)				
8 or lower, 44 (40%)	3 (21.4)	41 (21.1)				
Missing <sup>§</sup> , 6 (5%)	1 (7.1)	5 (5.2)				

SE: sensitivity, SP: specificity, PPV: positive predicting value, NPV: negative predicting value

\* Answer to the question: Should this patient be discussed with an orthopaedic surgeon to consider prophylactic surgery, based on your clinical experience? <sup>§</sup> Due to missing data regarding pain



**Figure 2:** Competing risk analyses showing the cumulative fracture incidence in 110 metastatic lesions for A) axial cortical involvement  $\geq 30$  mm ( $n = 60$ ) versus  $< 30$  mm ( $n = 50$ ) ( $p = 0.012$ ) and B) radiotherapy treatment schedule single ( $n = 61$ ) versus multiple fractions ( $n = 49$ ) ( $p = 0.69$ )

## Discussion

In this study, three observers independently reviewed 110 metastatic femoral bone lesions of 100 patients on conventional radiographs aiming to validate the 30 mm threshold for axial cortical involvement<sup>11</sup> and reevaluate other published risk factors for assessment of impending femoral fracture.<sup>10</sup> Most risk factors were not associated with fracture occurrence. Fracture risk assessment using the 30 mm threshold for axial cortical involvement was associated with fracture occurrence and resulted in a PPV and NPV of 20% and 96%, respectively. These values are similar to those of the original study on 102 patients and 110 lesions in which the PPV and NPV were 23% and 97%, respectively.<sup>11</sup> Our results confirm that approximately one in every five patients with a metastatic lesion  $\geq 30$  mm will encounter a pathological fracture, and have a 5.3 times higher risk of fracture and should therefore be considered for prophylactic surgery. On the other hand, those with a lesion  $< 30$  mm are at low risk of fracture and may receive non-invasive radiotherapy to treat pain.

Overtreatment (low PPV), i.e. patients with a short life expectancy who would not have developed a fracture but receive prophylactic surgery, is disconcerting, as is undertreatment (low NPV), i.e. the occurrence of pathological fractures in low risk patients. To choose the most optimal patient-specific local treatment, the expected fracture risk and life expectancy should be taken into account.<sup>21</sup> Another issue is the choice of surgical procedure: patients with a relatively prolonged life expectancy are at risk for long-term complications with simple reconstructions, such as plate or intramedullary nail fixation. Therefore, more elaborate surgery, such as total hip arthroplasty or proximal/distal femoral replacement, should be considered.<sup>21</sup> The additional medical complications as a result of an actual pathological fracture, apart from anxiety and stress to the patient, are well known and, in general, worse than the complications related to prophylactic surgery.<sup>22-25</sup>

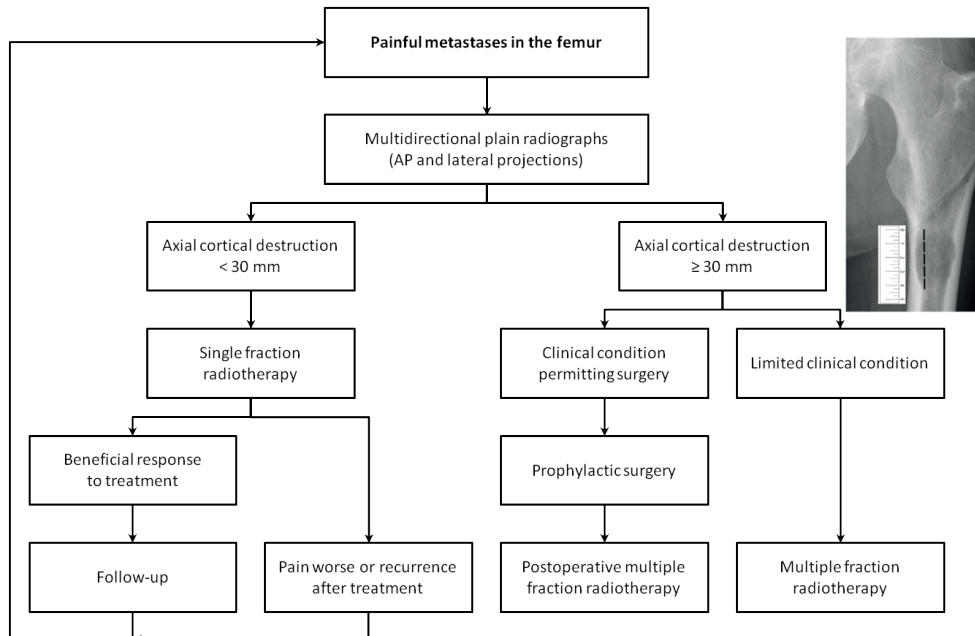
It is preferable to prevent all pathological femoral fractures, and therefore a high NPV is of importance for a reliable risk factor. In our study, Mirels' scoring system had an NPV of 93%. However, if used to determine actual treatment, unnecessary surgical interventions would have been performed in 83% of the surgically treated patients and therefore they would be needlessly exposed to surgical complication risks. In order to reduce the rate of overtreatment, we would argue that using axial cortical involvement  $\geq 30$  mm, with two missed fractures compared to three using Mirels' as well as a slightly higher NPV (96%), is preferable to assess fracture risk, despite the modest improvement. The variability of the measurements between the three observers was slightly larger than in the original study.<sup>11</sup> We suggest that, according to the standard work flow procedures, the radiologist assesses the fracture risk and subsequently advises the treating physicians, to avoid inconsistencies between different medical specialties.

This study has some limitations. The absence of conventional radiographs in 56 of the 156 patients (36%) potentially could have led to selection bias in the studied population. However, no significant differences in baseline variables between included and excluded patients were detected. Another plausible form of bias could have been introduced by the applied radiotherapy schedule, as patients with larger lesions would have received multiple fractions (MF) in order to induce remineralisation when following treatment guidelines. Our earlier paper suggested a nearly significant effect of MF on fracture postponement corrected for lesion size ( $p = 0.07$ ).<sup>11</sup> In the present study, no effect of radiotherapy schedule on fracture incidence and fracture risk was found. This is remarkable, since most patients with larger lesions are now treated with MF RT according to clinical guidelines. Only if the clinical condition is too poor, also in high risk lesions, patients receive single fraction RT. We did not record why choices for SF or MF were made in treatment plans of the included patients. Apparently, despite the clinical guidelines, individual decisions are made by treating physicians in consultation with patients based on other aspects than only lesion size. Another remark is that information about patients' activity level was lacking. Therefore, results could be masked by the reduction in bone strength in inactive patients, who, if they had been more active, could have developed a fracture. Obviously, patients included in this study did not receive prophylactic surgery, on account of either low fracture risk assessment, or because patients' physical condition or preferences did not allow invasive surgery. Detailed information about the reason for withholding surgery in larger lesions is not available. Nonetheless, one could argue that the reported fracture incidence in this study could be higher in all patients with symptomatic femoral metastases due to the selection of only low risk patients or those in poor health. Another limitation is the inclusion of patients with multiple myeloma. Although bone lesions derived from multiple myeloma are not considered as true bone metastases, they have similar features as lytic bone metastases from solid tumours and assessment of risk of fracturing is of similar importance.

Recent literature has shown that CT imaging has a very good sensitivity for diagnosing metastatic bone lesions.<sup>26</sup> Therefore, the value of axial cortical involvement and other risk factors measured on CT scans instead of conventional radiographs has to be studied to further improve fracture risk prediction. Besides clinical parameters, some studies focus on developing biomechanical computer models, such

as finite element (FE) models, to objectively judge fracture risks. Although several obstacles exist, such as possible inaccuracy due to variations in type of CT scanner and scanner settings,<sup>27</sup> first results are promising.<sup>13,28</sup> However, generation of FE models is time consuming, and it should be investigated whether the clinical value of FE models outweighs the easy-to-use measurement of axial cortical involvement.

In conclusion, this study validated the previously reported risk factor axial cortical involvement with a threshold of 30 mm to predict fracture risk in cancer patients with femoral metastases.<sup>11</sup> Our results were similar to the earlier reports, and therefore, until a more accurate and practically feasible method is developed, this clinical parameter remains an easy method to assess fracture risk, to aid patients and clinicians to choose the most optimal individual treatment modality (Figure 3).



**Figure 3:** Treatment algorithm regarding patients with femoral bone metastases

## Acknowledgements

This work was supported by the Dutch Cancer Society / Alpe d'HuZes (UL2013-6286), the Dutch Science Foundation NWO-STW (NPG.06778), Fonds NutsOhra (1102-071), and the Dutch Cancer Society (KUN 2012-5591).

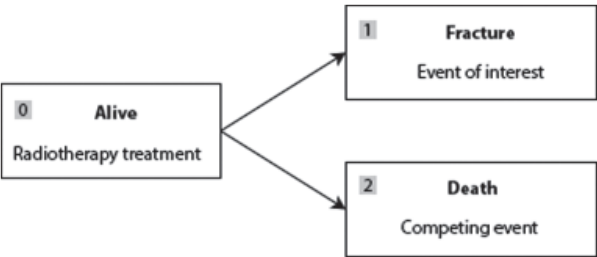
## References

- Schulman KL, Kohles J. 2007. Economic burden of metastatic bone disease in the U.S. *Cancer* 109:2334-2342.
- Rich SE, Chow R, Raman S, Liang Zeng K, Lutz S, Lam H, Silva MF, Chow E. 2018. Update of the systematic review of palliative radiation therapy fractionation for bone metastases. *Radiother Oncol* 126:547-557.
- Willeumier JJ, van der Linden YM, van de Sande MAJ, Dijkstra PDS. 2016. Treatment of pathological fractures of the long bones. *EFORT open reviews* 1:136-145.
- Koswig S, Budach V. 1999. Remineralisation und Schmerzlinderung von Knochenmetastasen nach unterschiedlich fraktionierter Strahlentherapie (10 mal 3 Gy vs. 1 mal 8 Gy). Eine prospektive Studie. *Strahlenther Onkol* 175:500-508.
- Beals RK, Lawton GD, Snell WE. 1971. Prophylactic internal fixation of the femur in metastatic breast cancer. *Cancer* 28:1350-1354.
- Fidler M. 1981. Incidence of fracture through metastases in long bones. *Acta Orthop Scand* 52:623-627.
- Harrington KD. 1982. New trends in the management of lower extremity metastases. *Clin Orthop Relat Res*:53-61.
- Dijkstra PD, Oudkerk M, Wiggers T. 1997. Prediction of pathological subtrochanteric fractures due to metastatic lesions. *Arch Orthop Trauma Surg* 116:221-224.
- Keene JS, Sellinger DS, McBeath AA, Engber WD. 1986. Metastatic breast cancer in the femur. A search for the lesion at risk of fracture. *Clin Orthop Relat Res*:282-288.
- Van der Linden YM, Dijkstra PD, Kroon HM, Lok JJ, Noordijk EM, Leer JW, Marijnen CA. 2004. Comparative analysis of risk factors for pathological fracture with femoral metastases. *J Bone Joint Surg Br* 86:566-573.
- Van der Linden YM, Kroon HM, Dijkstra SP, Lok JJ, Noordijk EM, Leer JW, Marijnen CA. 2003. Simple radiographic parameter predicts fracturing in metastatic femoral bone lesions: results from a randomised trial. *Radiother Oncol* 69:21-31.
- Mirels H. 1989. Metastatic disease in long bones. A proposed scoring system for diagnosing impending pathologic fractures. *Clin Orthop*:256-264.
- Eggermont F, Derikx LC, Verdonchot N, Van der Geest ICM, De Jong MAA, Snyers A, Van der Linden YM, Tanck E. 2018. Can patient-specific finite element models better predict fractures in metastatic bone disease than experienced clinicians? Towards introducing computational modelling into daily clinical practice. *Bone Joint Res* 7:430-439.
- Eggermont F, Derikx LC, Verdonchot N, Hannink G, Kaatee R, Tanck E, van der Linden YM. 2017. Limited short-term effect of palliative radiation therapy on quantitative computed tomography-derived bone mineral density in femora with metastases. *Adv Radiat Oncol* 2:53-61.
- Schemper M, Smith TL. 1996. A note on quantifying follow-up in studies of failure time. *Control Clin Trials* 17:343-346.
- Putter H, Fiocco M, Geskus RB. 2007. Tutorial in biostatistics: competing risks and multi-state models. *Stat Med* 26:2389-2430.
- Dutz A, Lock S. 2019. Competing risks in survival data analysis. *Radiother Oncol* 130:185-189.
- Gray RJ. 1988. A Class of K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk. *Ann Stat* 16:1141-1154.

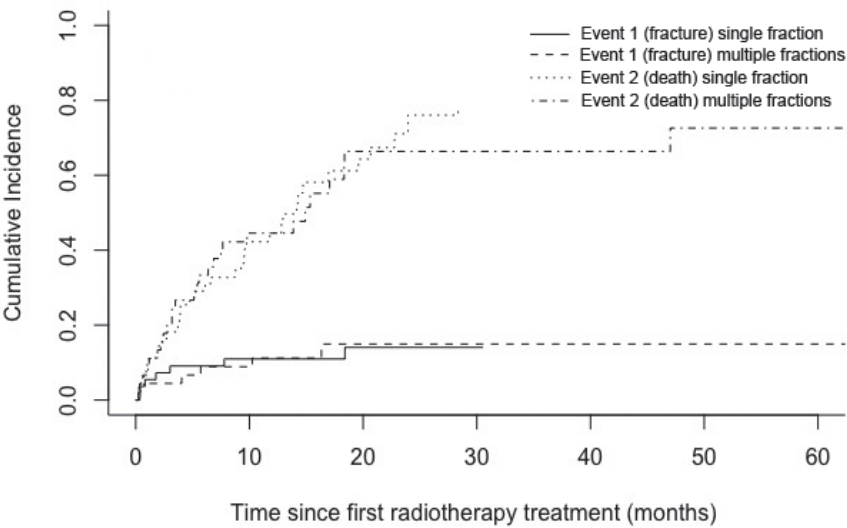
19. de Wreede LC, Fiocco M, Putter H. 2010. The mstate package for estimation and prediction in non- and semi-parametric multi-state and competing risks models. *Comput Methods Programs Biomed* 99:261-274.
20. de Wreede LC, Fiocco M, Putter H. 2011. mstate: An R Package for the Analysis of Competing Risks and Multi-State Models. *J Stat Softw* 38:1-30.
21. Willeumier JJ, van der Linden YM, van der Wal CWPG, Jutte PC, van der Velden JM, Smolle MA, van der Zwaal P, Koper P, Bakri L, de Pree I, Leithner A, Fiocco M, Dijkstra PDS. 2018. An Easy-to-Use Prognostic Model for Survival Estimation for Patients with Symptomatic Long Bone Metastases. *Journal of Bone and Joint Surgery-American Volume* 100:196-204.
22. Blank AT, Lerman DM, Patel NM, Rapp TB. 2016. Is Prophylactic Intervention More Cost-effective Than the Treatment of Pathologic Fractures in Metastatic Bone Disease? *Clin Orthop Relat Res* 474:1563-1570.
23. Kreul SM, Sorger JI, Rajamanickam VP, Heiner JP. 2016. Updated Outcomes of Prophylactic Femoral Fixation. *Orthopedics* 39:e346-352.
24. Ward WG, Holsenbeck S, Dorey FJ, Spang J, Howe D. 2003. Metastatic disease of the femur: surgical treatment. *Clin Orthop Relat Res*:S230-244.
25. Willeumier JJ, Kaynak M, van der Zwaal P, Meylaerts SAG, Mathijssen NMC, Jutte PC, Tsagozis P, Wedin R, van de Sande MAJ, Fiocco M, Dijkstra PDS. 2018. What Factors Are Associated With Implant Breakage and Revision After Intramedullary Nailing for Femoral Metastases? *Clin Orthop Relat Res* 476:1823-1833.
26. Hamaoka T, Madewell JE, Podoloff DA, Hortobagyi GN, Ueno NT. 2004. Bone imaging in metastatic breast cancer. *J Clin Oncol* 22:2942-2953.
27. Eggermont F, Derikx LC, Free J, Van Leeuwen R, Van der Linden YM, Verdonchot N, Tanck E. 2018. Effect of Different CT Scanners and Settings on Femoral Failure Loads Calculated by Finite Element Models. *J Orthop Res* 36:2288-2295.
28. Goodheart JR, Cleary RJ, Damron TA, Mann KA. 2015. Simulating activities of daily living with finite element analysis improves fracture prediction for patients with metastatic femoral lesions. *J Orthop Res* 33:1226-1234.



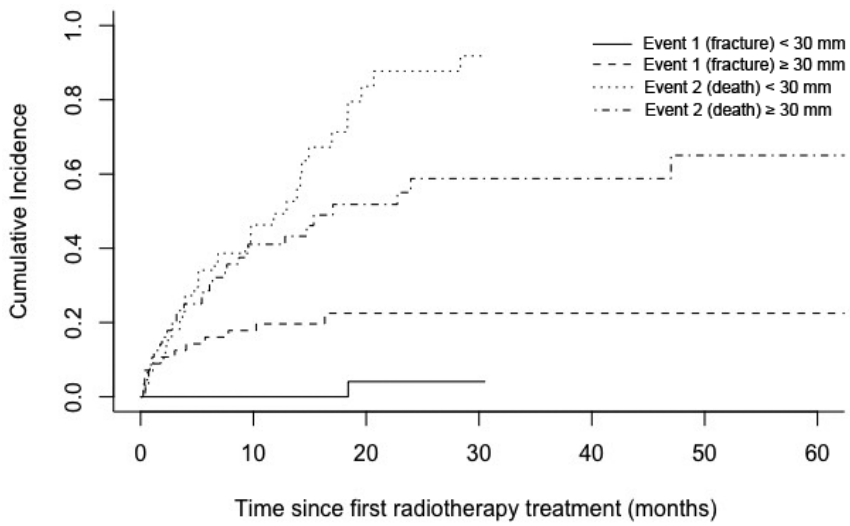
Supplementary material



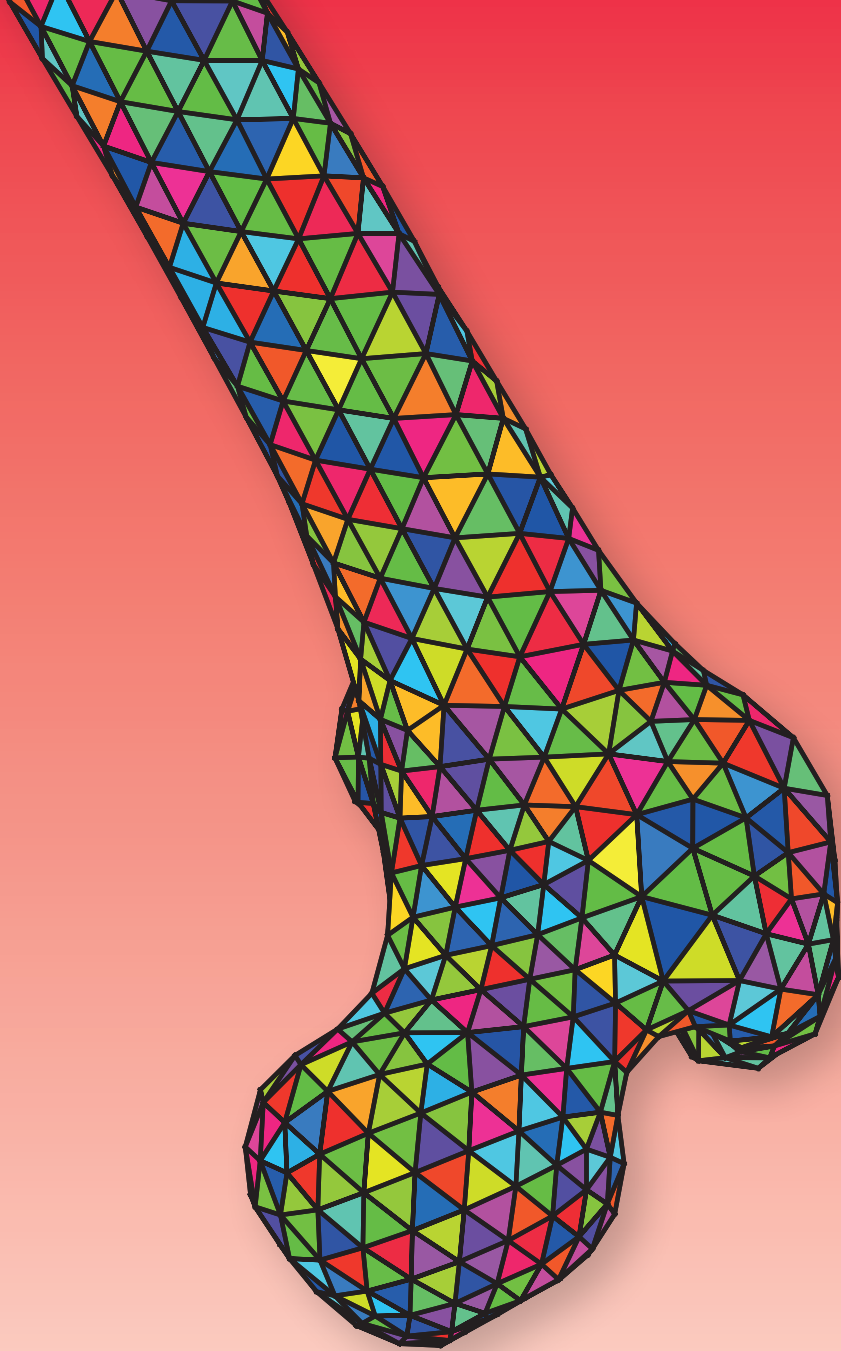
**Supplementary Data 1.** Potential states of a patient. In the initial state, the patient is alive and event-free (0). The first event may be cortical fracture (event of interest 1) or death before the patient experienced cortical fracture (competing event 2).



**Supplementary Data 2.** Estimated cumulative incidence of the competing events



**Supplementary Data 3.** Cumulative incidence for the two competing events for cortical involvement



# CHAPTER 3

## LIMITED SHORT-TERM EFFECT OF PALLIATIVE RADIATION THERAPY ON QUANTITATIVE COMPUTED TOMOGRAPHY-DERIVED BONE MINERAL DENSITY IN FEMORA WITH METASTASES



---

Florieke Eggermont, Loes Derikx, Nico Verdonchot, Gerjon Hannink, Robert Kaatee, Esther Tanck,  
Yvette van der Linden

Advances in Radiation Oncology 2017; 2(1): 53-61.



## Introduction

Bone metastases are most frequently seen in patients with primary tumors in the breast, prostate, lung, kidney, or thyroid.<sup>1,2</sup> These lesions can have a lytic, blastic, or mixed radiological appearance. Lytic lesions result from disproportionate bone resorption by osteoclasts and leads to progressive destruction of the bone tissue and a subsequent risk of pathological fracturing.<sup>1,3</sup> Blastic lesions are characterized by excessive bone formation and are hypothesized to decrease bone strength because the newly formed bone has a reduced structural integrity.<sup>1,3</sup>

External beam radiation therapy (RT) plays an important role in the palliative care of patients with bone metastases because it relieves pain.<sup>4-7</sup> Additionally, some studies report a beneficial effect of RT on bone mineral density (BMD),<sup>8-12</sup> which is also the clinical experience of medical specialists. In contrast, this beneficial effect was not confirmed in a recent systematic review.<sup>13</sup> Moreover, the relationship between pain relief and BMD is unclear. Some authors suggest there is no relationship,<sup>7,11</sup> whereas others state that improved BMD contributes to long-lasting pain relief,<sup>14</sup> increased bone stability, and decreased risk of fractures.<sup>15</sup>

In patients with cancer and bone metastases in the femur, the ability to walk and remain mobile is very important for the overall quality of life. Therefore, it is important to assess not only pain but also the risk for fractures when determining RT dose schedules. If the expected risk for fractures is low, RT is administered in one, relatively high dose (single fraction (SF)) to relieve pain. Patients with a high expected risk for fractures who are not eligible for or do not want surgery can receive RT in multiple fractions (MF) to induce remineralization and prevent pathological fracturing.<sup>16,17</sup> Previous research suggested that 24 Gy in 4 fractions postponed pathological fractures when compared with a single dose of 8 Gy.<sup>17</sup> However, studies on the effect of RT doses on BMD are limited, and the reported responses differ between studies.<sup>8-12</sup> To date, only Koswig *et al.* compared SF and MF in terms of BMD and found a greater response after MF RT.<sup>8</sup>

Most studies included an analysis of lytic lesions in the vertebrae<sup>8-11</sup> but the effect of RT on BMD in blastic or mixed lesions was often not considered. Although the femur is also frequently affected,<sup>18,19</sup> few studies analyzed the effect of RT on BMD in femoral lesions.<sup>8,9,12</sup> Also, little is known about the effect of RT on BMD within the entire field of RT as it relates the femur.

Therefore, the aim of this study is to determine the effect of SF and MF RT on BMD in patients with cancer and femoral bone metastases. Specifically, we studied 2 regions of interest: the proximal femur within the radiation field (region of interest (ROI)-PF) and the metastatic lesion (ROI-ML). For both regions, we investigated whether there was an overall effect of SF and MF RT on BMD and whether these effects were different in the femora among lytic, blastic, and mixed lesions.

## Methods

### Patients

Between 2006 and 2009, patients who received palliative RT for femoral metastases in 3 RT institutes in The Netherlands (Radiotherapeutic Institute Friesland, Leiden University Medical Center and Radboud University Medical Center) were asked to participate in this prospective study. Institutional approval was obtained from the ethics committees of all participating centers. This study is part of a larger study on the prediction of fracture risks with use of Finite Element modeling.<sup>20,21</sup>

Patients received either SF ( $1 \times 8$  Gy) or MF (5 or  $6 \times 4$  Gy) RT in accordance with the Dutch clinical guidelines that state that lesions with cortical involvement of more than 3 cm have an increased risk of fracture and will be considered for prophylactic surgery.<sup>16,17</sup> SF is typically applied to treat pain that is related to smaller, uncomplicated lesions with a low expected risk for fracture and has a 60% to 80% chance of pain relief.<sup>4-7</sup> In patients who have larger lesions (i.e., in principle requiring stabilization) and refuse surgery or have a deteriorating clinical condition, radiation oncologists typically prescribe a higher total dose to induce remineralization. Surgery may be too hazardous for these patients; hence, higher radiation doses are selected with the hope that they prevent pathologic fracturing.

Patients were included in the study if they had a Karnofsky Performance Score of  $>60$ , no clinical or radiologic evidence of pathologic femoral fractures, no planned or prior palliative surgery to the femur, no radionuclide treatment 30 days prior to inclusion, and no previous RT to the femur. In total, 66 patients gave written informed consent. With use of follow-up questionnaires after 4 and 10 weeks and after 4, 5, and 6 months, patients were actively followed for 6 months or until they sustained a femoral fracture or died. At the 6-month follow-up, the study database was updated on the basis of hospital records and then closed.

### Measurements and follow-up

Baseline patient characteristics including sex, age, body weight, Karnofsky Performance Score, time since primary tumor diagnosis and since first metastasis, primary tumor, and concurrent systemic therapy were registered by the treating radiation oncologist at the time of intake. During the RT planning session, patients underwent their first quantitative computed tomography (QCT) scan. Subsequent QCT scans were taken after 4 and 10 weeks. Patients who received MF RT also underwent an additional QCT scan after 1 week on the final day of their RT schedule to determine a potential immediate effect of RT on BMD.<sup>8,10</sup> At the same time points, patients completed questionnaires on pain (ie, Brief Pain Inventory<sup>22</sup>), level of activity, and quality of life (including sections from the Longitudinal Aging Study Amsterdam Physical Activity Questionnaire,<sup>23</sup> Short Form-36,<sup>24</sup> and the Western Ontario and McMaster Universities Arthritis Index<sup>25</sup>). These patient-reported outcomes will be published separately.

The institutes were instructed to perform the QCT scans in accordance with a standardized protocol with the following settings: 120 kVp, 220 mA, slice thickness 3 mm, pitch 1.5, spiral and standard reconstruction, and in-plane resolution 0.9375 mm. The protocol required scanning of at least the

proximal half of the femur, including the painful metastases. A solid calibration phantom (Image Analysis, Columbia, KY) that contained 4 known calcium hydroxyapatite (CaHA) concentrations of 0, 50, 100, and 200 mg/cm<sup>3</sup> was placed under the patient in the scanner. The densities in this phantom were used to calibrate each Hounsfield unit (HU) to CaHA density. This CaHA density is a calcium-equivalent density that is a measure of BMD; in the remainder of this work, we will refer to it as BMD.

### Registration

To analyze the effect of RT on BMD of the proximal femur region (ROI-PF) over time, the proximal half of the femur for each patient was segmented from one of the QCT scans (Mimics 11.0 and 14.0, Materialise, Leuven, Belgium).

Patients' QCT scans were registered with fully automated rigid registration algorithms for medical images (elastix<sup>26,27</sup>). For this, numerous alignments of 2 CT scans were calculated until the best fit was found,<sup>28</sup> resulting in an objective and accurate registration (Figure 1A and B). All voxels that represent the femoral geometry were included in the analyses.

Furthermore, an experienced radiation oncologist segmented the lesions and scored each metastatic femur as lytic, blastic, or mixed on the first QCT scan. To account for obscure edges, regions of metastatic lesions (ROI-ML) were defined by expanding the lesions by 6 mm in all directions. If the procedure included voxels outside the femur, these voxels were ignored (Figure 1C).

### Analysis

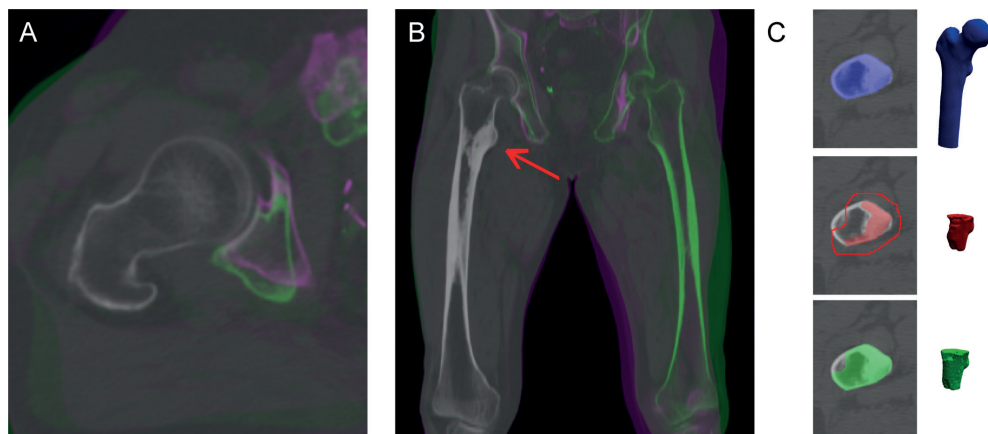
Baseline characteristics were compared between patients who received SF and MF and patients with lytic, blastic, and mixed lesions with use of the Mann-Whitney U, Fisher exact, Pearson  $\chi^2$ , and Kruskal-Wallis tests, where applicable.

Mean BMD in mg/cm<sup>3</sup> was calculated for each ROI-PF and ROI-ML at each time point. Linear mixed models were used to study the effect of RT on the BMD of all ROI-PF and ROI-ML over time. This analysis allowed for the inclusion of patients with missing data. Lesion type and size were added to the models. To address a potential effect of confounding by indication, we tested whether other baseline characteristics (eg, performance score, primary tumor, concurrent systemic therapy (anticancer and/or bisphosphonates)) also affected BMD. However, none of these other baseline characteristics influenced the effect of RT on BMD, and they were removed from the final models.<sup>29</sup>

A random intercept was included to disregard the variability in initial BMD between patients. The interaction between lesion type and time was significant and therefore added to the model. All other interactions were not significant. *P*-values below 0.05 were considered statistically significant. Statistical analyses were performed using Stata/SE 11.2 (StataCorp LP, College Station, TX).

The statistical model tested BMD in mg/cm<sup>3</sup> but not in percentages. However, for interpretational and visual purposes, the data were converted to percentages by marking BMD on the first QCT scan as 100% and calculating BMD of the subsequent QCT scans relative to the first measurement.





**Figure 1:** Transversal (A) and sagittal (B) false color overlay of two registered quantitative computed tomography scans of the right femur, showing an accurate registration of the right femur (no green/purple color visible) and inaccurate registration of the pelvis and left femur (green/purple colors visible). The red arrow indicates the lesion. (C) Segmented geometry of a proximal femur (ROI-PF, blue), lesion (red), and 6-mm margin (line), and the region of interest that was used for analysis of the lesion (ROI-ML, green).

## Results

### Patients

Of the 66 eligible patients, 24 were excluded from this analysis because only 1 QCT scan was available ( $n = 20$ ), the first QCT scan was missing ( $n = 1$ ), or lesions were unidentifiable ( $n = 3$ ). Hence, 42 patients were included for analysis, 5 of whom received RT to both femora, leading to a total of 47 femora for analysis. Three femora had 2 separate lesions and 1 femur had 3 lesions, which resulted in 52 lesions and comprised 24 lytic, 8 blastic, and 20 mixed lesions.

Not every patient underwent all QCT scans at all time points because of death, fracture, or deteriorating condition. At baseline, all 42 QCT scans were obtained, but after 4 and 10 weeks, only 30 and 27 QCT scans were acquired, respectively. Of the 26 patients who receive MF RT, 25 underwent a QCT scan on the final day of RT. Against protocol instructions, less than half of the femur was scanned in 12 cases (range, 41%-49% of the femoral length). Additionally, 8 ROI-PF were larger than 50% to include lesions in the distal half of the femur (range, 55%-89% of the femoral length).

Baseline characteristics were not significantly different between patients who received SF or MF RT, but sex, age, and primary tumor differed among patients with lytic, blastic, or mixed lesions (Table 1).

Table 1: Patient baseline characteristics.

	Single fraction (1x8Gy) n=16	Multiple fractions (5-6x4Gy) n=26	p-value	Lytic n=17	Blastic n=8	Mixed n=17	p-value
<b>Gender<sup>a</sup></b>							
Male	12 (75%)	16 (62%)	0.5	7 (41%)	7 (87%)	14 (82%)	0.02
Female	4 (25%)	10 (38%)		10 (59%)	1 (13%)	3 (18%)	
<b>Age in years<sup>b</sup></b>							
Median (Range)	68.5 (39 – 89)	66 (52 – 85)	0.8	62 (39-84)	69 (61-75)	70 (46-89)	0.5
<b>Body weight in kg<sup>b</sup></b>							
Median (Range)	75 (44 – 92)	79 (57 – 106)	0.5	73 (57-106)	77.5 (44-83)	84.5 (56-95)	0.5
<b>Performance score(KPS *)<sup>b</sup></b>							
Median (Range)	80 (60 – 100)	80 (60 – 100)	0.1	80 (60-90)	80 (70-90)	80 (60-100)	0.3
<b>Time since primary tumor diagnosis in years<sup>b</sup></b>							
Median (Range)	4.0 (0.4 – 17.6)	3.0 (0.1 – 23.8)	0.6	1.9 (0.1-23.8)	5.9 (0.2-12.3)	2.8 (0.4-17.6)	0.3
<b>Time since first metastasis in years<sup>b</sup></b>							
Median (Range)	1.2 (0 – 7.3)	1.0 (0.3 – 7.8)	0.8	0.8 (0-7.8)	2.6 (0-7.3)	1.2 (0-5.4)	0.004
<b>Primary tumor<sup>c</sup></b>							
Breast	3 (19%)	5 (19%)	0.6	5 (30%)	1 (13%)	2 (12%)	
Lung	1 (6%)	5 (19%)		4 (23%)	1 (13%)	1 (6%)	
Prostate	9 (56%)	10 (38%)		1 (6%)	6 (74%)	12 (70%)	
Other <sup>c</sup>	3 (19%)	6 (23%)		7 (41%)	0 (0%)	2 (12%)	
<b>Lesion type<sup>c</sup></b>							
Lytic	4 (25%)	13 (50%)	0.2				
Blastic	3 (19%)	5 (19%)					
Mixed	9 (56%)	8 (31%)					
<b>Affected femur<sup>d</sup></b>			1				0.8
Unilateral	14 (87%)	23 (89%)		14 (82%)	8 (100%)	15 (88%)	
Bilateral	2 (13%)	3 (11%)		3 (18%)	0 (0%)	2 (12%)	
<b>Concurrent systemic therapy<sup>e</sup></b>			0.6				0.2
No concurrent therapy	1 (6%)	5 (19%)		5 (29%)	1 (13%)	0 (0%)	
Systemic therapy + Bisphosphonates	4 (25%)	6 (23%)		3 (18%)	1 (13%)	6 (35%)	
Systemic therapy - Bisphosphonates	10 (63%)	12 (46%)		7 (41%)	5 (61%)	10 (59%)	
Bisphosphonates only	0 (0%)	0 (0%)		0 (0%)	0 (0%)	0 (0%)	
Missing	1 (6%)	3 (12%)		2 (12%)	1 (13%)	1 (6%)	

\*KPS = Karnofsky Performance Score. <sup>a</sup>Other = Kidney, Rectum, Kahler's disease, Urethra, Cervix or aCUP (cancer of unknown primary origin). <sup>b</sup>Tested with Fisher's Exact.<sup>c</sup>Tested with Mann-Whitney U for differences between single and multiple fractions and with Kruskal-Wallis for differences between lytic, blastic and mixed lesions. <sup>d</sup>Tested with Pearson's Chi-Square.

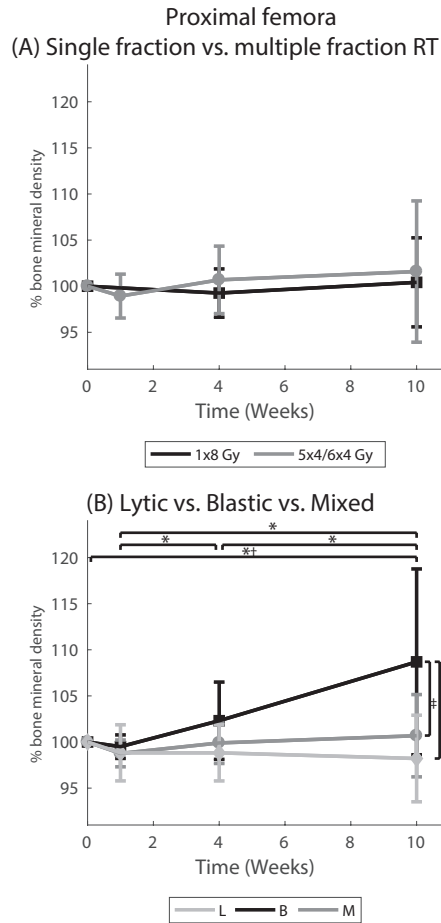
**Bone mineral density**

Table S-1 depicts the mean BMD (in  $\text{mg}/\text{cm}^3$ ) of the proximal femora (ROI-PF) and metastatic lesions (ROI-ML) from the scans that were available (Supplementary Material).

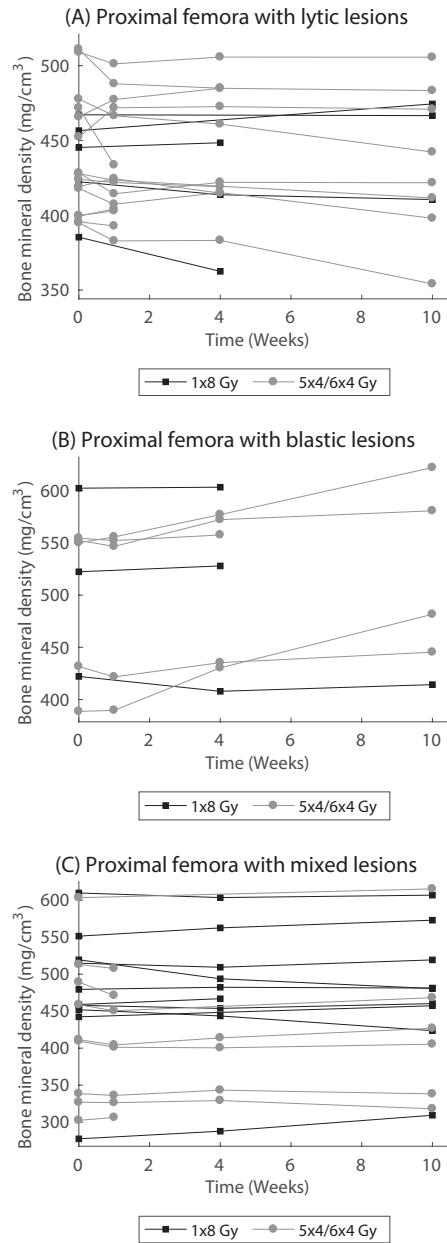
***Effect of RT on BMD in proximal femur (ROI-PF)***

At baseline, mean BMD in ROI-PF was  $471.5 \text{ mg}/\text{cm}^3$  (standard deviation (SD),  $77.4 \text{ mg}/\text{cm}^3$ ) in patients who were assigned to SF RT and  $445.7 \text{ mg}/\text{cm}^3$  (SD,  $70.7 \text{ mg}/\text{cm}^3$ ) in patients who were assigned to MF RT. After 10 weeks, BMD increased 0% after SF and 2% after MF RT and was not different between SF and MF RT over all time points (Figure 2A). An interaction was found between lesion type and time, which indicates that femora affected by different lesion types responded diversely to RT over time (Figure 2B). This lesion-dependent effect over time holds for both radiation schedules, as there was no difference in BMD between ROI-PF treated with SF and MF RT.

Independent of RT schedule, ROI-PF that included lytic lesions showed a significant decrease of 2% in BMD between QCT scans at baseline and after 10 weeks, ( $-9.2 \text{ mg}/\text{cm}^3$ , 95% confidence interval (CI),  $-18.0$ – $-0.4$ ,  $P = 0.04$ ). An increase in BMD to 109% at 10 weeks ( $37.9 \text{ mg}/\text{cm}^3$ , 95% CI  $24.7$ – $51.0$ ,  $P < 0.001$ ) was observed in ROI-PF that contained blastic lesions. No significant differences over time were found for ROI-PF with mixed lesions. Furthermore, 10 weeks after RT, BMD in ROI-PF that contained blastic lesions was significantly higher than in ROI-PF that contained lytic ( $106.4 \text{ mg}/\text{cm}^3$ , 95% CI  $39.1$ – $173.7$ ,  $P = 0.002$ ) or mixed lesions ( $87.2 \text{ mg}/\text{cm}^3$ , 95% CI  $24.8$ – $149.6$ ,  $P = 0.006$ ). Figure 3 depicts the BMD of all ROI-PF and shows a widespread individual response to RT.



**Figure 2:** Mean  $\pm$  standard deviation bone mineral density (in percentage relative to quantitative computed tomography scan 1) of all proximal femur regions (ROI-PF) over time, for (A) single fraction ( $1 \times 8$  Gy) versus multiple fractions ( $5 \times 4 \times 4$  Gy), and (B) each lesion type. \*Significant difference for blastic lesions. †Significant difference for lytic lesions. ‡Significant difference at 10 weeks.



**Figure 3:** Mean bone mineral densities (in  $\text{mg}/\text{cm}^3$ ) of the proximal femur regions (ROI-PF) that contain lytic (A), blastic (B), or mixed (C) lesions for each femur over time.

**Effect of RT on BMD in metastatic lesion (ROI-ML)**

On average, ROI-ML volume was 88 cm<sup>3</sup> (SD, 61 cm<sup>3</sup>). Volumes of lytic (mean, 58 cm<sup>3</sup>; SD, 44 cm<sup>3</sup>), blastic (mean, 143 cm<sup>3</sup>; SD, 78 cm<sup>3</sup>), and mixed (mean, 103 cm<sup>3</sup>; SD, 53 cm<sup>3</sup>) ROI-ML were significantly different ( $P = .004$ ).

At baseline, the mean BMD in ROI-ML treated with SF RT was 474.9 mg/cm<sup>3</sup> (SD, 150.1 mg/cm<sup>3</sup>), and 394.6 mg/cm<sup>3</sup> (SD, 190.3 mg/cm<sup>3</sup>) when treated with MF RT. Mean BMD of ROI-ML increased by 1% in SF and 7% in MF RT after 10 weeks, but this was not significantly different (Figure 4A). A significant interaction between lesion type and time was found, indicating that the effect of RT on BMD was different for the 3 lesion types (Figure 4B).

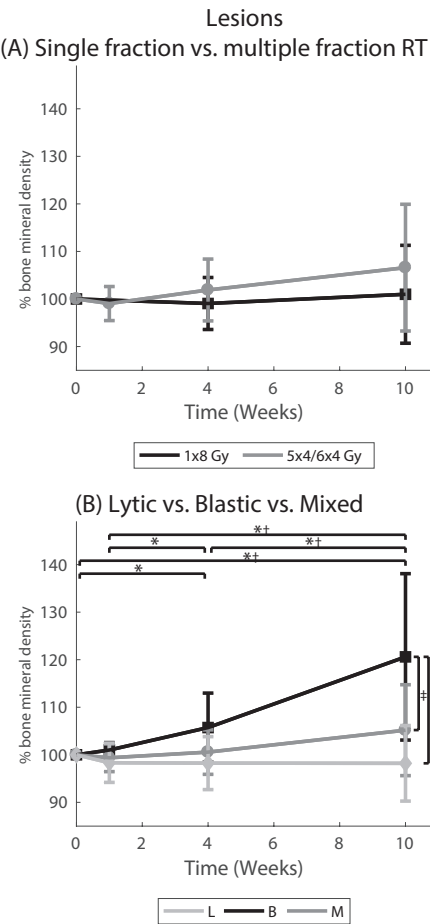
No changes in BMD were observed in lytic ROI-ML over time. BMD in blastic and mixed ROI-ML significantly increased by 21% and 5%, respectively, between baseline and 10 weeks, (blastic: 72.8 mg/cm<sup>3</sup>, 95% CI 50.5-95.0,  $P < 0.001$ ; mixed: 22.0 mg/cm<sup>3</sup>, 95% CI 9.2-34.9,  $P = 0.001$ ).

At baseline, BMD in blastic ROI-ML was significantly higher than in lytic (161.4 mg/cm<sup>3</sup>; 95% CI, 6.6-316.2;  $P = 0.04$ ) and mixed ROI-ML (150.8 mg/cm<sup>3</sup>; 95% CI, 7.2-294.5;  $P = 0.04$ ). This difference remained significant and increased over time. Figure 5 shows the effect of RT on BMD in every ROI-ML and illustrates each ROI-ML responding differently to RT.

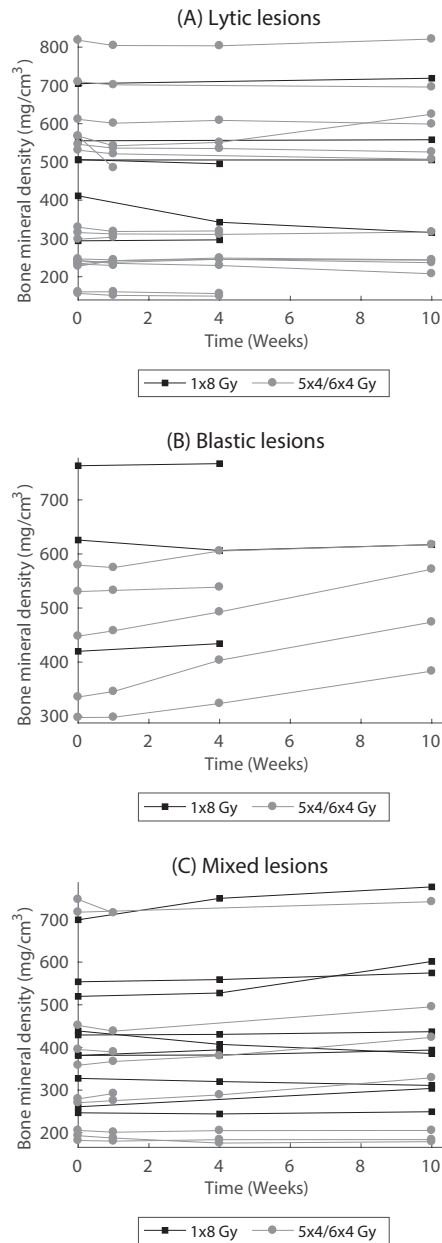
**Discussion**

The aim of this study was to determine the overall effect of palliative RT on bone mineral density in the proximal femur and metastatic lesions in patients with cancer and painful bone metastases. Additionally, we investigated whether these effects were different in femora with lytic, blastic, and mixed lesions.

In the proximal femora regions (ROI-PF), no differences in BMD were found between SF and MF RT. BMD decreased in ROI-PF that contained lytic lesions, increased in ROI-PF with blastic lesions, and did not change in ROI-PF with mixed lesions over time. After 10 weeks, differences in BMD in the radiation field (ROI-PF) were smaller than those in the metastatic lesions. This may indicate that the effect in ROI-PF was mainly due to BMD changes in the lesions (ROI-ML) and suggests that the irradiated femoral bone inside the radiation field but outside the lesion is unaffected or less affected by RT. A few studies found a smaller BMD increase in irradiated normal-appearing bone surrounding lytic metastases in vertebrae compared with the regions with vertebral lesions.<sup>10,11</sup> However, to our knowledge, the effect of RT on total bone volume in the radiation field has not been studied previously. Moreover, previous studies did not include blastic or mixed lesions.



**Figure 4:** Mean  $\pm$  standard deviation bone mineral density (in percentage relative to quantitative computed tomography scan 1) of the metastatic lesions (ROI-ML) over time for (A) single fraction ( $1 \times 8$  Gy) versus multiple fractions ( $5-6 \times 4$  Gy), and (B) each lesion type. \*Significant difference for blastic lesions. †Significant difference for mixed lesions. ‡Significant difference at 10 weeks.



**Figure 5:** Mean bone mineral densities (in  $\text{mg}/\text{cm}^3$ ) of the lytic (A), blastic (B), and mixed (C) metastatic lesions (ROI-ML) for each lesion over time.

When focusing on ROI-ML, BMD did not differ between SF and MF RT. Additionally, BMD in lytic ROI-ML did not change, which contradicts the findings in previous work.<sup>8-12</sup> Koswig *et al.*<sup>8</sup> observed a decrease



in BMD immediately after RT for both SF ( $1 \times 8$  Gy) and MF ( $10 \times 3$  Gy), followed by an increase in BMD after 6 months of 120% for SF and 173% for MF. After 10 weeks (comparable with our last time point), BMD increased approximately 106% and 137%, respectively. A similar response after MF (40 Gy) RT in vertebral lesions was observed by Reinbold *et al.*<sup>10</sup> who showed a decrease of 20% in BMD at the end of RT, followed by an increase of more than 60% after 3 months.<sup>10</sup> The observation that RT ultimately increases BMD in lytic lesions is supported by other studies.<sup>9,11</sup> Actually, Koswig *et al.* only observed significant differences in lesions that originated from breast cancer,<sup>8</sup> which are known to be responsive to RT.<sup>30</sup> When considering only metastases that arise from breast cancer in our study, no differences between SF and MF RT were seen. Hence, a beneficial effect of MF RT over SF RT on BMD was not observed in our study.

Our results, which contrast with those of previous studies, have several possible explanations. First, it should be noted that concurrent treatment with bisphosphonates may enhance the effect of RT on BMD.<sup>11-13</sup> Although we did not find any interaction with medication, it may potentially have caused biased effects in the earlier studies that did not test this.<sup>8,10</sup> Second, the radiation doses in the previous studies were typically higher compared with the doses administered in our study. We administered doses of a total of 8 or 20 to 24 Gy, whereas other studies applied doses of up to 40 Gy.<sup>10-12</sup> Third, we included metastatic lesions that originated from various primary tumors, some of which are known to be less responsive to RT than others.<sup>8,30</sup> In addition, most other studies included no or few femoral bone metastases but mostly vertebral and pelvic metastases, which are suggested to have a better BMD response to RT compared with metastases in the extremities.<sup>8,30</sup> Only one study included solely femoral metastases, and density changes in lesions were evaluated on the basis of X-ray test results with a reported response in 42% of patients. However, the study's follow-up ranged from 1 to 28 months, and higher response rates were found in patients who were followed for a longer period of time.<sup>12</sup> The follow-up period in our study may have been too short to determine the long-term effects of RT on BMD. Finally, the most relevant difference between earlier studies and the current one is probably the detailed QCT-approach. We studied lesional volumes 3-dimensionally, which provided a more extensive analysis of RT effects on metastatic lesions. We performed several sensitivity analyses that proved that our 3-dimensional image registration was accurate. All previous research studied 2-dimensional ROI on the basis of X-ray test results<sup>12</sup> or single CT scan image,<sup>8-11</sup> and temporal registration was accomplished by reproducing the patient position on the CT scanner<sup>8-10</sup> or drawing ROI in each scan by hand.<sup>11</sup> It can be questioned whether the same accuracy can be obtained with manual registration compared with our fully automated registration. The latter is not dependent on the arbitrary position of a limited number of landmarks but uses the overlap of a large number of voxels that are taken from the different images; therefore, they should be better than manual registration.<sup>28</sup> For these reasons, we consider our study results to be reliable.

This study also has some limitations. As previously shown<sup>31</sup>, accurate identification of the margins of the actual bone lesions was difficult. Therefore, we added a rim of 6 mm around the edges to decrease the chance of omitting lesional tissue in the analysis, even though this may include nonaffected bone tissue. Also, categorization of lesions into pure lytic, pure blastic, and a mixed-type category using CT

scans was complicated. Although the lesions were categorized in accordance with guidelines,<sup>32</sup> some caution should be taken when interpreting differences between lesion types. Furthermore, the total number of patients included in the study was limited, and not all QCT scans were acquired for each patient. However, the main reasons for missing scans were death, fracture, and deteriorating condition, which is inevitable when analyzing data from patients with cancer who are in the palliative phase of their disease.

It is difficult to extrapolate BMD effect to femoral bone strength. Mean BMD did not change in lytic ROI-ML, which indicates that there was no progression of disease or remineralization of the bone tissue; hence, there was no change in bone strength. In contrast, BMD in blastic and mixed ROI-ML increased over time. However, the effect of these BMD increases on bone strength is difficult to interpret because denser blastic lesions could flag either disease progression or formation of new high-density bone tissue. Additionally, in mixed lesions, both blastic and lytic processes occur. These processes may cancel out a potential effect on bone mineral density. Therefore, the way changes in BMD affect femoral bone strength in blastic and mixed lesions should be investigated further.

## Conclusions

In conclusion, higher total RT doses in patients with cancer and femoral metastases did not lead to significantly higher BMD up to 10 weeks after palliative RT, which brings into question the clinical relevance of MF over SF to stabilize femoral bone within this time period. Additionally, 10 weeks after RT, a significant increase in BMD was observed in blastic and mixed lesions but not in lytic lesions. Whether this implies progression or remineralization is unclear, especially since there was no control group of patients who received no RT. Also, the subsequent clinical effect of these changes on femoral bone strength remains unknown and needs to be investigated in the future.

## Acknowledgements

The authors would like to thank Tom Knoop and Wouter Gevers for the development of the framework for registration and analysis, and the Department of Radiation Oncology in Nijmegen, RIF Leeuwarden and the Department of Clinical Oncology in Leiden for their help.

This project was funded by the Dutch Cancer Society (KUN 2012-5591) and the Dutch Science Foundation NWO-STW (NPG.06778).

## References

1. Coleman RE. 1997. Skeletal complications of malignancy. *Cancer* 80:1588-1594.
2. Coleman RE. 2006. Clinical features of metastatic bone disease and risk of skeletal morbidity. *Clin Cancer Res* 12:6243s-6249s.
3. Healey JH, Brown HK. 2000. Complications of bone metastases: surgical management. *Cancer* 88:2940-2951.
4. Chow E, Zeng L, Salvo N, Dennis K, Tsao M, Lutz S. 2012. Update on the systematic review of palliative radiotherapy trials for bone metastases. *Clin Oncol* 24:112-124.
5. Lutz S, Berk L, Chang E, Chow E, Hahn C, Hoskin P, Howell D, Konski A, Kachnic L, Lo S, Sahgal A, Silverman L, von Gunten C, Mendel E, Vassil A, Bruner DW, Hartsell W, American Society for Radiation O. 2011. Palliative radiotherapy for bone metastases: an ASTRO evidence-based guideline. *Int J Radiat Oncol Biol Phys* 79:965-976.
6. Bedard G, Hoskin P, Chow E. 2014. Overall response rates to radiation therapy for patients with painful uncomplicated bone metastases undergoing initial treatment and retreatment. *Radiother Oncol* 112:125-127.
7. van der Linden Y, Roos D, Lutz S, Fairchild A. 2009. International variations in radiotherapy fractionation for bone metastases: geographic borders define practice patterns? *Clin Oncol* 21:655-658.
8. Koswig S, Budach V. 1999. Remineralisation und Schmerzlinderung von Knochenmetastasen nach unterschiedlich fraktionierter Strahlentherapie (10 mal 3 Gy vs. 1 mal 8 Gy). Eine prospektive Studie. *Strahlenther Onkol* 175:500-508.
9. Chow E, Holden L, Rubenstein J, Christakis M, Sixel K, Vidmar M, Finkelstein J, Hayter C, Loblaw A, Wong R, Szumacher E, Danjoux C. 2004. Computed tomography (CT) evaluation of breast cancer patients with osteolytic bone metastases undergoing palliative radiotherapy--a feasibility study. *Radiother Oncol* 70:291-294.
10. Reinbold WD, Wannenmacher M, Hodapp N, Adler CP. 1989. Osteodensitometry of vertebral metastases after radiotherapy using quantitative computed tomography. *Skeletal Radiol* 18:517-521.
11. Foerster R, Eisele C, Bruckner T, Bostel T, Schlamp I, Wolf R, Debus J, Rief H. 2015. Bone density as a marker for local response to radiotherapy of spinal bone metastases in women with breast cancer: a retrospective analysis. *Radiat Oncol* 10:368.
12. Harada H, Katagiri H, Kamata M, Yoshioka Y, Asakura H, Hashimoto T, Furutani K, Takahashi M, Sakahara H, Nishimura T. 2010. Radiological response and clinical outcome in patients with femoral bone metastases after radiotherapy. *J Radiat Res* 51:131-136.
13. Groenen KH, Pouw MH, Hannink G, Hosman AJ, van der Linden YM, Verdonschot N, Tanck E. 2016. The effect of radiotherapy, and radiotherapy combined with bisphosphonates or RANK ligand inhibitors on bone quality in bone metastases. A systematic review. *Radiother Oncol* 119:194-201.
14. Saarto T, Janes R, Tenhunen M, Kouri M. 2002. Palliative radiotherapy in the treatment of skeletal metastases. *Eur J Pain* 6:323-330.
15. Smith HS. 2011. Painful osseous metastases. *Pain Physician* 14:E373-403.
16. Van der Linden YM, Dijkstra PD, Kroon HM, Lok JJ, Noordijk EM, Leer JW, Marijnen CA. 2004. Comparative analysis of risk factors for pathological fracture with femoral metastases. *J Bone Joint Surg Br* 86:566-573.

17. Van der Linden YM, Kroon HM, Dijkstra SP, Lok JJ, Noordijk EM, Leer JW, Marijnen CA. 2003. Simple radiographic parameter predicts fracturing in metastatic femoral bone lesions: results from a randomised trial. *Radiother Oncol* 69:21-31.
18. Steenland E, Leer JW, van Houwelingen H, Post WJ, van den Hout WB, Kievit J, de Haes H, Martijn H, Oei B, Vonk E, van der Steen-Banasik E, Wiggenraad RG, Hoogenhout J, Warlam-Rodenhuis C, van Tienhoven G, Wanders R, Pomp J, van Reijn M, van Mierlo I, Rutten E. 1999. The effect of a single fraction compared to multiple fractions on painful bone metastases: a global analysis of the Dutch Bone Metastasis Study. *Radiother Oncol* 52:101-109.
19. Toma CD, Dominkus M, Nedelcu T, Abdolvahab F, Assadian O, Krepler P, Kotz R. 2007. Metastatic bone disease: a 36-year single centre trend-analysis of patients admitted to a tertiary orthopaedic surgical department. *J Surg Oncol* 96:404-410.
20. Derikx LC, van Aken JB, Janssen D, Snyers A, van der Linden YM, Verdonschot N, Tanck E. 2012. The assessment of the risk of fracture in femora with metastatic lesions: comparing case-specific finite element analyses with predictions by clinical experts. *J Bone Joint Surg Br* 94:1135-1142.
21. Derikx LC, Groenen K, van Bon GA, van der Linden YM, Snyers A, Verdonschot N, Tanck E. 2011. Patient-specific finite element models discriminate between patients with and without a pathological fracture in metastatic bone disease. 57th Annual Meeting of the Orthopaedic Research Society. Long Beach, California, United States.
22. Fairbank JC, Couper J, Davies JB, O'Brien JP. 1980. The Oswestry low back pain disability questionnaire. *Physiotherapy* 66:271-273.
23. Stel VS, Smit JH, Pluijm SM, Visser M, Deeg DJ, Lips P. 2004. Comparison of the LASA Physical Activity Questionnaire with a 7-day diary and pedometer. *J Clin Epidemiol* 57:252-258.
24. Ware JE, Jr., Sherbourne CD. 1992. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 30:473-483.
25. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. 1988. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol* 15:1833-1840.
26. Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. 2010. elastix: A Toolbox for Intensity-Based Medical Image Registration. *IEEE Trans Med Imaging* 29:196-205.
27. Shamonin DP, Bron EE, Lelieveldt BPF, Smits M, Klein S, Staring M, In AsDN. 2014. Fast parallel image registration on CPU and GPU for diagnostic classification of Alzheimer's disease. *Frontiers in Neuroinformatics* 7.
28. Knoop TH, Derikx LC, Verdonschot N, Slump CH. 2015. A novel framework for the temporal analysis of bone mineral density in metastatic lesions using CT images of the femur. *SPIE Medical Imaging: International Society for Optics and Photonics*; pp. 94143A-94143A-94111.
29. Budtz-Jorgensen E, Keiding N, Grandjean P, Weihe P. 2007. Confounder selection in environmental epidemiology: assessment of health effects of prenatal mercury exposure. *Ann Epidemiol* 17:27-35.
30. Rieden K, Adolph J, Lellig U, zum Winkel K. 1989. [The radiotherapeutic effect on bone metastases in relation to the frequency of metastases, sites of

metastases and histology of the primary tumor].  
Strahlenther Onkol 165:380-385.

31. Dijkstra PD, Oudkerk M, Wiggers T. 1997.  
Prediction of pathological subtrochanteric  
fractures due to metastatic lesions. Arch Orthop  
Trauma Surg 116:221-224.
32. Mulder JD, Kroon HM, Schutte HE, Taconis WK.  
1993. The diagnosis of bone tumours. Radiologic  
atlas of bone tumours. Amsterdam: Elsevier  
publishers; pp. 28-31.



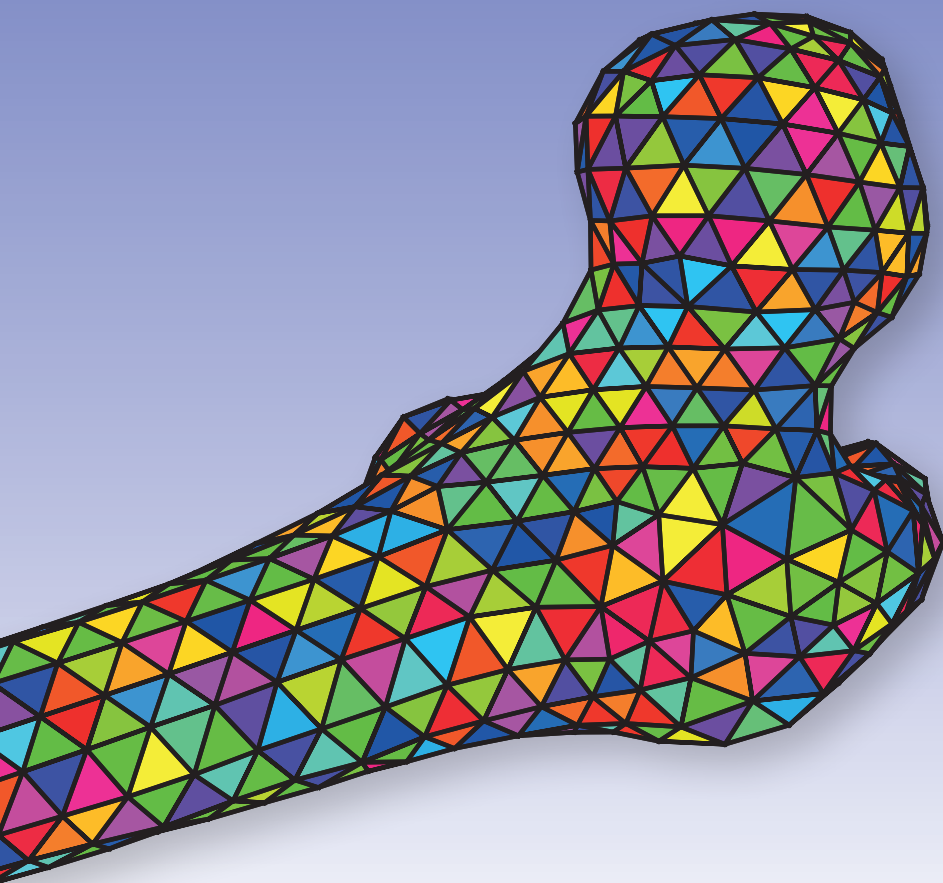
Supplementary material

Table S-1: Mean bone mineral densities (BMD, in mg/cm<sup>3</sup>) of all proximal femora and lesions over time

Femur	Femur	RT	Lesion	Primary tumor	Proximal femora (ROI-PF)			Metastatic lesions (ROI-ML)				
					QCT1	QCT2	QCT3	QCT1	QCT2	QCT3	QCT4	QCT4
113R	P01R	1x8 Gy	Lytic	Breast	445	-	448	294	-	296	x	x
120L	P02L	1x8 Gy	Lytic	Breast	457	-	x	505	-	x	505	505
120L*	P02L*	1x8 Gy	Lytic	Breast	*	-	*	705	-	x	719	719
120R	P02R	1x8 Gy	Lytic	Breast	467	-	x	555	-	x	558	558
236R	P03R	1x8 Gy	Lytic	Lung	422	-	414	411	-	342	315	315
102L	P04L	1x8 Gy	Lytic	Kahler's disease	385	-	363	506	-	495	x	x
217L	P05L	1x8 Gy	Blastic	Breast	602	-	603	763	-	767	x	x
106R	P06R	1x8 Gy	Blastic	Prostate	422	-	408	626	-	606	617	617
229R	P07R	1x8 Gy	Blastic	Prostate	522	-	528	420	-	434	x	x
103R	P08R	1x8 Gy	Mixed	Prostate	459	-	467	381	-	394	x	x
119R	P09R	1x8 Gy	Mixed	Prostate	479	-	482	429	-	431	437	437
203R	P10R	1x8 Gy	Mixed	Prostate	442	-	x	261	-	x	304	304
213R	P11R	1x8 Gy	Mixed	Prostate	277	-	288	520	-	528	601	601
235L	P12L	1x8 Gy	Mixed	Prostate	519	-	494	439	-	407	386	386
305L	P13L	1x8 Gy	Mixed	Prostate	610	-	604	554	-	559	575	575
306R	P14R	1x8 Gy	Mixed	Prostate	551	-	563	699	-	749	776	776
107L	P15L	1x8 Gy	Mixed	Rectum	458	-	454	247	-	244	249	249
107R	P15R	1x8 Gy	Mixed	Rectum	515	-	509	381	-	382	394	394
104L	P16L	1x8 Gy	Mixed	aCUP	452	-	444	328	-	320	311	311
208R	P17R	6x4 Gy	Lytic	Breast	428	414	422	315	312	311	317	317
208R*	P17R*	6x4 Gy	Lytic	Breast	*	*	*	546	536	535	526	526
208R*	P17R*	6x4 Gy	Lytic	Breast	*	*	*	611	601	609	599	599
221L	P18L	6x4 Gy	Lytic	Breast	395	383	383	242	235	229	208	208
302R	P19R	6x4 Gy	Lytic	Breast	478	467	461	568	542	551	625	625
202L	P20L	6x4 Gy	Lytic	Lung	511	488	485	329	318	320	x	x

207R	P21R	6x4 Gy	Lytic	Lung	400	404	x	x	298	303	x	x
211L	P22L	6x4 Gy	Lytic	Lung	396	393	x	x	233	229	x	x
218R	P23R	6x4 Gy	Lytic	Prostate	400	403	x	x	246	244	x	x
117L	P24L	6x4 Gy	Lytic	Kidney	509	501	506	506	818	804	804	821
123L	P25L	6x4 Gy	Lytic	Kidney	466	477	485	483	233	242	249	244
123R	P25R	6x4 Gy	Lytic	Kidney	452	472	473	471	227	242	245	237
304R	P26R	6x4 Gy	Lytic	Rectum	418	407	415	398	241	237	246	243
105L	P27L	6x4 Gy	Lytic	Kahler's disease	424	422	419	x	156	151	149	x
105R	P27R	6x4 Gy	Lytic	Kahler's disease	419	424	415	x	160	160	156	x
121R	P28R	6x4 Gy	Lytic	Urethra	428	423	x	412	531	521	x	507
121R*	P28R*	6x4 Gy	Lytic	Urethra	*	*	*	*	709	702	x	696
101L	P29L	6x4 Gy	Lytic	Cervix	472	434	x	x	567	485	x	x
233R	P30R	6x4 Gy	Blastic	Lung	550	556	577	622	448	458	493	572
108R	P31R	6x4 Gy	Blastic	Prostate	555	552	558	x	530	533	539	x
214L	P32L	6x4 Gy	Blastic	Prostate	389	390	430	482	335	346	403	474
216R	P33R	6x4 Gy	Blastic	Prostate	552	547	572	581	579	575	605	617
225L	P34L	6x4 Gy	Blastic	Prostate	432	422	435	445	297	298	323	383
219L	P35L	5x4 Gy	Mixed	Breast	513	508	x	x	395	389	x	x
212R	P36R	6x4 Gy	Mixed	Breast	412	404	414	427	270	275	289	329
234R	P37R	6x4 Gy	Mixed	Lung	459	451	x	468	452	438	x	495
222R	P38R	5x4 Gy	Mixed	Prostate	410	401	400	405	358	367	380	424
222R*	P38R*	5x4 Gy	Mixed	Prostate	*	*	*	*	193	188	176	179
232L	P39L	5x4 Gy	Mixed	Prostate	489	471	x	x	747	715	x	x
122L	P40L	6x4 Gy	Mixed	Prostate	302	306	x	x	279	292	x	x
204L	P41L	6x4 Gy	Mixed	Prostate	327	326	329	318	182	180	184	184
204R	P41R	6x4 Gy	Mixed	Prostate	339	336	343	338	206	201	205	205
227L	P42L	6x4 Gy	Mixed	Prostate	603	x	x	615	717	x	x	742

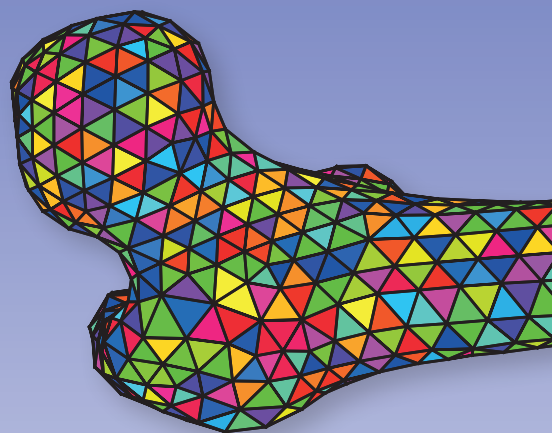




# CHAPTER 4

**CAN PATIENT-SPECIFIC FINITE ELEMENT MODELS BETTER  
PREDICT FRACTURES IN METASTATIC BONE DISEASE THAN  
EXPERIENCED CLINICIANS?**

***TOWARDS INTRODUCING COMPUTATIONAL MODELLING  
INTO DAILY CLINICAL PRACTICE***



---

Florieke Eggermont\*, Loes Derikx\*, Nico Verdonshot, Ingrid van der Geest, Marianne de Jong,  
An Snyers, Yvette van der Linden, Esther Tanck

\*Joint first authorship

Bone and Joint Research 2018; 7(6):430–439.



## Introduction

Cancers of the breast, prostate, lung, kidney, and thyroid can metastasize to bone.<sup>1-3</sup> These metastases can cause pain and, when left untreated, carry a risk of developing complications such as pathological fractures or, in case of vertebral metastases, spinal cord compression.<sup>1,2,4</sup> Pathological fractures in extremities affect quality of life as they hamper the patient's mobility and self-care.

Femoral metastases with a low risk of fracture can be treated conservatively with local radiotherapy. Metastases with a high risk of fracture require prophylactic surgery to retain stability of the bone.<sup>5</sup> This is an invasive procedure requiring anaesthesia, which is generally complex in cancer patients with limited life expectancy and deteriorating condition. Thus, the decision to proceed with either a non-invasive treatment or a prophylactic surgical treatment should be carefully made.

However, current clinical practice lacks an accurate tool to guide clinicians to the correct treatment decision. Numerous studies have evaluated lesion or patient factors on the probability of impending fractures; however, none has shown a sufficient predictive power.<sup>5</sup> A potential tool to improve clinical fracture risk assessments is finite element (FE) modelling, which has been shown to predict human femoral bone strength fairly accurately.<sup>6-10</sup> Our group has shown that the FE model accurately calculated failure load and fairly predicted fracture locations in cadaver femurs with and without artificial lesions compared with mechanical experiments.<sup>11-13</sup> Moreover, we demonstrated that ranking on FE failure load better resembled the experimentally measured failure loads than rankings by experienced clinicians.<sup>11</sup>

In this prospective cohort study, we investigated whether our subject-specific FE models are able to identify patients at risk of pathological femoral fractures resulting from metastatic bone disease. For this purpose, we included patients referred for radiotherapy to treat painful femoral metastases. Against expectations, some of these patients sustained pathological fractures in the femur during follow-up. We calculated the femoral failure loads and compared those between patients who did or did not sustain a fracture. In addition, we compared the FE predictions with assessments by experienced clinicians. We hypothesized that the FE models more accurately identify patients with a high fracture risk than experienced clinicians.

## Patients and Methods

### Study design

Between August 2006 and September 2009, all patients referred for palliative radiotherapy of the femur to three participating radiotherapy institutes in the Netherlands (Radboud university medical center (institute 1), Radiotherapeutic Institute Friesland (institute 2), and Leiden University Medical Center (institute 3) were asked to participate in this prospective cohort study. Globally, 20% to 25% of the eligible patients participated. Ethical approval was obtained from all participating centres. These patients received palliative radiotherapy following Dutch clinical guidelines. Lesions with an

axial cortical involvement < 30 mm have an expected low risk of fracture (< 5%) and were treated with a single dose of 8 Gy.<sup>5</sup> If the axial cortical involvement was > 30 mm, the risk of fracture is estimated at 23%.<sup>14</sup> These patients were referred for prophylactic surgery and therefore excluded from this study.<sup>5</sup> If the patient's condition was such that surgery was undesirable or impossible, the patient was referred for multiple fraction radiotherapy (e.g. 5 or 6 × 4 Gy) to induce remineralization of the bone.<sup>15</sup> These patients were included in this study. Further inclusion criteria are depicted in Table 1.<sup>16</sup> During the study period, 62 patients gave their consent. The patients were grouped according to the predominant appearance of their bone metastases (i.e. lytic or blastic). We excluded patients who had predominant blastic lesions (n = 16). Although blastic lesions generally lead to a decreased structural bone strength,<sup>17</sup> in this study the femoral bone strength was overestimated, probably due to the high degree of mineralization which resulted in unrealistically strong material properties in the FE model (see supplementary information). Additionally, patients who had no body weight (BW) recorded (n = 5) or sustained a femoral fracture more than a year after inclusion (n = 1) were excluded. One patient sustained a femoral fracture following a fall and was therefore excluded. This led to inclusion of 39 patients with predominant lytic bone lesions in this study.

Baseline characteristics of all patients were recorded before radiotherapy. Furthermore, quantitative computed tomography (QCT) scans of the femoral region were retrieved prior to, at 28, and at 70 days after radiotherapy. Patients referred for multiple fraction radiotherapy underwent an additional QCT scan on the final day of their radiation schedule to capture the potential short-term effect of multiple fraction radiotherapy.<sup>18</sup> Through follow-up questionnaires and hospital records, patients were actively followed for six months or until a fracture occurred or until death, as competing risk, whichever occurred first. Based on having sustained a fracture, the patients were divided into either the fracture (F) group or the non-fracture (NF) group. Additionally, after two years, data on fractures and death were updated with the use of hospital records.

**Table 1:** Inclusion criteria

**Inclusion criteria**

- Proven malignancy
- Karnofsky score<sup>16</sup> ≥ 60
- No clinical or radiological evidence of pathological fracturing of the femur
- No prior palliative surgery for the current treatment site of the femur
- No planned surgical intervention of the femoral bone
- No systemic radiotherapy 30 days prior to entry into the study
- No previous radiotherapy to the current treatment site of the femur
- Patient is able and willing to fill out baseline and follow-up forms on pain and quality of life
- Patient is willing to undergo additional CT scans for the femoral region

### Different CT scanners

Recent work by Carpenter *et al.*<sup>19</sup> has shown that the use of different CT scanners can have a significant effect on bone mineral density measurements and subsequent calculated failure loads, which is

difficult to correct for. In the current study, the three institutes used two different types of CT scanner, Philips Big Bore Brilliance (institute 1) and Philips AcQSim CT (institute 2 and 3), both manufactured by Philips Medical Systems, Eindhoven, The Netherlands. Although QCT scan settings were protocolized as far as possible, inter-scanner effects may have been present in the input to our FE models, which could potentially lead to incorrect or at least incomparable FE failure loads. Therefore, apart from a group analysis, we also analyzed the data individually for the three institutes to circumvent such inter-scanner differences. It should be noted that our previous *ex vivo* validation study was conducted using the scanning equipment of institute 1.<sup>11-13</sup>

## FE modelling

Patient-specific femoral FE models were generated, for the greater part, using the workflow reported previously.<sup>11</sup> Summarizing, QCT images were generated using a standard protocol (as far as allowed by clinical practice), with the following settings: 120 kVp, 220 mA, slice thickness 3 mm, pitch 1.5, spiral and standard reconstruction, in-plane resolution 0.9375 mm. The patient-specific femoral geometry was segmented from the most recent CT images available and converted to a 3D surface mesh (Mimics 11.0 and 14.0, Materialise, Leuven, Belgium) and a solid mesh consisting of tetrahedral elements (average element volume 1.0 mm<sup>3</sup>; Patran 2005r2, MSC Software Corporation, Santa Ana, California), subsequently. A solid calibration phantom containing known calcium equivalent densities (Image Analysis, Columbia, Kentucky) was scanned along with the patient at the level of the proximal femur. Since pilot tests showed that calibration in diaphyseal slices was most accurate due to beam hardening in more proximal slices, we performed a mean diaphyseal slice calibration to convert the grey values to calcium equivalent densities, ash densities, and non-linear isotropic material behaviour, respectively, based on the material model of Keyak *et al.*<sup>7</sup> In this material model, the post-failure material behaviour for each element is represented by an initial perfectly plasticity phase, followed by a strain softening phase and finally an indefinite perfectly plastic phase.<sup>7</sup> However, in case patients' limbs were supported by a cushion to diminish pain during CT scanning, an air gap between calibration phantom and patient was present, leading to an artefact in the calibration phantom at the diaphyseal level. In such cases, we used more proximal slices for calibration.

The FE simulations of the proximal femur were performed using MSC.MARC (2007r1; MSC Software Corporation). The FE models were loaded by displacing a cup on the head of the femur in the axial direction, while distally fixed at the knee joint centre by two bundles of high-stiffness (200 000 000 N/m) springs (Figure 1), which roughly resembles single-legged stance. Force-displacement curves were made based on displacement and contact normal forces that were registered for each increment. The maximum total reaction force determined the failure load of the femur, which was normalized for BW. The failure location was defined by elements that had plastically deformed at the moment of structural failure, and was compared with the post-fracture radiograph.



---

**Figure 1:** Boundary conditions for the finite element model. The model was distally fixed by springs with a very high stiffness and the load was applied by means of a cup on the head of the femur, which incrementally displaced in a distal direction.

### Clinical assessment

To compare the FE predictions with clinical fracture risk assessments, we generated digitally reconstructed radiographs (DRRs) from the CT scans in this study.<sup>20</sup> We asked two radiation oncologists with broad expertise in palliative radiotherapy, who regularly discuss and refer patients to the orthopaedics department (C1 and C2), and one experienced orthopaedic oncology surgeon (C3) to individually assess the DRRs just as in daily practice, without providing any further information.

First, they indicated whether the patient carried a high risk of fracture requiring elective surgery. Subsequently, we asked them to judge whether the cortical disruption caused by the metastasis was > 30 mm.<sup>14</sup>

### Statistical analysis

Baseline data were compared between fracture and non-fracture group on the femur level using chi-squared test, Fisher's exact test, or Mann–Whitney U test, where applicable. We compared the failure load corrected for BW between the fractured and the non-fractured femurs using Mann–Whitney U tests. For all tests, the level of significance was defined as  $p < 0.05$ .

To compare the clinical assessments with the predictions by the FE model, a critical FE failure load was defined for the whole group, as well as for each institute separately, classifying a patient to a high or a low fracture risk. More specifically, diagnostic accuracy values (sensitivity and specificity, and positive and negative predictive values (PPV and NPV)) were calculated for different thresholds of a critical failure load using increments of  $0.5 \times \text{BW}$ . The threshold with the highest sum of specificity and sensitivity was chosen. For comparison with clinical assessments, we used the critical failure loads of the separate institutes.

## Results

### Patients

In all, 39 patients with predominant lytic painful bone metastases were included in this study. One of the patients sustained a fracture of the femur one month after follow-up, and was included in the F group. This F group consisted of seven patients sustaining nine fractures (Tables 2 and 3). Two of these fractures occurred in the contralateral femur that was not irradiated (the irradiated femurs of these patients were included in the NF group). Additionally, two cases with an unknown cause of fracture were included. The NF group comprised a total of 34 patients with 38 treated non-fractured femurs (Table 3). There were no significant differences in baseline characteristics between the F and NF group (Table 4).



**Table 2.** Characteristics of the patients who sustained a fracture during follow-up.

Patient	Sex	Age at inclusion (yrs)	Femur	Treatment dose (fractions, n)	Time to fracture (days)	Type of fracture	Activity while fracture occurred
Patient 1	M	70	Right femur (F1)	24 Gy/6	123	Neck of femur fracture	Walking
			Left femur (F2)	24 Gy/6	123	Neck of femur fracture	Walking
Patient 2	F	53	Right femur (F3)	8 Gy/1	92	Pertrochanteric fracture	Spontaneously
			Left femur (F4)	-	92	Neck of femur fracture	Spontaneously
Patient 3	M	64	Left femur (F5)	24 Gy/6	7	Subtrochanteric fracture	Spontaneously
Patient 4	F	66	Left femur (F6)	24 Gy/6	13	Neck of femur fracture	Spontaneously
Patient 5	F	62	Right femur (F7)	24 Gy/6	3	Unknown	Unknown
Patient 6	M	89	Right femur (F8)	8 Gy/1	237	Diaphyseal fracture	Spontaneously
Patient 7	F	80	Left femur (F9)	-	133	Neck of femur fracture	Unknown

**Table 3:** Femurs and patients included in each of the institutes.

	Institute 1 <sup>a</sup>		Institute 2 <sup>b</sup>		Institute 3 <sup>c</sup>	
	femurs	patients	femurs	patients	femurs	patients
<b>F</b>	5	3	3	3	1	1
<b>NF</b>	11	8	23	22	4	4

<sup>a</sup> One patient in institute 1 fractured both femurs, but was only radiated on the right side; <sup>b</sup> One patient in institute 2 was radiated on both sides, but only fractured her right femur, leaving her left femur in the NF group. <sup>c</sup> One patient in institute 3 fractured her non-treated femur, leaving her treated femur in the NF group.

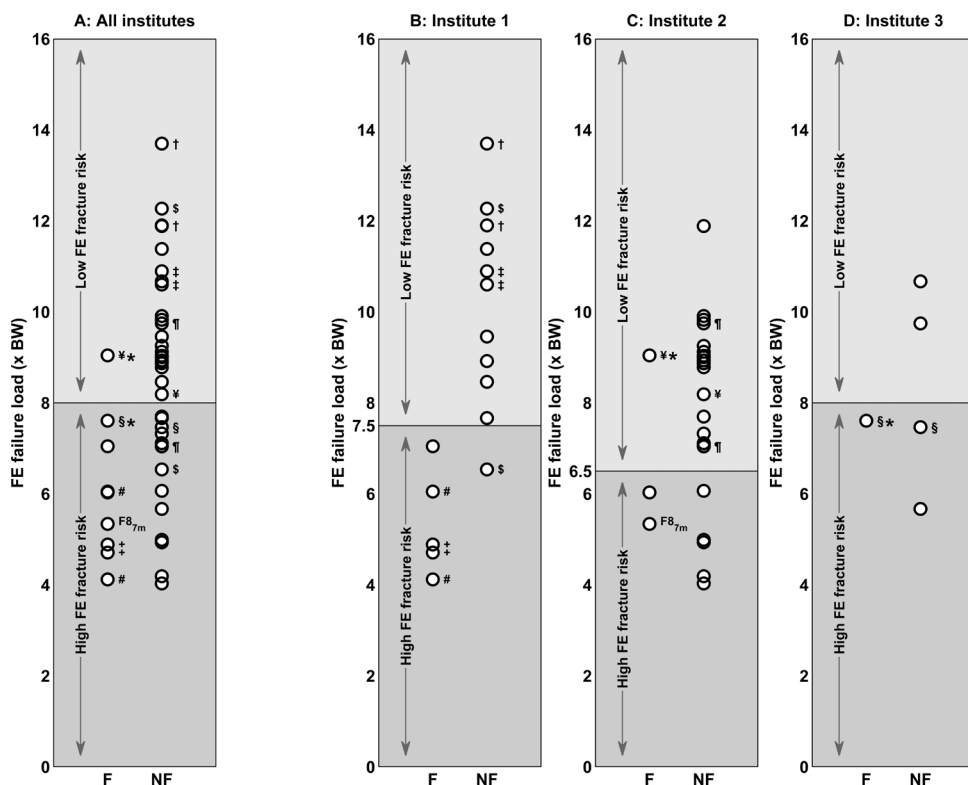
**Table 4:** Baseline characteristics.

	Fracture group (F) (n = 9) <sup>a</sup>	Non-fracture group (NF) (n = 38) <sup>a</sup>	p-value
<b>Gender, n (%)</b>			
Male	4 (44%)	20 (53%)	0.7 <sup>e</sup>
Female	5 (56%)	18 (47%)	
<b>Age in years</b>			
Median (IQR)	66.0 (57.5 to 75.0)	62.5 (52.8 to 76.5)	0.5 <sup>f</sup>
<b>Body weight in kg</b>			
Median (IQR)	73.0 (63.0 to 76.5)	76.0 (57.8 to 87.3)	0.6 <sup>f</sup>
<b>Radiation schedule, n (%) <sup>b</sup></b>			
SF	4 (44%)	15 (39%)	1 <sup>e</sup>
MF	5 (56%)	23 (61%)	
<b>KPS</b>			
Median (IQR)	80.0 (70.0 to 80.0)	80.0 (70.0 to 90.0)	0.6 <sup>f</sup>
<b>Time since primary tumour in years</b>			
Median (IQR)	3.6 (1.7 to 6.5)	3.3 (0.8 to 5.5)	0.8 <sup>f</sup>
<b>Time since first metastasis in years</b>			
Median (IQR)	3.2 (0.1 to 3.6)	1.1 (0.1 to 2.8)	0.5 <sup>f</sup>
<b>Primary cancer site, n (%)</b>			
Breast	2 (22%)	11 (29%)	0.3 <sup>g</sup>
Lung	2 (22%)	5 (13%)	
Prostate	2 (22%)	12 (32%)	
Kidney	0 (0%)	4 (11%)	
Rectum	0 (0%)	2 (5%)	
Multiple myeloma	3 (33%)	2 (5%)	
Urethra	0 (0%)	1 (3%)	
aCUP <sup>c</sup>	0 (0%)	1 (3%)	
<b>Time to death since inclusion in months <sup>d</sup></b>			
Median (IQR)	11.0 (4.0 to 13.0)	8.0 (3.0 to 17.0)	0.9 <sup>f</sup>
<b>Institute, n (%)</b>			
1	5 (56%)	11 (29%)	0.3 <sup>g</sup>
2	3 (33%)	23 (61%)	
3	1 (11%)	4 (11%)	
<b>Lesion type, n (%)</b>			
Not visible	1 (11%)	3 (8%)	0.4 <sup>g</sup>
Lytic	1 (11%)	13 (34%)	
Mixed	7 (78%)	22 (58%)	

IQR: interquartile range. <sup>a</sup> Fracture group: 9 femurs in 7 patients. Non-fracture group: 38 femurs in 34 patients. 2 patients had one fractured and one non-fractured femur. <sup>b</sup> Two femurs in the fracture group were not treated with radiotherapy. <sup>c</sup> Cancer of Unknown Primary origin. <sup>d</sup> Date of death missing for three non-fracture patients. <sup>e</sup> p-value from Fisher's Exact test. <sup>f</sup> p-value from Mann-Whitney U test. <sup>g</sup> p-value from Pearson Chi-Square.

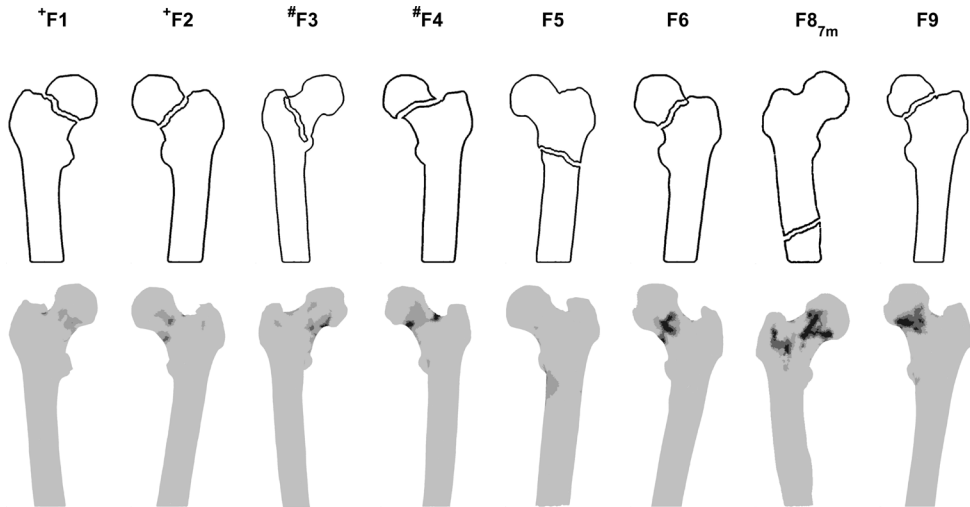
## FE models

Figure 2 shows the BW-corrected failure loads for all femurs in this study together (Figure 2A), as well as split to the three institutes (Figure 2B-D). The median failure load of all fractured femurs together was significantly lower compared with failure load of all non-fractured femurs (6.03 (interquartile range (IQR) 4.80 to 7.33) vs 8.93 (IQR 7.10 to 9.85),  $p = 0.002$ ). After splitting to individual institute to avoid inter-scanner differences, the median failure load of fractured femurs was significantly lower in institute 1 compared with failure load of the non-fractured femurs (4.89 (IQR 4.41 to 6.55) vs 10.60 (IQR 8.46 to 11.90),  $p = 0.001$ ). This was not the case for the femurs from institutes 2 and 3 (6.03 vs 8.78 (IQR 7.04 to 9.12),  $p = 0.5$  and 7.61 vs 8.61 (IQR 6.12 to 10.44),  $p = 1$ ).



**Figure 2:** Femoral failure load for patients who did (F) or did not (NF) sustain a femoral fracture during follow-up, corrected for body weight (BW), A) in all institutes without considering inter-scanner differences, and in B) institute 1, C) institute 2, and D) institute 3 separately. It should be noted that one femur (F87m) fractured one month after follow-up. The institutional thresholds were used to compare the predictive power of the finite element (FE) model versus experienced clinicians. \*Femur fractured during unknown activity; all other symbols (#, +, †, \$, ‡, ¶, ¥, §) indicate paired femurs.

We compared the actual fracture location with those predicted by the FE models for eight out of nine femurs. In one case, no radiological information on clinical fracture location was available. Six out of eight FE fracture locations resembled the actual fractures on post-fracture radiographs (Figure 3). In the two other cases, the FE models predicted femoral neck fractures, whereas these patients clinically presented with a pertrochanteric and diaphyseal fracture, respectively.



**Figure 3:** Schematic overview of clinical fracture locations (upper panel), indicated by an experienced clinician who was blind to the predicted fracture locations, and the fracture locations at failure (mid-coronal plane) predicted by the finite element models (lower panel). Femurs indicated with + and # are paired femurs. F8 fractured one month after follow-up (7m). There was no clinical information about fracture location available for F7.

### Clinical assessment

A critical failure load was defined for the whole group, as well as for each institute. The critical failure loads were  $8.0 \times BW$  for the whole group and  $7.5 \times BW$ ,  $6.5 \times BW$ , and  $8.0 \times BW$  for institutes 1, 2, and 3, respectively (Figure 2). When each institute was analyzed individually, the sensitivity remained the same, while the specificity increased from 0.63 to 0.79.

For comparison with clinical assessments, we used the critical failure loads of the separate institutes. More patients were correctly identified with a high fracture risk by the FE model than by clinicians who relied on their clinical experience (Figure 4), resulting in higher sensitivity of the FE model (0.89) compared with the clinicians (ranging from 0.00 to 0.33; Table 5). The FE model identified 16 femurs with a high fracture risk, eight of which actually fractured during follow-up (PPV = 0.50). The PPV for clinicians ranged between 0.00 and 0.50. Of the 38 non-fractured femurs, the FE model correctly identified 30 femurs as having a low fracture risk (specificity = 0.79). The specificity values for clinicians

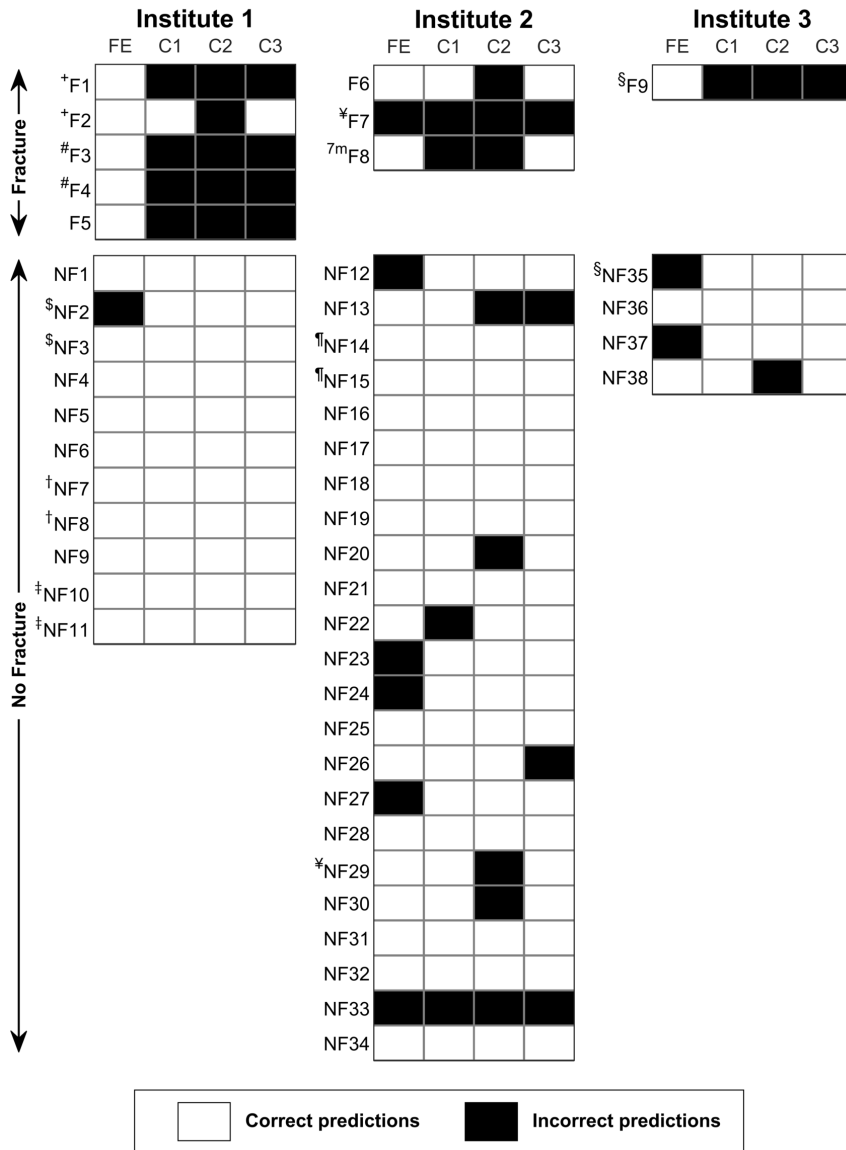
were slightly higher and ranged between 0.84 and 0.95, although the 95% confidence intervals overlapped. The FE model identified 31 femurs with a low fracture risk, of which 30 indeed did not fracture (NPV = 0.97). NPV for clinicians were lower and ranged between 0.78 and 0.85.

When the experienced clinicians were asked to base their decision on 30 mm axial cortical disruption (Table 5), their diagnostic accuracy values were comparable with the predictions based on clinical experience.

**Table 5:** Summary statistics for the prediction accuracy of the FE model and the experienced clinicians when relying on their experience and when judging whether cortical involvement was larger than 30 mm.<sup>14</sup> 95% confidence intervals are given between brackets.

		F	NF	SE	SP	PPV	NPV	
FE model whole group <sup>a</sup>	F predicted	8	14	<b>0.89</b>	<b>0.63</b>	<b>0.36</b>	<b>0.96</b>	
	NF predicted	1	24	(0.52-1.00)	(0.46-0.78)	(0.26-0.48)	(0.79-0.99)	
FE model split to institute <sup>b</sup>	F predicted	8	8	<b>0.89</b>	<b>0.79</b>	<b>0.50</b>	<b>0.97</b>	
	NF predicted	1	30	(0.52-1.00)	(0.63-0.90)	(0.34-0.66)	(0.82-0.99)	
Experience	C1	F predicted	2	2	<b>0.22</b>	<b>0.95</b>	<b>0.50</b>	<b>0.84</b>
		NF predicted	7	36	(0.03-0.60)	(0.82-0.99)	(0.14-0.86)	(0.78-0.88)
	C2	F predicted	0	6	<b>0.00</b>	<b>0.84</b>	<b>0.00</b>	<b>0.78</b>
		NF predicted	9	32	(0.00-0.34)	(0.69-0.94)		(0.76-0.80)
	C3	F predicted	3	3	<b>0.33</b>	<b>0.92</b>	<b>0.50</b>	<b>0.85</b>
		NF predicted	6	35	(0.07-0.70)	(0.79-0.98)	(0.19-0.81)	(0.78-0.90)
30 mm cortical involvement	C1	F predicted	2	1	<b>0.22</b>	<b>0.97</b>	<b>0.67</b>	<b>0.84</b>
		NF predicted	7	37	(0.03-0.60)	(0.86-1.00)	(0.17-0.95)	(0.79-0.88)
	C2	F predicted	1	6	<b>0.11</b>	<b>0.84</b>	<b>0.14</b>	<b>0.80</b>
		NF predicted	8	32	(0.00-0.48)	(0.69-0.94)	(0.02-0.55)	(0.75-0.84)
	C3	F predicted	0	1	<b>0.00</b>	<b>0.97</b>	<b>0.00</b>	<b>0.80</b>
		NF predicted	9	37	(0.00-0.34)	(0.86-1.00)		(0.80-0.81)

<sup>a</sup> Without considering inter-scanner differences. <sup>b</sup> Prediction of the FE models based on different thresholds for each institute (inst 1: 7.5 x BW, inst 2: 6.5 x body weight (BW), inst 3: 8.0 x BW). F = fracture, NF = non-fracture, SE = sensitivity, SP = specificity, PPV = positive predictive value, NPV = negative predictive value



**Figure 4:** Correct and incorrect fracture predictions by the finite element (FE) model and the experienced clinicians (C1, C2, C3) for institute 1, institute 2, and institute 3. Clinicians judged the reconstructed radiographs of the patients based on their experience, without any further guidelines prescribed. For the FE predictions thresholds of  $7.5 \times BW$ ,  $6.5 \times BW$ , and  $8.0 \times BW$  for institute 1, 2, and 3, respectively, were used to indicate fracture (F) or non-fracture (NF). Results are shown per group (F and NF). Symbols (+, #, \$, †, ‡, ¥, ¶, §) indicate paired femurs. F8 fractured one month after follow-up (7m).

## Discussion

Previously, we have shown that FE models calculated the femoral load to failure comparably with those measured in mechanical experiments.<sup>11</sup> In the current study, we applied these FE models *in vivo* by comparing the model predictions with clinical follow-up data in a prospective cohort of patients with cancer and painful femoral metastases who were referred for palliative radiotherapy. We verified whether the model could have predicted the pathological fractures that some of the patients with painful bone metastases unexpectedly sustained during follow-up.

We showed a difference in median failure load between patients who sustained a pathological fracture and those who did not when we analyzed the whole group together as well as in institute 1. However, this difference was not present in the two other institutes, probably because of the low number of fractures. Additionally, for two femurs in the latter institutes, the activity during which the fracture occurred was not recorded. We could therefore not confirm whether these fractures were pathological. If these fractures were traumatic, the high predicted failure loads as calculated by the FE model would have been expected, with improved predictions as a result. Nevertheless, the results from institute 1 show that FE models were able to comprehend many factors that contribute to the *in vivo* load capacity of metastatic femurs, such as the bone quality and the bone geometry, or compromise it, such as the location and the size of the lesion. Goodheart *et al.*<sup>10</sup> recently found that FE models can be used to distinguish between metastatic femurs that would and would not fracture. Positive findings were also shown in the field of osteoporosis (e.g. Keyak *et al.*,<sup>21</sup> Kopperdahl *et al.*<sup>22</sup>), where FE strength was found to highly correlate with fracture<sup>21</sup> and FE bone strength remained predictive for fracture after correction for total hip areal bone mineral density (aBMD) in men and women.<sup>22</sup>

In the present study, the FE predictions demonstrated higher sensitivity compared with clinical assessments. This suggests a better identification of patients who will sustain a fracture by the FE model, resulting in prevention of more pathological fractures. Specificity of the FE model was relatively high but slightly lower compared with the clinicians. However, NPV were very high (97%), indicating that if the FE model predicts a low fracture risk, a fracture almost never occurs. As a result, the FE model could be clinically used to prevent unnecessary surgery. Diagnostic values of the 30 mm cortical involvement from a previous study (sensitivity 86%, specificity 58%, PPV 23%, NPV 97%)<sup>5</sup> were quite different from the clinical assessment of the current study (Table 5), showing that these values may be dependent on the clinicians and/or the studied patient group.

In six out of eight cases, the predicted fracture location resembled the actual clinical fracture location. In cases F3 and F8, the FE fracture locations did not resemble the clinical fracture lines. F8 suffered from mixed metastases and clinically fractured through a lesion with higher CT density, which could explain why the FE model did not predict the correct fracture location. Since F3 fractured spontaneously, the axial load applied in this study might be inappropriate to simulate the correct fracture line in this femur. Modelling more and realistic loading conditions may further improve the predicted fracture location. Although the results in this study are promising, some limitations should be mentioned here. First of all, we realize that the sample size in this study is limited, especially after splitting to institute.

As a result, in institutes 2 and 3, the critical failure loads were based on only a few fractures ( $n = 3$  and  $n = 1$ ), indicating the need for larger data sets in the near future. For that purpose, a solution to overcome inter-scanner differences should be developed, something we are currently working on.<sup>23</sup>

A second limitation in this study relates to the modelling of metastatic tissue. We excluded patients with predominant blastic femoral lesions from our analyses, as blastic lesions generally show very high CT intensities. In the current FE model, these CT intensities would have been converted to material behaviour using relationships that are defined based on experiments with human tissue affected by metastases as well as healthy bone.<sup>7</sup> Therefore, the empirical relationships have to be adapted for blastic metastatic tissue. So far, differences in microarchitecture have been described for metastases (e.g. Sone *et al.*<sup>24</sup>), but the mechanical behaviour has not yet been established unequivocally.<sup>25,26</sup> Moreover, adapted material models did not yet improve the predictive power of FE models with metastatic lesions.<sup>27</sup> Hence, further research is required to determine the mechanical behaviour of different types of metastatic tissue.

Third, the use of strain softening as a material property can cause mesh sensitivity and its use to capture localization can be questioned. In the past, our group performed a sensitivity analysis by varying mesh density with or without applying a correction for element size. Based on a fit with experimental results, we decided to use the current mesh density (average element volume of  $1 \text{ mm}^3$ ) without applying the correction. Subsequently, we have been using the same protocol<sup>11</sup> to minimize the differences between bones.

As a fourth limitation, it should be mentioned that the clinicians pointed out that the quality of the DRRs was suboptimal compared with the conventional radiographs they normally use, which may have affected their assessments.

In conclusion, we showed that patient-specific FE models are a potential tool to improve clinical fracture risk predictions in patients with metastatic bone disease. The FE models provided an accurate identification of patients with high fracture risk in one of the three institutes. Future work in a larger patient population should confirm the higher predictive power of the FE models compared with current clinical guidelines. However, a robust solution to overcome inter-scanner differences should be developed before the FE models can be extensively used for clinical fracture risk assessments in a multicentre setting. In the future, the individual FE outcome may help patients and their clinicians to weigh the chance of fracturing against choosing the most appropriate treatment, which is either non-invasive radiotherapy to treat pain, or surgery to restore stability.

## Acknowledgements

This work was supported by the Dutch Cancer Society (KUN 2012-5591), the Dutch Science Foundation NWO-STW (NPG.06778), and Fonds NutsOhra (1102-071). The authors would like to thank Wouter Gevers for his help in generating the digitally reconstructed radiographs, and Femke Peters, MD, PhD, and Tom Rozema, MD for their input to this study.



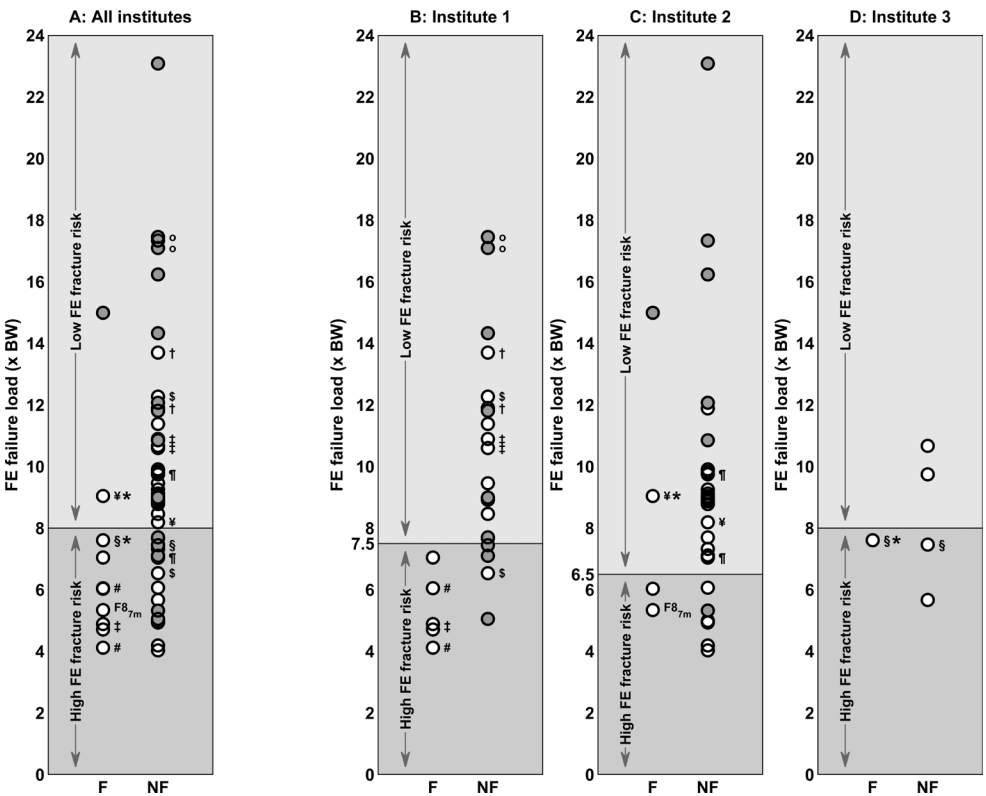
## References

1. Coleman RE. 1997. Skeletal complications of malignancy. *Cancer* 80:1588-1594.
2. Gralow JR, Biermann JS, Farooki A, Fornier MN, Gagel RF, Kumar RN, Shapiro CL, Shields A, Smith MR, Srinivas S, Van Poznak CH. 2009. NCCN Task Force Report: Bone Health in Cancer Care. *J Natl Compr Canc Netw* 7 Suppl 3:S1-32.
3. Johnson SK, Knobf MT. 2008. Surgical interventions for cancer patients with impending or actual pathologic fractures. *Orthop Nurs* 27:160-171.
4. Coleman RE. 2006. Clinical features of metastatic bone disease and risk of skeletal morbidity. *Clin Cancer Res* 12:6243s-6249s.
5. Van der Linden YM, Dijkstra PD, Kroon HM, Lok JJ, Noordijk EM, Leer JW, Marijnen CA. 2004. Comparative analysis of risk factors for pathological fracture with femoral metastases. *J Bone Joint Surg Br* 86:566-573.
6. Bessho M, Ohnishi I, Matsuyama J, Matsumoto T, Imai K, Nakamura K. 2007. Prediction of strength and strain of the proximal femur by a CT-based finite element method. *J Biomech* 40:1745-1753.
7. Keyak JH, Kaneko TS, Tehranzadeh J, Skinner HB. 2005. Predicting proximal femoral strength using structural engineering models. *Clin Orthop*:219-228.
8. Lenaerts L, van Lenthe GH. 2009. Multi-level patient-specific modelling of the proximal femur. A promising tool to quantify the effect of osteoporosis treatment. *Philos Transact Ser A Math Phys Eng Sci* 367:2079-2093.
9. Schileo E, Taddei F, Cristofolini L, Viceconti M. 2008. Subject-specific finite element models implementing a maximum principal strain criterion are able to estimate failure risk and fracture location on human femurs tested in vitro. *J Biomech* 41:356-367.
10. Goodheart JR, Cleary RJ, Damron TA, Mann KA. 2015. Simulating activities of daily living with finite element analysis improves fracture prediction for patients with metastatic femoral lesions. *J Orthop Res* 33:1226-1234.
11. Derikx LC, van Aken JB, Janssen D, Snyers A, van der Linden YM, Verdonchot N, Tanck E. 2012. The assessment of the risk of fracture in femora with metastatic lesions: comparing case-specific finite element analyses with predictions by clinical experts. *J Bone Joint Surg Br* 94:1135-1142.
12. Derikx LC, Vis R, Meinders T, Verdonchot N, Tanck E. 2011. Implementation of asymmetric yielding in case-specific finite element models improves the prediction of femoral fractures. *Comput Methods Biomech Biomed Engin* 14:183-193.
13. Tanck E, van Aken JB, van der Linden YM, Schreuder HW, Binkowski M, Huizenga H, Verdonchot N. 2009. Pathological fracture prediction in patients with metastatic lesions can be improved with quantitative computed tomography based computer models. *Bone* 45:777-783.
14. Van der Linden YM, Kroon HM, Dijkstra SP, Lok JJ, Noordijk EM, Leer JW, Marijnen CA. 2003. Simple radiographic parameter predicts fracturing in metastatic femoral bone lesions: results from a randomised trial. *Radiother Oncol* 69:21-31.
15. Koswig S, Budach V. 1999. Remineralisation und Schmerzlinderung von Knochenmetastasen nach unterschiedlich fraktionierter Strahlentherapie (10 mal 3 Gy vs. 1 mal 8 Gy). Eine prospektive Studie. *Strahlenther Onkol* 175:500-508.
16. Karnofsky D, Burchenal J. 1949. The clinical evaluation of chemotherapeutic agents in cancer. In: MacLeod C editor. *Evaluation of Chemotherapeutic Agents*. New York: Columbia University Press; pp. 191-205.

17. Healey JH, Brown HK. 2000. Complications of bone metastases: surgical management. *Cancer* 88:2940-2951.
18. Eggermont F, Derikx LC, Verdonschot N, Hannink G, Kaatee R, Tanck E, van der Linden YM. 2017. Limited short-term effect of palliative radiation therapy on quantitative computed tomography-derived bone mineral density in femora with metastases. *Adv Radiat Oncol* 2:53-61.
19. Carpenter RD, Saeed I, Bonaretti S, Schreck C, Keyak JH, Streeper T, Harris TB, Lang TF. 2014. 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* 36:1225-1232.
20. Jacobs F, Sundermann E, De Sutter B, Christiaens M, Lemahieu I. 1998. A fast algorithm to calculate the exact radiological path through a pixel or voxel space. *J Comput Inf Technol* 6:89-94.
21. Keyak JH, Sigurdsson S, Karlsdottir GS, Oskarsdottir D, Sigmarsdottir A, Kornak J, Harris TB, Sigurdsson G, Jonsson BY, Siggeirsdottir K, Eiriksdottir G, Gudnason V, Lang TF. 2013. Effect of finite element model loading condition on fracture risk assessment in men and women: the AGES-Reykjavik study. *Bone* 57:18-29.
22. Kopperdahl DL, Aspelund T, Hoffmann PF, Sigurdsson S, Siggeirsdottir K, Harris TB, Gudnason V, Keaveny TM. 2014. Assessment of incident spine and hip fractures in women and men using finite element analysis of CT scans. *J Bone Miner Res* 29:570-580.
23. Eggermont F, Derikx LC, Free J, Van der Linden Y, Verdonschot N, Tanck E. 2016. Effect of Different CT Scanners and Settings on Bone CT values for Finite Element Predicted Failure Load. Oral Presentation at European Society of Biomechanics. Lyon.
24. Sone T, Tamada T, Jo Y, Miyoshi H, Fukunaga M. 2004. Analysis of three-dimensional microarchitecture and degree of mineralization in bone metastases from prostate cancer using synchrotron microcomputed tomography. *Bone* 35:432-438.
25. Kaneko TS, Bell JS, Pejicic MR, Tehranzadeh J, Keyak JH. 2004. Mechanical properties, density and quantitative CT scan data of trabecular bone with and without metastases. *J Biomech* 37:523-530.
26. Kaneko TS, Pejicic MR, Tehranzadeh J, Keyak JH. 2003. Relationships between material properties and CT scan data of cortical bone with and without metastatic lesions. *Med Eng Phys* 25:445-454.
27. Keyak JH, Kaneko TS, Rossi SA, Pejicic MR, Tehranzadeh J, Skinner HB. 2005. Predicting the strength of femoral shafts with and without metastatic lesions. *Clin Orthop* 439:161-170.

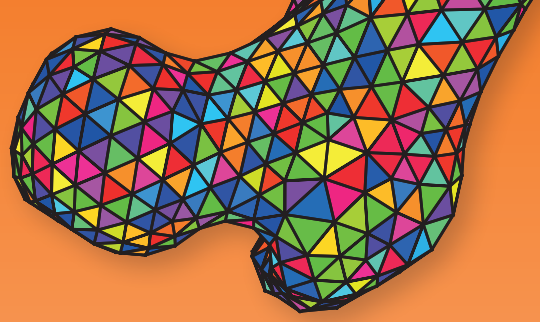
Supplementary material

This figure shows the femoral failure load for patients who did (F) or did not (NF) sustain a femoral fracture during follow-up, corrected for body weight (BW), including the femurs that were affected with blastic metastases, A) in all institutes without considering inter-scanner differences, and in B) institute 1, C) institute 2, and D) institute 3 separately. Femurs with blastic lesions are indicated in grey; femurs with lytic or mixed type lesions are indicated in white. It should be noted that one femur (F87m) fractured one month after follow-up. The thresholds were used to compare the predictive power of the finite element (FE) model *versus* experienced clinicians. \*Femur fractured during unknown activity; all other symbols (#, +, †, \$, ‡, ¶, ¥, §) indicate paired femurs. Femurs with blastic lesions (grey) often have substantially higher failure loads compared with lytic and mixed lesions (white). This is probably due to the fact that these lesions have a high degree of mineralization that would result in inaccurately strong material properties in the FE model, although blastic lesions are generally thought to lead to decreased structural bone strength. Hence, our current FE model is not (yet) able to calculate femurs with blastic lesions.









# CHAPTER 5

## **THE EFFECT OF DIFFERENT CT SCANNERS, SCAN PARAMETERS AND SCANNING SETUP ON HOUNSFIELD UNITS AND CALIBRATED BONE DENSITY: A PHANTOM STUDY**

---

Florieke Eggermont\*, Jeffrey Free\*, Loes Derikx, Ruud van Leeuwen, Yvette van der Linden,  
Wim Jansen, Esther Raaijmakers, Esther Tanck and Robert Kaatee

\*Joint first authorship

Biomedical Physics & Engineering Express 2018; 4(5), 055013.



## Introduction

Patients with femoral bone metastases have an increased risk of sustaining pathological fractures. It is difficult to differentiate between low and high fracture risk, resulting in over- and undertreated patients.<sup>1</sup> In experimental settings, subject-specific finite element (FE) models have been used to calculate bone strength,<sup>2-5</sup> which may be promising for predicting fracture risk in patients with metastatic lesions.<sup>6</sup> Computed tomography scans (CTs), acquired for radiotherapy treatment planning, are often used to build these FE models. From the CTs, the patient-specific bone geometry and bone density can be derived if a calibration phantom with known calcium hydroxyapatite concentrations (CaHA) is scanned concurrently. Using this phantom, Hounsfield units (HU) are converted into calcium equivalent densities, as a measure of bone density, and are related to the mechanical material properties of bone.<sup>7,8</sup> Currently, the FE models for fracture prediction are being evaluated in our Dutch multicentre study including patients with cancer and bone metastases.

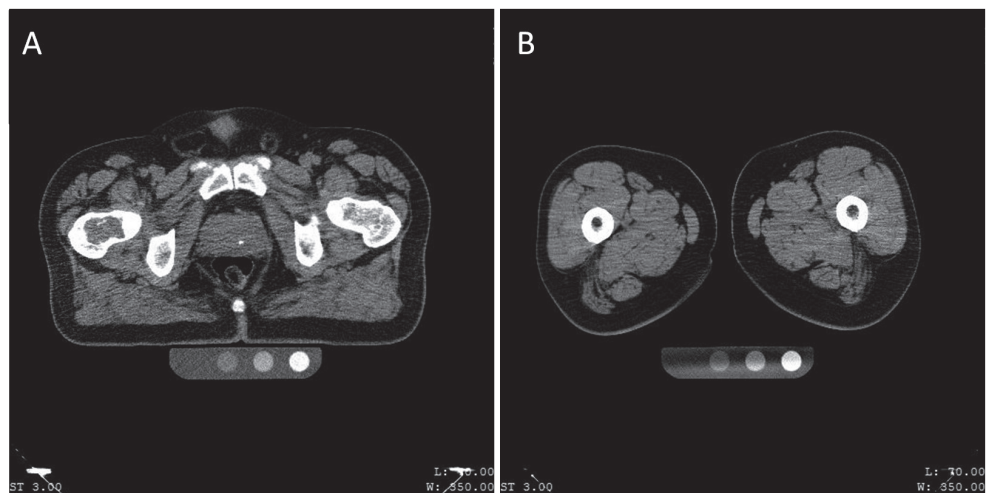
Since modelling mechanical properties in the FE models is dependent on the HU of the CT scan, these should be comparable between scanners. Unfortunately, HU depend on several factors. Firstly, scanner-type<sup>9-12</sup> and scan and reconstruction protocols, including patient and phantom setup,<sup>9,10</sup> have been shown to affect HU. In our patient study, effort was made by the participating centres to minimize HU differences between scanners by using similar standard protocols. However, each CT scanner-type had its own implementation for image reconstruction, which consequently affects HU. Additionally, protocols could be violated, either unintentionally or to optimize preparation of subsequent treatment of the patient. Therefore, it is important to investigate both the effect of different CT scanners and the effect of changes in CT protocols on HU and CaHA.

Secondly, local HU depend on the surrounding material. The accuracy of HU is influenced by for example X-ray scatter and beam hardening.<sup>13,14</sup> As a result, HU of bone in thicker patients will be different from the HU of bone with the same density in thinner patients.<sup>9,15</sup> Furthermore, the surroundings of the calibration phantom could influence the HU within this phantom<sup>16</sup> and consequently the calibration to CaHA. In addition, when scanning patients with painful femoral metastases, knee support cushions are used to reduce the patient's pain during treatment. In such cases, an air gap is present between the legs and the calibration phantom, which can cause a shading artefact in the CT scans that affects the calibration curve (Figure 1).

It is clear that for a correct prediction of fracture risk in our patient study, an accurate conversion of HU to CaHA and mechanical properties of bone is needed. Therefore, HU should be comparable between CT scanners and reconstruction settings, on a level where differences do not significantly affect the outcomes of FE models. Carpenter *et al.*<sup>17</sup> concluded that sources of error in HU and CaHA from different CT manufacturers should be further investigated.

Therefore, the aims of this study were to (1) investigate the effect of different reconstruction parameters on the measured HU; (2) study the inter-scanner differences before (HU) and after calibrating to CaHA; (3) investigate the impact of the patient position within the field of view (FOV) on HU and CaHA; and (4) examine the effect of an air gap between patient and the calibration phantom on CaHA.





**Figure 1:** Two slices of the same CT scan, made with the standard protocol as used in the multicentre patient study<sup>8</sup> (see Table 1). (A): a proximal slice with no air gap between the calibration phantom and the patient, showing no artefact in the calibration phantom. (B): a distal slice with an air gap between the calibration phantom and the patient, showing the shading artefact affecting the calibration phantom.

## Methods

### CT scanners

The four different radiotherapy centres in the Netherlands that accrue patients for the patient study and also participated in the current study, used four different CT scanners from three manufacturers. These comprised two Philips Brilliance Big Bore (Philips Medical Systems, Eindhoven, The Netherlands), one GE Optima CT580 (GE Healthcare, Milwaukee, WI, USA) and one Toshiba Aquillion Large Bore (Toshiba Medical Systems, Tokyo, Japan). All CT scanners were regularly calibrated according to standardized protocols and manufacturer's specifications. The clinically used scan and reconstruction parameters were the same as used in the multicentre patient study (Table 1).<sup>8</sup>

**Table 1:** Clinically used scan and reconstruction parameters of the multicentre patient study.

Scanner type	Philips-1 Brilliance Big Bore	Philips-2 Brilliance Big Bore	GE optima CT580	Toshiba Aquilion Large Bore
kVp	120	120	120	120
mA	variable	variable	variable	variable
Scan slice thickness	0.75 mm × 16	0.75 mm × 16	1.25 mm × 16	0.5 mm × 16
Reconstruction slice thickness	3 mm	3 mm	2.5 mm	3 mm
Reconstruction field of view	480 mm	480 mm	480 mm	480 mm
Pitch	0.813	0.692	0.9375	0.69
Reconstruction kernel	B	B	Standard	FC17
Reconstruction mode	FBP	FBP	FBP	FBP

FBP = Filtered Back Projection

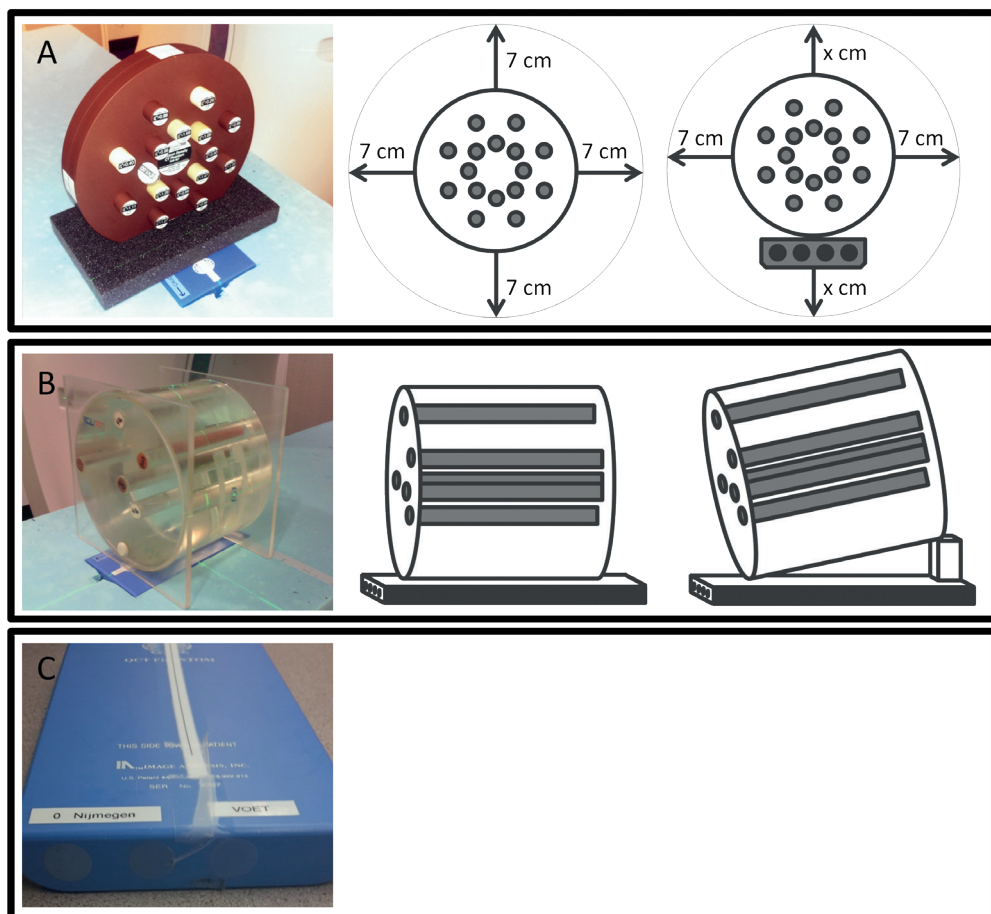
## Phantoms

Three different phantoms were used in this study. All experiments were performed with the same phantoms on each CT scanner, to avoid inter-phantom differences.<sup>11</sup>

Firstly, the Gammex 467 phantom (RMI Gammex, Middleton, WI, USA, Figure 2A) was used, containing 16 inserts with different densities (Table 2). Five inserts contained bone equivalent material (electron density range 1.2 to 2.19 g/cm<sup>3</sup>). The inserts were arranged as suggested in the user guide.

Secondly, since the inserts of the Gammex phantom were quite short, a Perspex water filled phantom with longer inserts was used to investigate the effect of an air gap between patient and calibration phantom. This phantom was developed in-house, consisting of a Perspex cylinder (diameter 300 mm, length 200 mm, Figure 2B), filled with water and containing one RMI insert of fat, one of muscle and three of bone equivalent densities (RMI Gammex, Middleton, WI, USA; Table 2).

Thirdly, the previous described phantoms could be scanned atop a calibration phantom (Image Analysis, Columbia, KY, USA; Figure 2C) containing four known calcium hydroxyapatite concentrations (0, 50, 100 and 200 mg per cm<sup>3</sup> solid water), which could be used to calibrate HU to CaHA. CaHA is a measure of bone density and is used to assign mechanical material properties to the FE models.



**Figure 2:** Overview of the phantoms and corresponding measurements. (A): the Gammex phantom in centred position atop the calibration phantom, and schematic overviews of the displacements in the FOV with and without the calibration phantom. When the calibration phantom was scanned along, the upward and downward displacement was as large as possible while still fitting within the FOV. (B): the custom-made phantom, and schematic overviews of the scans made with and without tilt. (C): the calibration phantom.

**Table 2:** Overview of the phantoms and associated inserts.

Gammex 467 Rod materials	Perspex water filled + RMI inserts			Calibration phantom		
	HU <sub>reference</sub> <sup>a</sup> [g/cm <sup>3</sup> ]	°ρ <sub>reference</sub> [g/cm <sup>3</sup> ]	Rod materials	HU <sub>reference</sub> <sup>b</sup> [g/cm <sup>3</sup> ]	°ρ <sub>reference</sub> [g/cm <sup>3</sup> ]	Rod materials CaHA [mg/cm <sup>3</sup> ]
LN-300 Lung	-684.9	0.32	Fat (Adipose) RMI 453	-101.1	0.9	Rod 0-CaHA 0
LN-450 Lung	-568.2	0.43	Muscle RMI 452	45.3	1.05	Rod 50-CaHA 50
AP6 Adipose	-90	0.91	Inner bone RMI 456	235.8	1.24	Rod 100-CaHA 100
BR-12 Breast	-44.1	0.96	Cortical bone RMI 450-1	1230.6	2.23	Rod 200-CaHA 200
SolidWater-1, 2, 3, 4	-0.8	1	Cortical bone RMI 450-2	1259.8	2.26	
Water	0	1				
BRN-SR2 Brain	25.6	1.03				
LV1 Liver	79.2	1.08				
IB Inner Bone	196.4	1.2				
B200 Bone Mineral	218.7	1.22				
CB2-30% CaCO <sub>3</sub>	438.4	1.44				
CB2-50% CaCO <sub>3</sub>	790	1.79				
SB3 Cortical Bone	1188.6	2.19				

<sup>a</sup>From Gammex user guide (120 kVp—measured with a GE CT/i™ scanner).

<sup>b</sup>Average insert-HU derived from scans from all participating centres without air gap.

### HU and CaHA measurements

For the Gammex phantom, the average HU and standard deviation (SD) of the inserts were determined using equally-sized cylindrical volumes of interest (VOIs) of 6 cm<sup>3</sup> positioned in the centre of the inserts using Mirada RTx Advanced 1.6 (Mirada Medical Ltd, Oxford, United Kingdom). In the scans with the custom-made phantom, average HU and SD were obtained in VOIs of 6 cm<sup>3</sup> positioned in the centre of the inserts. For both the Gammex and custom-made phantom, only results from inserts which contained bone-like materials (*IB Inner Bone* to *SB3 Cortical Bone* and *Inner bone RMI 456* to *Cortical bone RMI 450-2*, respectively) are reported in this paper.

In the scans with the calibration phantom underneath the scanned phantom, a calibration curve was determined using a linear fit between the HU and known CaHA of the four calibration inserts within the calibration phantom. CaHA can be calculated using the intercept (*a*) and the slope (*b*) from the linear fit where  $CaHA = a + (b * HU)$ .

Because mean HU could be depending on placement of the VOI and the size of the VOI, we additionally investigated the accuracy of the mean HU and SD measurement using VOIs by shifting the VOI of the *SB3 Cortical Bone* insert by 2 mm to the left and downwards, and by decreasing the VOI volume by 50% on the scans that determined the effect of reconstruction parameters (see 2.4 Effect of reconstruction parameters). Moreover, we performed intra scanner reliability measurements on all CT scanners by scanning the Gammex phantom twice using the standard protocol. Between the two scans, the Gammex phantom was removed and repositioned on the table.

To investigate the quality of the images for the scans that were made to determine the effect of reconstruction parameters (see 2.4 Effect of reconstruction parameters), we report the contrast-to-noise ratio (CNR) and the SD. The CNR is defined as the difference between the mean HU of the insert and the mean HU of the background, divided by the SD of the background. We defined differences less than 5 HU as small.

### Effect of reconstruction parameters

To investigate the effect of different reconstruction parameters, the Gammex phantom was scanned on each of the scanners using a clinical (standard) reconstruction protocol (3 mm slice thickness, 480 mm FOV, standard reconstruction kernel) and with variations in reconstructed slice thickness (1 mm), FOV (550 mm) and reconstruction kernel (detail using increased edge enhancement, Table 3). For each variation, the difference in HU relative to the standard protocol was calculated ( $\Delta HU$ ). The kernels used in the standard protocol are used clinically for, but not limited to, the pelvic region. The detail kernel was mainly used for scanning of the brain.

**Table 3:** Overview of the investigated reconstruction protocols.

Scan parameters	kVp: 120, mAs: 250, detector slice thickness: 1 mm, Orientation: Head First Supine			
Reconstruction protocols	Standard protocol	Variation slice thickness	Variation FOV	Variation Kernel
<b>Philips-1 Brilliance Big Bore</b>	Slice: 3 mm	Slice: 1 mm	Slice: 3 mm	Slice: 3 mm
	FOV: 480 mm	FOV: 480 mm	FOV: 550 mm	FOV: 480 mm
	Pixel Spacing: 0.9375	Pixel Spacing: 0.9375	Pixel Spacing: 1.0742	Pixel Spacing: 0.9375
	Kernel: B	Kernel: B	Kernel: B	Kernel: UB
<b>Philips-2 Brilliance Big Bore</b>	Slice: 3 mm	Slice: 1 mm	Slice: 3 mm	Slice: 3 mm
	FOV: 480 mm	FOV: 480 mm	FOV: 550 mm	FOV: 480 mm
	Pixel Spacing: 0.9375	Pixel Spacing: 0.9375	Pixel Spacing: 1.0742	Pixel Spacing: 0.9375
	Kernel: B	Kernel: B	Kernel: B	Kernel: D
<b>GE Optima CT580</b>	Slice: 2.5 mm	Slice: 1.25 mm	Slice: 2.5 mm	Slice: 2.5 mm
	FOV: 480 mm	FOV: 480 mm	FOV: 550 mm	FOV: 480 mm
	Pixel Spacing: 0.9375	Pixel Spacing: 0.9375	Pixel Spacing: 1.0742	Pixel Spacing: 0.9375
	Kernel: standard	Kernel: standard	Kernel: standard	Kernel: detail
<b>Toshiba Aquilion Large Bore</b>	Slice: 3 mm	Slice: 1 mm	Slice: 3 mm	Slice: 3 mm
	FOV: 480 mm (LL)	FOV: 480 mm (LL)	FOV: 550 mm (LL)	FOV: 480 mm (LL)
	Pixel Spacing: 0.935	Pixel Spacing: 0.935	Pixel Spacing: 1.074	Pixel Spacing: 0.935
	Kernel: FC17	Kernel: FC17	Kernel: FC17	Kernel: FC43

FOV = Field of View. Slice = Reconstructed slice thickness.

### Inter-scanner differences

To investigate the inter-scanner differences, the Gammex phantom was scanned with and without the calibration phantom on each CT scanner, using the standard protocols (Table 3). The Gammex phantom was aligned with the isocentre of the CT scanner. Differences (measured values—reference values) in HU ( $\Delta$ HU) and CaHA ( $\Delta$ CaHA) were determined relative to the HU reported in the phantom user guide. The reference CaHAs were calculated by calibrating the HU reference values using the average calibration curve of all scanners.

### Position in the field of view

To study the effect of patient position in the scanner on HU and CaHA, the Gammex phantom was scanned at five different positions in the FOV (i.e. isocentre, left, right, up and down) with and without the calibration phantom underneath (Figure 2A). The four Gammex phantom displacements relative to the isocentre were 7 cm. When the Gammex phantom was displaced to left or right, the calibration

phantom remained centred, to mimic a patient lying off axis on the table. Upwards and downwards, both the Gammex and calibration phantoms were displaced by changing table height, while still fitting in the FOV. Upwards the displacement was also 7 cm and downwards between 3 cm and 7 cm, depending on whether the phantoms still fitted in the FOV. Scans were made according to the standard protocol. For each of the five inserts containing bone equivalent materials, the reference scan was defined as the scan in which the insert was closest to the isocentre. The differences in HU between the reference scan and the other scans were calculated, and were calibrated to CaHA based on the calibration phantom of the scan with the Gammex phantom positioned in the isocentre.

### Air gap between patient and calibration phantom

To investigate the effect of an air gap, the custom-made phantom was scanned twice using the standard protocol: once flat atop the calibration phantom (reference scan), and once while one side of the custom-made phantom was lifted 3 cm to induce tilt and, consequently, an air gap increasing in the longitudinal scan direction (Figure 2B). The relative difference in CaHA between the scan with the tilted phantom and the reference scan (Table 4) was determined at three positions, for air gaps of 0.5, 1.5 and 2.5 cm, respectively.

**Table 4:** CaHA reference values.

Scanner type	Philips-1 Brilliance Big Bore	Philips-2 Brilliance Big Bore	GE Optima CT580	Toshiba Aquilion Large Bore
Fat (Adipose) RMI 453	-99	-98	-69	-93
Muscle RMI 452	32	36	58	40
Inner bone RMI 456	207	206	226	210
Cortical bone RMI 450-1	1107	1099	1090	1117
Cortical bone RMI 450-2	1142	1122	1117	1136

## Results

### Measurement accuracy and image quality

We determined the effects of 2 mm left and downwards shift of the VOI and the effect of decreasing the VOI volume by 50% on the HU in the *SB3 Cortical Bone* insert. When using the scans obtained with the standard protocol, as well as scans with a different slice thickness or FOV, the effects of each of the changes in VOI on mean HU were small on all of the CT scanners (range -2.2 HU to 2.3 HU). On the scans with a different kernel, the effects of changing VOI placement were similar on the Philips-1, Philips-2 and GE scanners (range -2.3 to 2.4), while the effect was somewhat larger on the Toshiba scanner (range -7.9 HU to 5.4 HU).

To determine the reproducibility of the results, we performed intra scanner reliability measurements of all CT scanners, and found that differences in mean HU between CT scans on the same scanner made with the standard protocol were on average 0.2 HU, with a maximal difference of 2.2 HU on the Philips-1 scanner in the CB2–30% CaCO<sub>3</sub> insert.

We investigated the CT scan quality by calculating the CNR and SDs of all bone inserts as a measure of noise. For the scans made with the standard protocol, the CNR of all bone inserts was largest on the Toshiba scanner (range 30.4 to 169.8) and smallest on the Philips-1 scanner (range 11.2 to 62.3). CNRs were comparable between the Philips-1 and Philips-2 scanner (ranges 11.2 to 62.3 and 12.9 to 72.2 over all inserts, respectively). The SDs were largest on the Philips-2 scanner (range 19.6 to 29.8 HU over all inserts) and smallest on the Toshiba scanner (range 2.9 to 11.2 HU over all inserts). On the Philips-1 and GE scanners, the SDs were comparable (range 13.5 to 20.3 HU and 13.6 to 21.0 HU over all inserts, respectively). When slice thickness was changed from 3 mm to 1 mm, CNR decreased on average for all scanners with 30% (range -47.1 to -4.0%, -79.9 to -0.5) and the SD increased on all scanners with 62% on average (range 28% to 85%, 3.9 to 21.2 ΔHU). CNR increased with 30% on average of all bone inserts (range 6.0% to 70.1%, 0.8 to 47.9) after changing the FOV from 480 to 550 mm, while SD slightly decreased with 15% on average (range -26 to -7%, -4.2 to -1.1 ΔHU). Changing kernel to a detail kernel had different effects on the CT scanners. On the Toshiba scanner, the CNR decreased 18% on average (range -18.4 to -18.2, -31.2 to -5.5). An increase of 42% in CNR over all bone inserts was seen on the Philips-1 and Philips-2 scanner (range 8.3% to 68.3%, 1.1 to 42.5). On Philips-1 and Toshiba, the SD increased (average of 6%, range 63% to 67%, 8.5 to 13.5 ΔHU and average 49%, range 29% to 92%, 2.8 to 10.3 HU, respectively), while the SD decreased when kernel was changed on the Philips-2 scanner (average 26%, range -27 to -23%, -7.9 to -4.6 ΔHU). On GE, there was no clear effect on CNR (average 2%, range 1.4% to 1.7%, 0.3 to 1.7) or SD (average 0%, range -4% to 4%, -0.5 to 0.9 ΔHU) due to changing kernel.

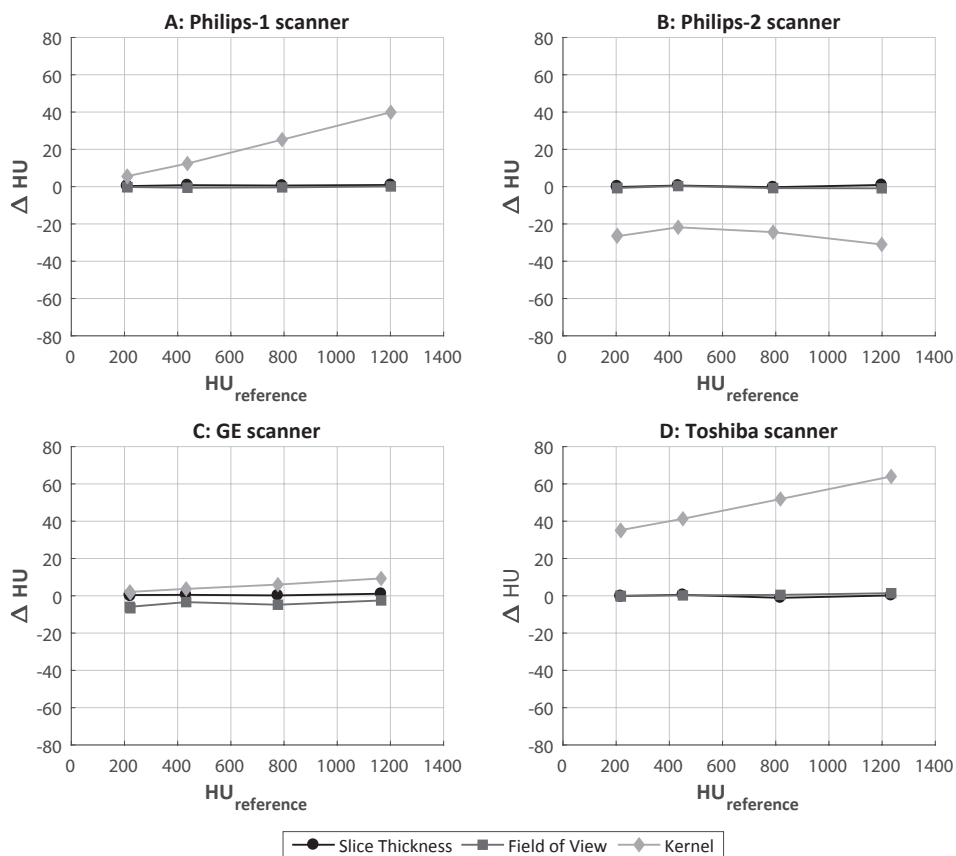
### Effect of reconstruction parameters

When changing the reconstruction slice thickness from 3 mm to 1 mm, the effect on HU was small on all scanners (<1.1 HU, Figure 3).

The variation in reconstruction FOV parameters (480 mm to 550 mm) showed a small effect on HU for Philips-1, Philips-2 and Toshiba scanners (≤1.4 HU, Figures 3A, B, D). On the GE scanner the effect was larger, with an average ΔHU of -4.7 HU (range -6.7 to -2.5 HU). The largest effect of FOV was found for the *IB Inner Bone* insert (Figure 3C).

The effect of varying reconstruction kernel (standard to detail) was clearly larger than the effects of the other reconstruction parameters, except for the GE scanner for which HU were relatively small for both reconstruction kernels (<9.3 HU, Figure 3C). The maximal ΔHU of the bone materials for the other scanners ranged from -31.0 HU to 64.0 HU, with the Toshiba scanner showing the largest average ΔHU of 45.5 HU (range 34.9 to 64.0 HU, Figure 3D).



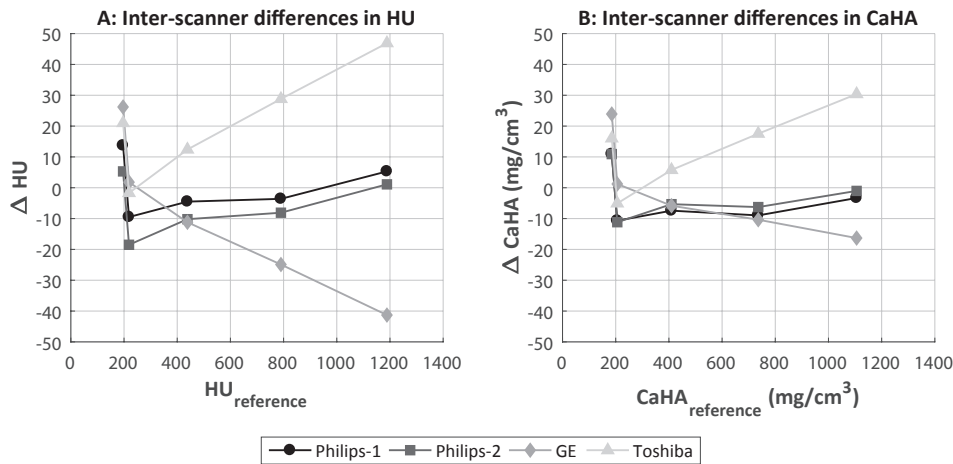


**Figure 3:** Effect of changes of three different reconstruction parameters (i.e. slice thickness, FOV and kernel) on differences in HU relative to the standard scan protocol on the Philips-1 (A), Philips-2 (B), GE (C) and Toshiba (D) scanners. Every dot represents an insert with bone-like materials of the Gammex phantom.  $\Delta$ HU represents the difference between the variation and the standard setting.

### Inter-scanner differences

Inter-scanner differences were largest between the Toshiba and GE scanners in the *SB3 Cortical Bone* insert (88.2 HU), while the smallest inter-scanner differences were found between Philips-2 and GE in the *CB2–30% CaCO<sub>3</sub>* insert (1.0 HU) (Figure 4A).

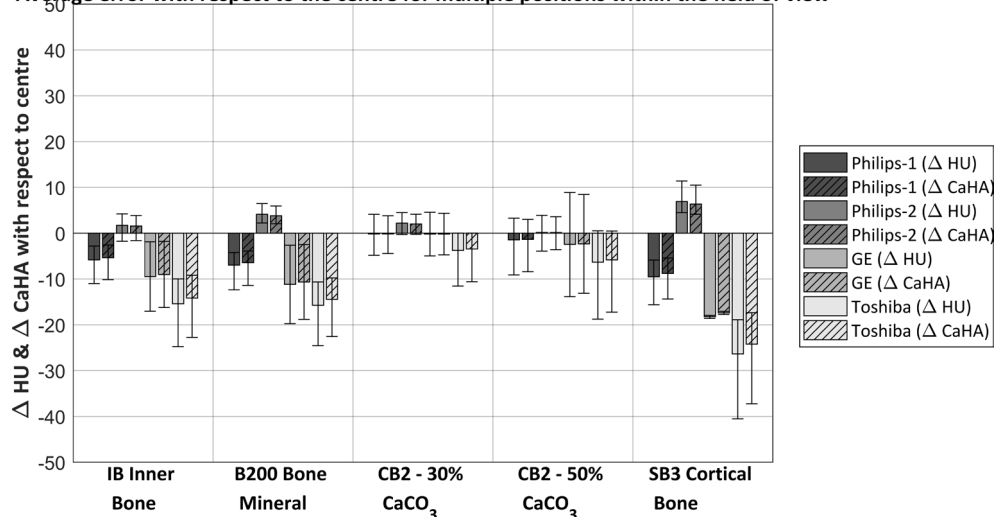
After calibrating the HU to CaHA, the inter-scanner variation reduced in all inserts (Figure 4B). The difference in CaHA between scanners remained largest between the Toshiba and GE scanners in the *SB3 Cortical Bone* insert, although the difference in CaHA was smaller than before calibration (46.7 CaHA). The smallest difference in CaHA was found between Philips-1 and Philips-2 in the *IB Inner Bone* insert (0.0 CaHA).



**Figure 4:** Inter-scanner differences in  $\Delta$ HU (A) and after calibration in  $\Delta$ CaHA (B). Every dot represents an insert with bone-like material of the Gammex phantom.  $\Delta$ HU and  $\Delta$ CaHA represent the difference between the measured HU and the reference HU obtained from the Gammex manual or the measured CaHA and reference CaHA obtained from an average calibration of the corresponding HUs.

### Position of the scanned object in the field of view

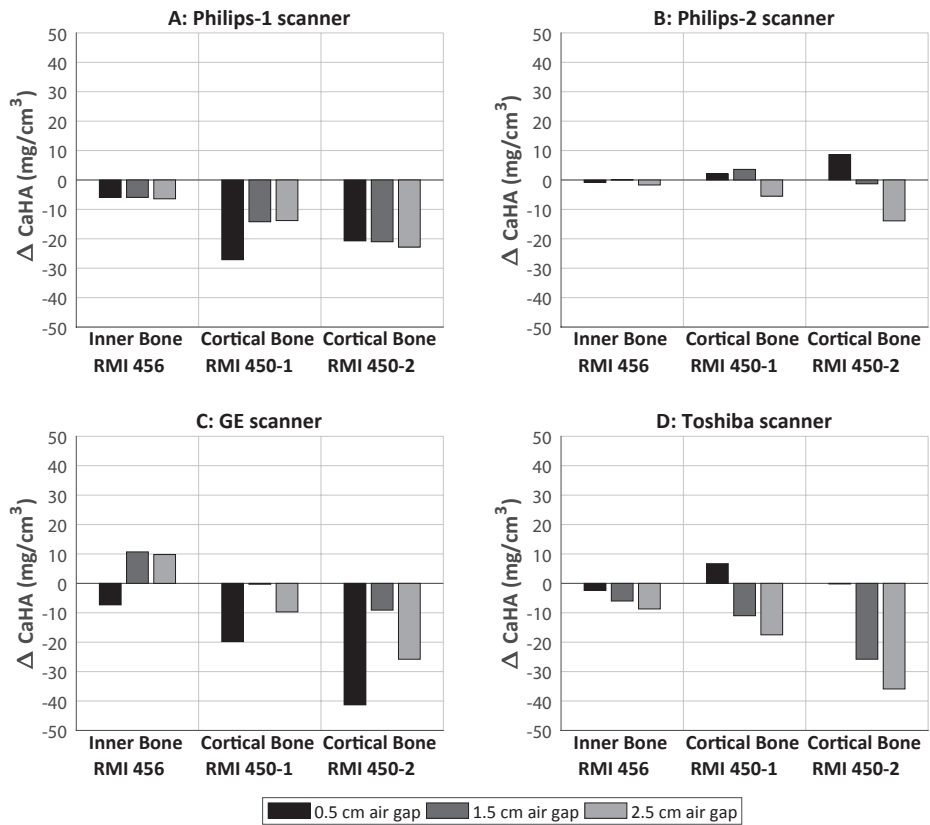
Changes in position in FOV resulted in the largest average  $\Delta$ HU and  $\Delta$ CaHA for the *SB3 Cortical Bone* insert on the Toshiba scanner (-26.4 HU and -24.2 CaHA, Figure 5), and the smallest average error for both the *CB2-CaCO<sub>3</sub>* inserts on the Philips-1 scanner (-0.2 HU and -0.2 CaHA). Calibration from HU to CaHA concentrations did not result in a large reduction of the error.

**Average error with respect to the centre for multiple positions within the field of view**

**Figure 5:** Effect of position within the FOV of each of the bone-like materials. Average deviation in  $\Delta$ HU (solid) and  $\Delta$ CaHA (dashed) over all four positions in the FOV (left, right, up, down) relative to the most centred insert. Error bars represent maximum and minimum effects.

### Tilt of the phantom on top of the calibration phantom

The largest effect of tilt was found in the *Cortical Bone RMI 450-2* insert on the GE scanner when inducing a 2.5 cm air gap, (-41.3 CaHA, Figure 6). The difference in  $\Delta$ CaHA between the different air gaps was the largest for the *Cortical Bone RMI 450-2* insert on the Toshiba scanner (-35.9 to -0.2 CaHA) and the smallest for the *Inner Bone RMI 456* insert from the Phillips-1 scanner (-5.9 to -5.9 CaHA). With an air gap of 0.5 cm the largest difference between scanners was 49.9  $\Delta$ CaHA between P2 and GE in the *Cortical Bone RMI 450-2* insert. An air gap of 1.5 and 2.5 cm showed the largest inter-scanner range between Philips-2 and Toshiba of 24.5 and 22  $\Delta$ CaHA, both in the *Cortical Bone RMI 450-2* inserts.



**Figure 6:** The effect of air gaps of 0.5, 1.5 and 2.5 cm, induced by tilting the custom-made phantom on top of the calibration phantom in  $\Delta\text{CaHA}$  compared to the reference scan without tilt in the three bone equivalent inserts (Inner Bone RMI 456, Cortical Bone RMI 450-1 and Cortical Bone RMI 450-2) on the Philips-1 (A), Philips-2 (B), GE (C) and Toshiba (D) scanners.

## Discussion

This study investigated the effect of different CT scanners, reconstruction protocols, scan positions, and air gaps on HU and/or calibrated CaHA.

First, we tested the effect of variations in reconstruction parameters on HU. When changing slice thickness or FOV, differences in HU with respect to the standard protocol were negligible ( $\leq 1.8$  HU, with exception of FOV on the GE scanner ( $\leq 6.3$  HU)). However, variations in reconstruction kernel had a larger effect on the HU ranging up to 66.0 HU. The precise cause of this effect remains unclear since reconstruction kernels are manufacturer-specific black boxes known to affect HU to, for example, increase visibility of small objects. Our results are in line with Birnbaum *et al.*,<sup>10</sup> who concluded that different kernels could result in large differences in HU.

When looking at the inter-scanner deviations the largest difference in HU between CT scanners was observed in the highest density insert. The inter-scanner differences increased in the higher density inserts representing cortical bone. Previously, several studies reported relevant inter-scanner differences.<sup>9-12</sup> It was stated that even scans from scanners of the same manufacturer could yield different attenuation values.<sup>9,10</sup> In addition, it has been reported that inter-scanner variations are larger for densities deviating more from water.<sup>11</sup> These findings are in correspondence with our results. After calibration to CaHA, the inter-scanner variation reduced for all inserts. Few studies reported the effect of calibration with the use of a calibration phantom on the inter-scanner differences. It has been suggested that calibration phantoms can correct for differences between CT scanners.<sup>7</sup> However, Carpenter *et al.*<sup>17</sup> showed that inter-scanner differences in bone density, after calibration, existed in both phantoms and patients, although they did not report on the inter-scanner differences before calibration. Nevertheless, our results showed that inter-scanner variations were reduced when the HU were calibrated to CaHA using a calibration phantom.

As a third aim, we investigated the effect of position in the FOV. In this case, the average error after calibration did not decrease substantially. Previous studies showed that changing phantom position within the FOV changed the HU of the scanned material significantly.<sup>9,15</sup> Therefore, our findings of increased error when shifting the Gammex phantom in any direction were as expected. Nevertheless, our results are an important finding as we have now quantified the expected effects of changes in position within the FOV of the specific CT scanners.

Our last objective was to determine the effect of air gaps on the calibration to CaHA. Tilting the phantom on top of the calibration phantom affected the CaHA error substantially. It appeared that different scanners responded differently to air gaps. Overall, air between the calibration phantom and the scanned object seemed to have an effect on the calibration curve, which is of importance in patient studies, since patient scans often include air gaps of approximately 3 cm over 20 cm length due to knee support. From this study, we suggest that calibration should only be done using slices without an air gap between calibration phantom and patient, for example in more proximal areas.

Remarkable for all results were the differences between the Philips-1 and Philips-2 scanners. The use of different pitches on the Philips-1 scanner resulted in small differences ( $< 8$  HU) for the bone inserts,

which was similarly reported by Hopper *et al.*<sup>18</sup> Therefore, the difference in pitch cannot explain the differences observed between the Philips scanners. Also small differences were found in beam quality measurements of kVp and half-value layer (HVL) between the Philips-1 (122.0 kVp, 8.77 mm Al) and Philips-2 (122.6 kVp, 8.92 mm Al) measured with a PTW Nomex Multimetric (PTW, Freiburg, Germany). Nevertheless, the use of two CT scanners from the same manufacturer resulted in interesting findings, showing that the same scanners could still yield different outcomes.<sup>9</sup>

There were some limitations to our study. First of all, as the calibration phantom did not contain high density calibration inserts, calibration was based on values between 0 and 200 CaHA, which were extrapolated to higher CaHA for example for cortical bone density. This extrapolation method makes the calibration of denser materials prone to errors. However, there are also advantages of missing high density inserts in the calibration phantom. It prevents unnecessary CT reconstruction artefacts (e.g. photon starvation) in patient scans<sup>19</sup> and ensures that the calibration phantom will not influence the dose calculation accuracy, when used in radiotherapy planning CT scans. In addition, we applied the same calibration method as used in our patient study,<sup>8</sup> and therefore we got a representative insight in the effects on the FE models. Secondly, the CaHA calculations could not be compared to a gold standard. Therefore, only relative differences based on HU from the Gammex manual or based on an average from all CT scanners could be reported. All HU presented in this paper were measured without calibration phantom underneath the Gammex to mimic the situation of making a CT scan, which can be seen as an uncalibrated bone density scan. To obtain the calibrated bone density (CaHA) scans, CT scans including the calibration phantom were made. The CaHA were retrieved from the calibration. The additional presence of the calibration phantom in the CT scan had a negligible effect on the HU. Based on the analysis of the Gammex phantom, we can state that the HU scale of the bone-like materials was linearly correlated with mass density for all CT scanners. We obtained standard deviations within each VOI as a measure of noise and determined the CNR, to determine the quality of the images. In general, we found that the CNR decreased if the SD increased, which can be expected. In contrast, both the SD and the CNR increased when changing kernel to detail on the Philips-1 scanner, suggesting that this detail reconstruction kernel had a larger effect on higher density materials than lower density materials. We found that the CNR was highest on the Toshiba scanner, while the noise was lower for the Toshiba scanner compared to the other scanners. The CT dose index volume (CTDI<sub>vol</sub>) can be a reason for this difference: the average CTDI<sub>vol</sub> value for the Toshiba scanner (66.4 mGy) was higher than for the other scanners (26.3, 19.9 and 33, 7 mGy for the Philips-1, Philips-2 and GE, respectively). Differences in image quality can also be explained other factors, such as the use of different detector materials, post processing techniques and different implementation of variable dose by different manufacturers. Moreover, we investigated the effect of changes in VOI placement, which were relatively small, indicating that slight changes the VOI placement would not yield large differences in our results. Additionally, we performed intra scanner reliability measurements on all CT scanners, and found that the reproducibility of our CT scans was good (<2.2 HU).

We realize that HU depend on the attenuation coefficients which depend on the atomic numbers, electron densities and on the beam energy  $E$ .<sup>20</sup> The mechanical properties of the patient-specific FE

models are assumed to be related on the effective atomic number of bone. Because of the dependence of HU on beam energy, we kept the energy-setting constant in our experiments. However, as a result, differences in the energy spectra of the different scanners could have been present.

In the current study, we focused on the effects of changes in CT scanner, CT protocol, position within FOV, and air between calibration phantom and patient, but future research may also look into factors such as beam hardening due to patient size, blurring due to patient motion, and inaccurate HU representation due to metal artefacts and patient motion.<sup>19</sup>

For the patient-specific FE models, HU from CT scans are converted to CaHA.<sup>3,4,7,8</sup> From these CaHA, mechanical material properties are calculated. Keyak *et al.*<sup>7</sup> described the relationship between CaHA and Ultimate Strength for trabecular and cortical bone. In the current study, the largest effects were found in the inserts that represent cortical bone, which is expected to have the largest effect on the structural strength of the femur, while the effects for inserts representing trabecular bone were smaller. However, because the strength also depends on shape and structure, it is difficult to predict the effect size of the determined  $\Delta\text{CaHA}$  on the failure load calculated by the FE model. To what extent the presented effects influence the fracture risk predictions, should be investigated in future research. Additionally, in the future, Dual Energy or Spectral CT could give more information on chemical composition. This might provide extra information for the FE models.

In conclusion, this study showed that several factors can induce variations in HU as well as calibrated CaHA. Different reconstruction kernels showed significant variation in HU between scanners, whereas slice thickness and FOV had negligible effect. Inter-scanner differences in HU reduced after calibrating to CaHA. Changes in position within the FOV should be avoided, as this resulted in relevant variations in HU and CaHA. Finally, air gaps between calibration phantom and patient could lead to errors in CaHA and should be prevented. To know what requirements are needed for CT scans used for predicting femoral fracture risk, the effects of the different variations on the calculations by the FE models should be quantified in future studies.

## Acknowledgements

The authors would like to thank Dr Arjen van der Schaaf for his input to the manuscript. Also, the authors thank all radiation therapists for their help with the CT scanners.

This project was partially funded by the Dutch Cancer Society (KUN 2012-5591).

## References

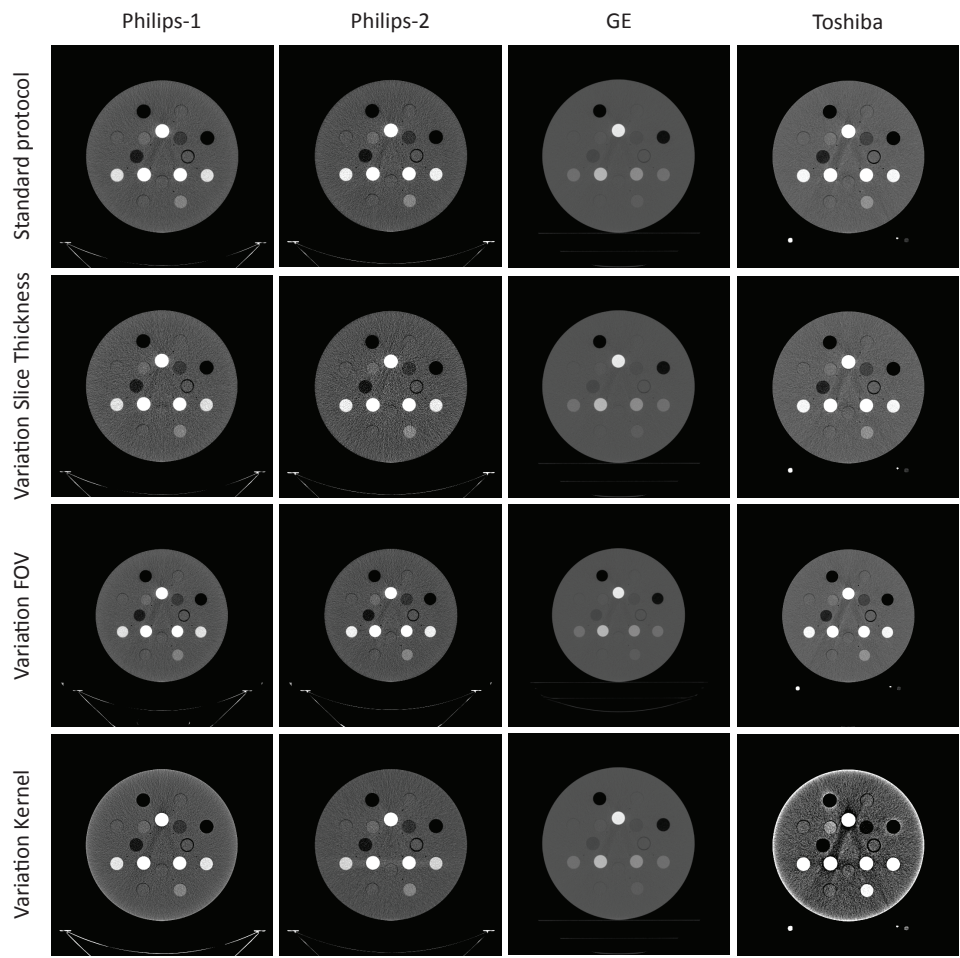
1. Van der Linden YM, Dijkstra PD, Kroon HM, Lok JJ, Noordijk EM, Leer JW, Marijnen CA. 2004. Comparative analysis of risk factors for pathological fracture with femoral metastases. *J Bone Joint Surg Br* 86:566-573.
2. Bessho M, Ohnishi I, Matsuyama J, Matsumoto T, Imai K, Nakamura K. 2007. Prediction of strength and strain of the proximal femur by a CT-based finite element method. *J Biomech* 40:1745-1753.
3. Tanck E, van Aken JB, van der Linden YM, Schreuder HW, Binkowski M, Huizenga H, Verdonchot N. 2009. Pathological fracture prediction in patients with metastatic lesions can be improved with quantitative computed tomography based computer models. *Bone* 45:777-783.
4. Derikx LC, van Aken JB, Janssen D, Snyers A, van der Linden YM, Verdonchot N, Tanck E. 2012. The assessment of the risk of fracture in femora with metastatic lesions: comparing case-specific finite element analyses with predictions by clinical experts. *J Bone Joint Surg Br* 94:1135-1142.
5. Dragomir-Daescu D, Salas C, Uthamaraj S, Rossman T. 2015. Quantitative computed tomography-based finite element analysis predictions of femoral strength and stiffness depend on computed tomography settings. *J Biomech* 48:153-161.
6. Derikx LC, Verdonchot N, Tanck E. 2015. Towards clinical application of biomechanical tools for the prediction of fracture risk in metastatic bone disease. *J Biomech* 48:761-766.
7. Keyak JH, Kaneko TS, Tehranzadeh J, Skinner HB. 2005. Predicting proximal femoral strength using structural engineering models. *Clin Orthop*:219-228.
8. Eggermont F, Derikx LC, Verdonchot N, Van der Geest ICM, De Jong MAA, Snyers A, Van der Linden YM, Tanck E. 2018. Can patient-specific finite element models better predict fractures in metastatic bone disease than experienced clinicians? Towards introducing computational modelling into daily clinical practice. *Bone Joint Res* 7:430-439.
9. Levi C, Gray JE, McCullough EC, Hattery RR. 1982. The unreliability of CT numbers as absolute values. *AJR Am J Roentgenol* 139:443-447.
10. Birnbaum BA, Hindman N, Lee J, Babb JS. 2007. Multi-detector row CT attenuation measurements: assessment of intra- and interscanner variability with an anthropomorphic body CT phantom. *Radiology* 242:109-119.
11. Sande EP, Martinsen AC, Hole EO, Olerud HM. 2010. Interphantom and interscanner variations for Hounsfield units--establishment of reference values for HU in a commercial QA phantom. *Phys Med Biol* 55:5123-5135.
12. Mackin D, Fave X, Zhang L, Fried D, Yang J, Taylor B, Rodriguez-Rivera E, Dodge C, Jones AK, Court L. 2015. Measuring Computed Tomography Scanner Variability of Radiomics Features. *Invest Radiol* 50:1-9.
13. Hsieh J, Molthen RC, Dawson CA, Johnson RH. 2000. An iterative approach to the beam hardening correction in cone beam CT. *Med Phys* 27:23-29.
14. Siewerdsen JH, Daly MJ, Bakhtiar B, Moseley DJ, Richard S, Keller H, Jaffray DA. 2005. A simple, direct method for x-ray scatter estimation and correction in digital radiography and cone-beam CT. *Med Phys* 33:187-197.
15. Hunter TB, Pond GD, Medina O. 1983. Dependence of substance CT number on scanning technique and position within scanner. *Comput Radiol* 7:199-203.



16. Boden SD, Goodenough DJ, Stockham CD, Jacobs E, Dina T, Allman RM. 1989. Precise measurement of vertebral bone density using computed tomography without the use of an external reference phantom. *J Digit Imaging* 2:31-38.
17. Carpenter RD, Saeed I, Bonaretti S, Schreck C, Keyak JH, Streeper T, Harris TB, Lang TF. 2014. 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* 36:1225-1232.
18. Hopper KD, Gouldy CA, Kasales CJ, TenHave TR, Fisher AL. 1997. The effect of helical CT on X-ray attenuation. *J Comput Assist Tomogr* 21:152-155.
19. Barrett JF, Keat N. 2004. Artifacts in CT: Recognition and avoidance. *Radiographics* 24:1679-1691.
20. Jackson DF, Hawkes DJ. 1981. X-Ray Attenuation Coefficients of Elements and Mixtures. *Phys Rep* 70:169-233.

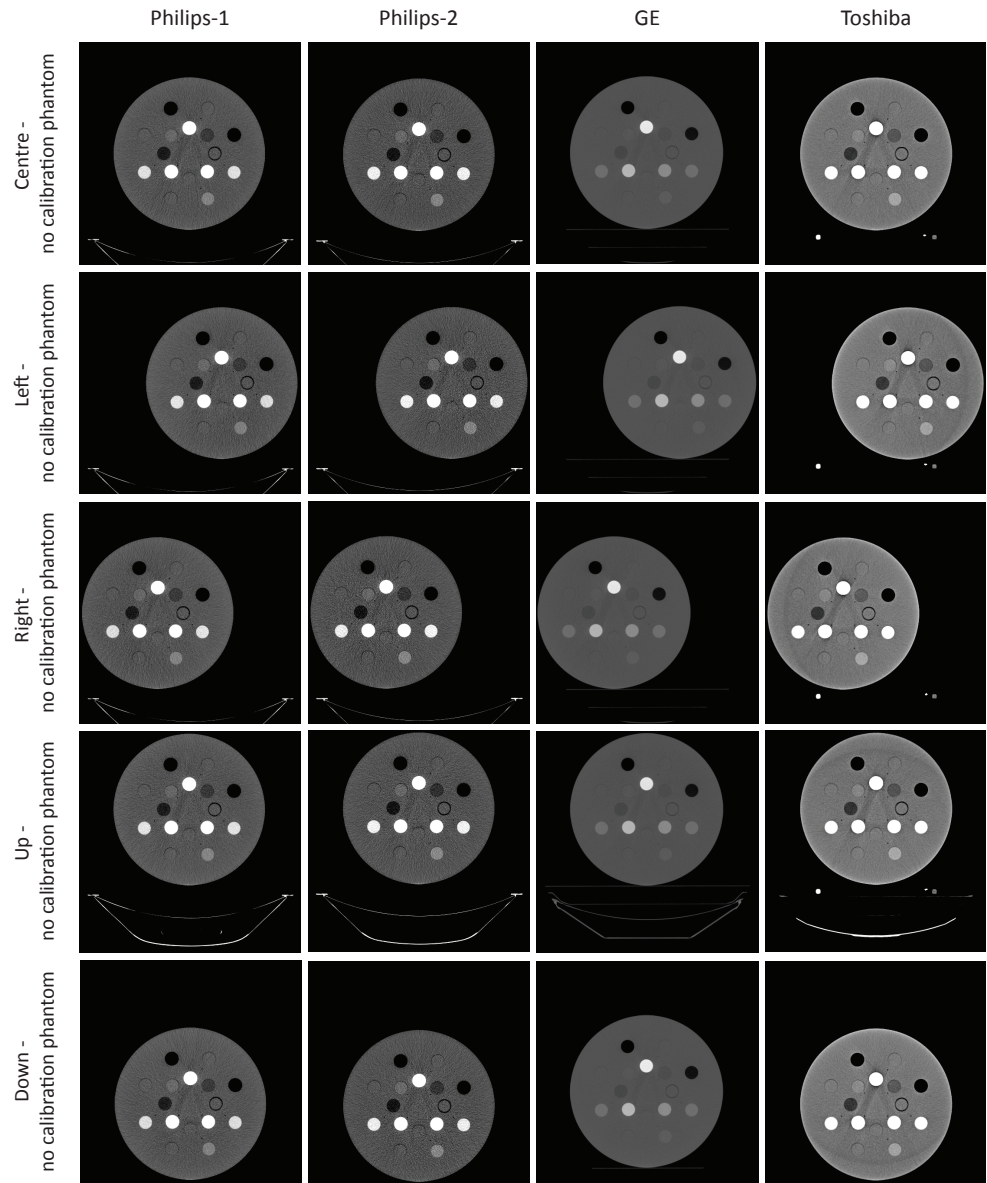
Supplementary material

FBP reconstructions of one slice of all CT scans  
*Effect of reconstruction parameters and inter-scanner differences*

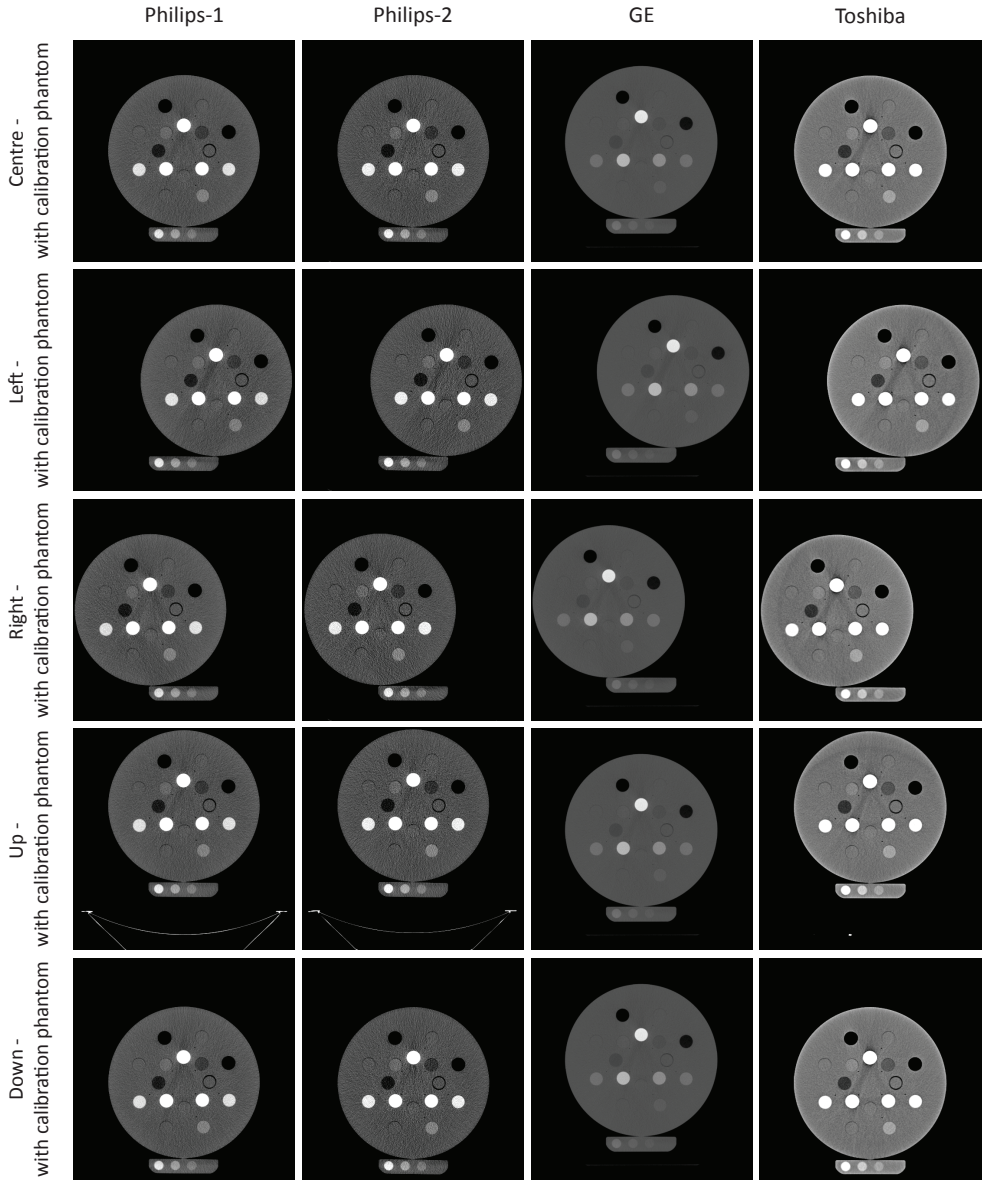


5

*Position of the scanned object in the field of view – no calibration phantom*

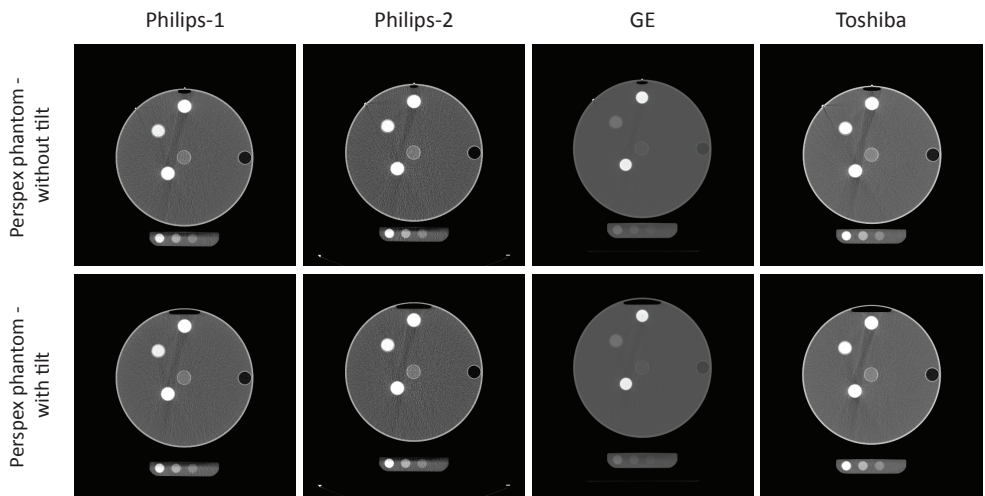


*Position of the scanned object in the field of view – with calibration phantom*

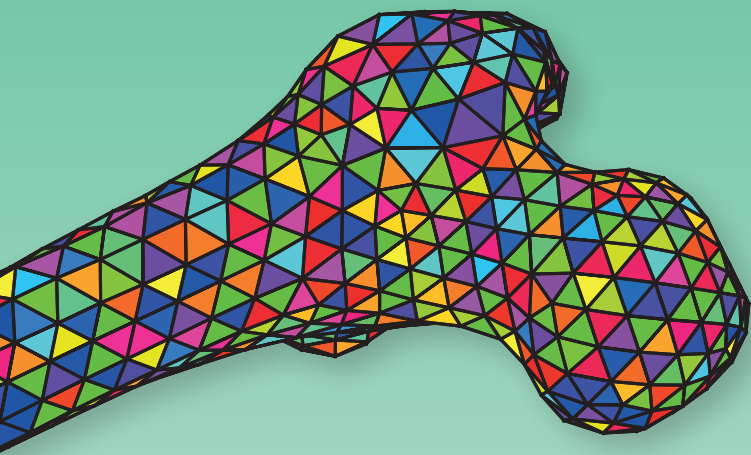


5

*Tilt of the phantom on top of the calibration phantom*









# CHAPTER 6

## **EFFECT OF DIFFERENT CT SCANNERS AND SETTINGS ON FEMORAL FAILURE LOADS CALCULATED BY FINITE ELEMENT MODELS**

---

Florieke Eggermont\*, Loes Derikx\*, Jeffrey Free, Ruud van Leeuwen, Yvette van der Linden,  
Nico Verdonschot, Esther Tanck

\*Joint first authorship

Journal of Orthopaedic Research 2018; 36(8), 2288-2295.





## Introduction

Bone metastases in patients with advanced cancer are very common. These metastases cause pain and induce a certain risk of pathological fracture. Predicting the fracture risk is important for deciding on treatment strategy, since patients with a low fracture risk are conservatively treated with local radiotherapy to relieve pain, whereas patients with a high fracture risk undergo stabilizing prophylactic surgery.<sup>1,2</sup> However, in present clinical practice, it appears to be difficult to distinguish between low and high fracture risk patients, causing a large number of over- and undertreated patients.<sup>1</sup>

Subject-specific finite element (FE) models are a promising tool in calculating strength of femora with (artificial) metastatic lesions. Experimentally, these FE models have shown promising results for calculation of fracture risk.<sup>3-5</sup> For such FE models, quantitative computed tomography (QCT) scans are used to segment the subject-specific femur geometry. Also, bone mineral density (BMD) is often calculated with the use of a calibration phantom under the patient. The subject-specific geometry and BMD are used as input for the FE models. Recently, studies using FE models showed promising results in discriminating patients with a low fracture risk from patients with a high fracture risk.<sup>6,7</sup>

However, it has been shown that scanning the same subject using comparable protocols on different CT scanners can result in different Hounsfield units (HU).<sup>8-11</sup> When comparing high versus low resolution CT scans, Dragomir-Daescu *et al.*<sup>12</sup> showed that there were differences in FE strength in a fall configuration of maximally 1,500 N (~45%). Another study showed that the FE calculated failure load could differ up to 2,500 N (~23%) when simulating a single-leg stance induced fracture based on CT scans of a healthy subject scanned on two different CT scanners.<sup>13</sup> The authors additionally showed that accounting for these inter-scanner differences is difficult.

Another problem arises when using clinical CT scans as input for the FE models. Potential changes in CT settings by deviating from a standard protocol may influence HU and subsequently the outcome of FE models.<sup>14</sup> Additionally, it appears that air between calibration phantom and patient induces an artifact in the calibration phantom. Such air gaps are common when scanning cancer patients, as patients' knees are often placed on a cushion to relieve pain. These air artifacts have been described before,<sup>11,15</sup> although it is unclear how it affects the calibration to *in vivo* BMD values and subsequent calculation of failure load.

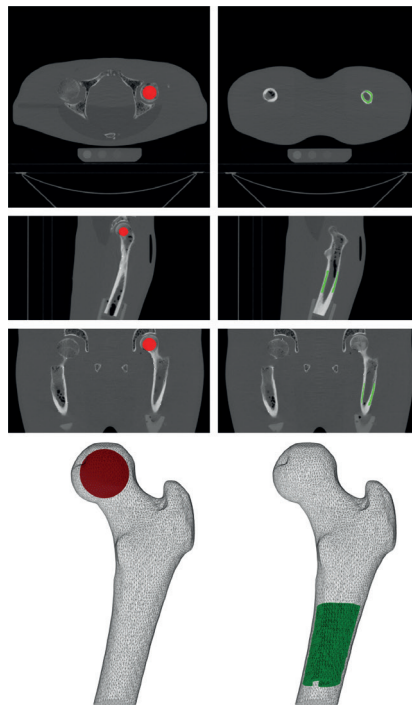
Since we are currently performing a multicenter patient study for *in vivo* validation of our FE models, we want to unravel these problems. As a first step, the effect of different CT scanners and CT protocols on HU and BMD using tissue characterization phantoms was recently investigated.<sup>11</sup> We found differences between CT scanners in HU in bone-equivalent regions within the phantom up to 10%, and these differences decreased to maximally 7% when HU were calibrated to BMD via a calibration phantom under the tissue characterization phantom. Additionally, variations in CT settings, mainly reconstruction kernel, resulted in differences in bone-equivalent HU up to 16%. Also, air between calibration phantom and tissue characterization phantom affected the calibration. These effects were scanner-dependent.

The next step is to determine how differences in CT equipment or protocols affect HU and BMD in a more physiological setting, for example, when scanning femoral tissue. In that case, FE failure loads can be calculated as well. Femoral tissue is more heterogeneous than inserts in a phantom, and other beam hardening and partial volume effects can be expected under physiological circumstances. Therefore, the aims of this study were to quantify the effect of (i) different CT scanners; (ii) different CT protocols (with variations in slice thickness, field of view (FOV), and reconstruction kernel); and (iii) air between calibration phantom and patient, on HU, BMD, and FE failure load.

## Methods

### Cadaveric Femora

Six fresh frozen femora (three male, three female; mean age 86.7 years, range 82–95 years) were obtained from the Anatomy department of the Radboud university medical center. Soft tissue was removed and the proximal femora were cut at 24 cm. All femora were embedded distally in polymethylmethacrylate according to previously described protocol.<sup>3</sup>



**Figure 1:** Axial, sagittal, and coronal slice and 3D plots of the trabecular (red) and cortical (green) ROIs.

### QCT Scanning

Four radiotherapy institutes participated in this study (Radboud university medical center Nijmegen, Leiden University Medical Center, Radiotherapeutisch Instituut Friesland Leeuwarden, Bernard Verbeeten Institute Tilburg). These institutes used CT scanners of three manufacturers (Philips Brilliance Big Bore (Philips Medical Systems, Eindhoven, The Netherlands, two institutes), GE Optima CT580 (GE Healthcare, Milwaukee, WI, USA) and Toshiba Aquilion/LB (Toshiba Medical Systems, Tokyo, Japan), abbreviated with P1, P2, GE, and To, respectively).

The femora were placed in an anatomically shaped model of the lower body,<sup>16</sup> mimicking the human body. This body model was positioned in the isocenter of the CT scanner atop a solid calibration phantom (Image Analysis, Columbia, KY), containing four known calcium hydroxyapatite concentrations (0, 50, 100, and 200 mg/cm<sup>3</sup> CaHA). The known densities in this phantom were used to calibrate HU to CaHA density, which is a measure of BMD.

Standard scans were acquired according to the standard patient study protocol, using the following settings: 120 kV, variable mA (calculated by the scanner software), 1 s rotation time, 3 mm slice thickness, FOV 480 mm, in plane resolution 0.9375 mm, standard reconstruction kernel (B on P1 and P2, standard on GE, and FC17 on To), pitch <1. To study the effect of different CT settings, the femora were scanned using the standard protocol, and with variations in slice thickness (1 mm), FOV (550 mm, in plane resolution 1.0742 mm), and reconstruction kernel (detailed: D on P1, UB on P2, detail on GE, and FC43 on To),<sup>11</sup> which were the most commonly applied deviations from the standard protocol in our patient study. All combinations were applied, resulting in a total of eight scans with different settings per CT scanner. Subsequently, the effect of an air gap between calibration phantom and lower body model was assessed by lifting the knees of the lower body model 5 and 10 cm, respectively. The latter scans were acquired on every CT scanner using the standard protocol, and mimicked a patient's knees being supported by a cushion.

### Cortical and Trabecular ROI

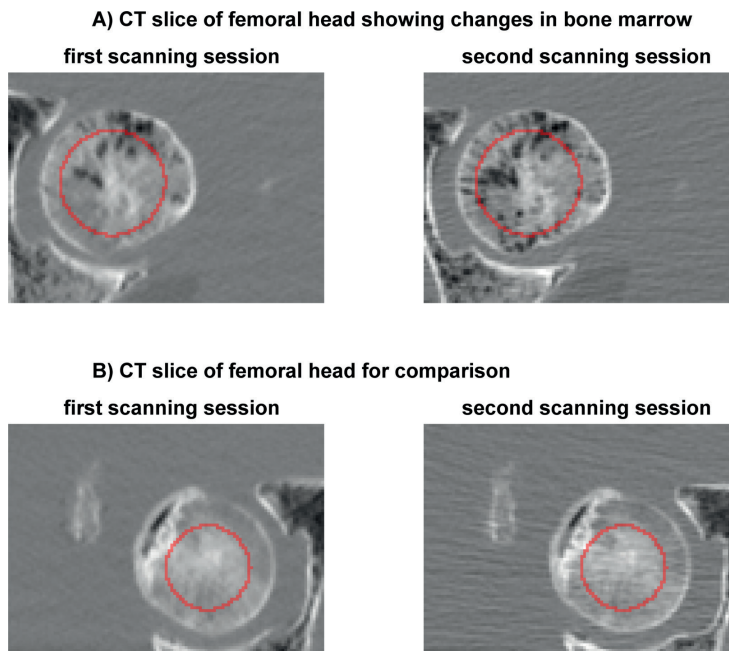
Subject-specific femoral geometry was obtained from the standard CT scans (3 mm, FOV 480, standard reconstruction kernel) by selecting the voxels containing femoral tissue in each slice (Mimics 14.0, Materialise, Leuven, Belgium). For all other scans, the femur geometry was registered to the CT scan using software containing algorithms for registration of medical images (elastix<sup>17,18</sup>). Subsequently, a cortical and a trabecular region of interest (ROI) were drawn (Figure 1). For the cortical ROI, 10 cm below the femoral head, voxels were selected along 6 cm of the cortex of the femoral shaft. For the trabecular ROI, 75% of the sphere that fitted the femoral head was used. ROIs were registered using the transformation of the femoral registration. Mean HU were obtained for each ROI and were calibrated to mean BMD.

In one of the femora, for unknown reasons, bone (marrow) composition in the femoral head seemed to have changed after the first scanning session (P1), when comparing it with scans of the next scanning sessions (P2, GE, and To, Figure 2). Therefore, this femur was excluded from analysis of the trabecular ROI. The cortical ROI was not affected. The femur was not excluded from FE analysis (see

next paragraph), as in most part of the femoral head post-yield material behavior was not implemented in the FE models, and hence, we expect the FE failure load not to be significantly affected.

### FE Models

FE models based on the standard CT scan of every femur were constructed as described previously.<sup>3</sup> The femoral geometry was converted into a solid mesh (average element volume  $1.4 \text{ mm}^3$ , Patran 2011, MSC Software Corporation, Santa Ana, CA). Using the calibration phantom, the HU of each element were calibrated to BMD. The BMD values were subsequently converted to non-linear isotropic bone material properties.<sup>19</sup> Two bundles of high-stiffness springs served as distal fixation and a displacement-driven load was applied on the femoral head via a simulated cup. To prevent artifacts as a result of the loading configuration, post-yield material behavior was not implemented in a region underneath the cup comprising the proximal elements of two third of the femoral head, and at the distal fixation comprising distal elements of a region as high as the radius of the shaft. FE simulations were performed using MSC.MARC (v2013.1, MSC Software Corporation). Incremental displacement and contact normal forces were registered and plotted in force-displacement curves. It was assumed that fracture occurred when maximum total reaction force was reached. At the corresponding increment, a clear fracture line of plastic elements was visible.



**Figure 2:** One femur seemed to have lost bone marrow in the femoral head after the first compared to the second scanning session (A). For comparison, we show another femur after the first and the second scanning session (B). The red line depicts the edge of the trabecular ROIs.

The femur geometry from the standard scan was registered onto the non-standard scans in order to obtain CT scan-specific material properties. All other aspects of the FE model, for example the geometry and alignment, were left unchanged. In this way, only the material properties differed between FE models, which enabled us to study the isolated effect of variations in CT images. This resulted in a total of 240 FE simulations (10 scans of six femora on four scanners, hence 240 calculated failure loads).

### Statistical Analysis

Linear mixed models were used to analyze effects of different CT scanners on HU and BMD in the cortical and trabecular ROI, and on FE calculated failure load. Slice thickness, FOV, and reconstruction kernel were added to the model as fixed factors to cover the effect of changes in CT settings. A random intercept was included to disregard the variability between femora. Only the interaction between CT scanners and reconstruction kernel was added to the model, as this increased the models' fits based on likelihood-ratio tests. All other interactions did not increase the fits and were therefore omitted from the final models. This includes the interactions between the effects of changing slice thickness, FOV or reconstruction kernel, indicating that there was no additional effect when two or more CT settings were changed.

Furthermore, linear mixed models were created to determine the effect of air between body model and calibration phantom on the outcome variables. As random factors, CT scanner and air gap (0, 5 or 10 cm) were added to the model. Again, a random intercept was included to account for the variability between the femora. For HU and failure load, no interactions were modeled, as they did not significantly improve the fit of the model without interaction based on likelihood-ratio tests. However, the interaction between CT scanner and air gap did significantly affect the fit of cortical and trabecular BMD, and was therefore included in these two models. *P*-values below 0.05 were considered statistically significant. All statistical analyses were performed using Stata/SE 11.2 (StataCorp LP, College Station, TX).

As descriptive statistics, median and range of cortical and trabecular HU and BMD and failure loads of the standard scans were calculated. These standard scans served as a default scan. To quantify inter-scanner effects, results of standard scans of P1, P2, GE, and To were expressed relative to the average of these four standard scans. The effect of variations in CT settings and air gap was quantified by normalizing the results of non-standard scans to the scanner-specific standard scans. Results are expressed as mean percentage  $\pm$  SD.

Results

Effect of CT Scanners

Medians and ranges of the standard scans for all outcome measures are depicted in Table 1. The differences in cortical HU between all CT scanners were significant (Table S-2). Scanning on GE derived the lowest cortical HU, while scanning on Toshiba resulted in highest cortical HU, leading to a difference of on average  $7 \pm 2\%$  (Figure 3). In the trabecular ROI, lowest HU were found when scanning on P2. On the GE scanner, HU in the trabecular ROI were largest, resulting in a maximal difference between GE and P2 of  $5 \pm 4\%$ .

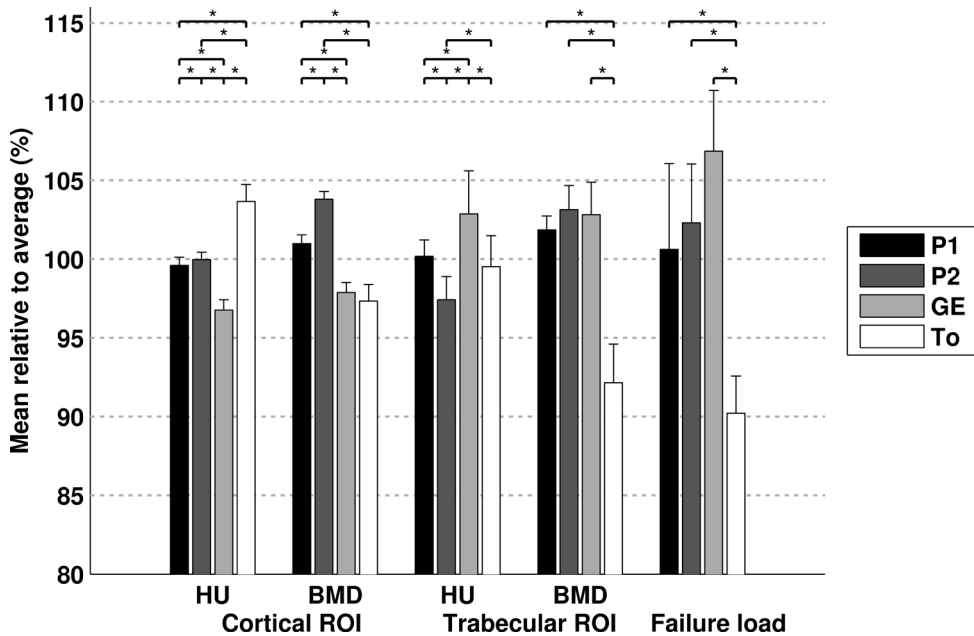
Calibrating HU to BMD changed the differences between CT scanners (Table S-2). In the cortical ROI, the highest BMD was found on P2, while on To BMD was the lowest, leading to a maximal variation of  $6 \pm 1\%$  (Figure 3). In the trabecular ROI, again P2 and To differed the most ( $11 \pm 4\%$ ).

FE calculated failure load was not significantly different between P1, P2, and GE, while all differences with respect to To were significant (Table S-2). Although not significantly different from P1 and P2, scanning on GE resulted in the highest calculated failure load, while the lowest failure load was calculated with scans of To. The maximal difference between GE and To in calculated failure load was  $17 \pm 5\%$  (Figure 3).

**Table 1:** Median of the standard scans (3 mm Slices, FOV 480, Standard Kernel) of all Femurs for all outcome measures

	P1	P2	GE	To
<b>Cortical ROI</b>				
<b>HU</b>	1150 (1043-1290)	1155 (1032-1289)	1118 (1004-1253)	1195 (1083-1341)
<b>BMD</b> <b>(mg/cm³)</b>	1009 (908-1133)	1036 (927-1154)	978 (874-1097)	971 (876-1089)
<b>Trabecular ROI</b>				
<b>HU</b>	278 (103-318)	267 (99-311)	280 (108-315)	278 (98-317)
<b>BMD</b> <b>(mg/cm³)</b>	245 (88-276)	241 (91-281)	242 (91-272)	222 (76-254)
<b>Failure load (N)</b>	4531 (1272-7152)	4655 (1340-7500)	4779 (1499-7224)	4179 (1218-6429)

In between brackets, the range of the outcomes among the different femurs is displayed. For the outcomes of the standard scan on each CT scanner for each femur separately, see Supplementary Table S-1.



**Figure 3:** Output of the scanners using the standard protocol (3 mm slices, FOV 480, standard kernel), in % relative to the average of all scanners (mean $\pm$ SD) for HU and BMD in the cortical and trabecular ROI, and simulated failure load. \*significant difference.

## Effect of CT Settings

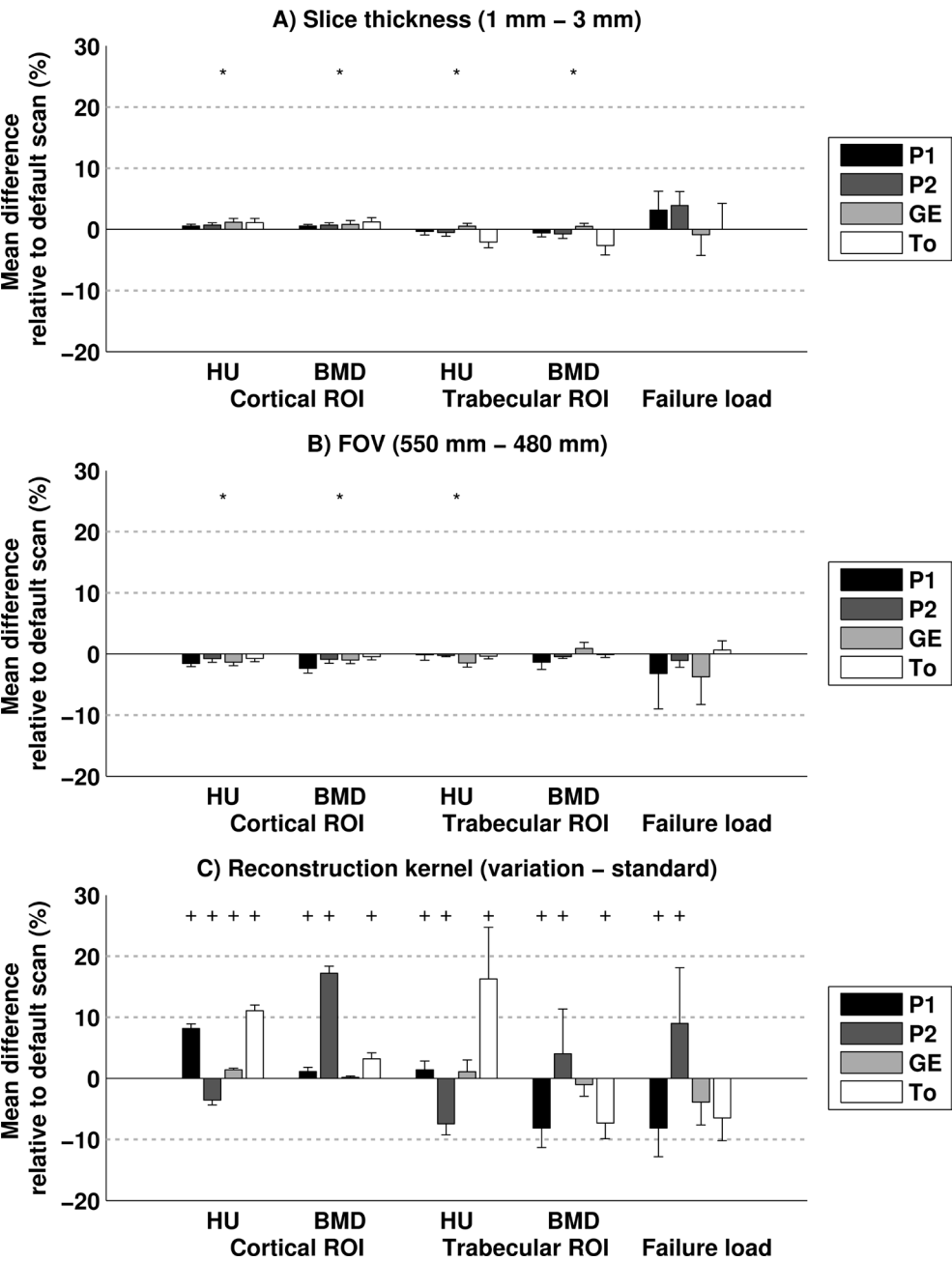
### Slice Thickness

The effect of changing the standard 3 mm slice thickness to 1 mm was small (Figure 4A, Table S-3). Changing slice thickness resulted in maximal  $1 \pm 1\%$  difference in cortical HU,  $2 \pm 1\%$  in trabecular HU,  $1 \pm 1\%$  in cortical BMD, and  $3 \pm 2\%$  in trabecular BMD. The effect of changing the slice thickness on calculated failure load was largest on the P2 scanner, with a  $4 \pm 2\%$  increase in failure load.

### FOV

The effects of varying FOV between 480 and 550 mm were small (Figure 4B, Table S-3). The largest effect of variations in FOV on HU was  $2 \pm 0\%$  on P1 for the cortical ROI and  $1 \pm 1\%$  on GE for the trabecular ROI. After calibration, the effects were largest on P1:  $2 \pm 1\%$  for the cortical ROI and  $1 \pm 1\%$  for the trabecular ROI. The effect of FOV on failure load was non-significant, but was largest in GE with on average  $4 \pm 5\%$  change in failure load.





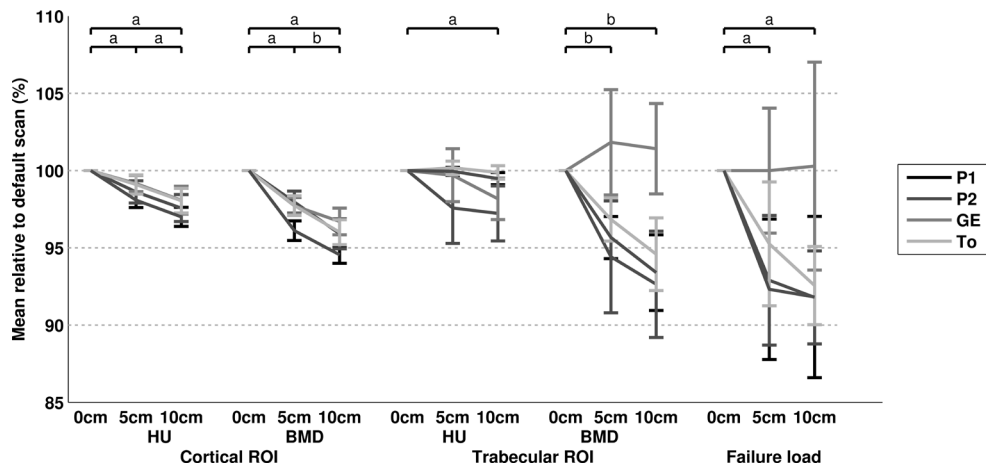
**Figure 4:** Difference between standard scan (3mm slices, FOV 480, standard kernel) and variation as percentage of the standard scan (mean±SD) of variations in slice thickness (A), FOV (B), and reconstruction kernel (C) for HU and BMD in the cortical and trabecular ROI, and simulated failure load. \* significant effect, holds for all of the CT scanners, as the interaction between CT scanner and slice thickness or FOV was not in the statistical model. +, significant effect.

### Reconstruction Kernel

The effect of changing the standard reconstruction kernel to a detailed reconstruction kernel was larger than the effects of variations in slice thickness or FOV on the P1, P2, and To scanners (Figure 4C, Table S-3). The effect of changing the reconstruction kernel on cortical HU was largest on To, resulting in an average increase of  $11 \pm 1\%$ . The largest effect on trabecular HU was  $16 \pm 8\%$  (increase) on To. When the HU were calibrated to BMD, the effects of reconstruction kernel changed, but did not disappear. The effect was largest on P2 ( $17 \pm 1\%$  increase) in the cortical ROI and on P1 ( $8 \pm 3\%$  decrease) in the trabecular ROI. For calculated failure load, the effect of changes in reconstruction kernel were significant on P2 with an average increase of  $9 \pm 9\%$ , and P1 with an average decrease of  $8 \pm 5\%$ . The effects of reconstruction kernel were not significant on GE and To.

### Effect of an Air Gap Between Body Model and Calibration Phantom

In the cortical ROI, the HU decreased significantly when there was an air gap between calibration phantom and body model (max  $3 \pm 1\%$  on P1; Figure 5, Table S-4). The effect of a 5 cm air gap was not significant in the trabecular ROI (max  $2 \pm 2\%$  on P2), while a 10 cm air gap resulted in a significant decrease in HU (max  $3 \pm 2\%$  on P2). The air gap resulted in a decreased cortical BMD on all scanners (max  $5 \pm 1\%$  on P1), and a decreased trabecular BMD on the P1, P2, and To scanner (max  $7 \pm 3\%$  on P2). Also, an air gap resulted in a decrease in failure loads (max  $8 \pm 3\%$  on P2).



**Figure 5:** Effect of an air gap (0, 5, and 10cm) between calibration phantom and body model as percentage (mean $\pm$ SD) relative to the standard scan (0cm air gap, 3mm slices, FOV 480, standard kernel) for cortical HU and BMD, trabecular HU and BMD, and simulated failure load. a, significant difference on all CT scanners. b, significant difference on P1, P2, and To.

## Discussion

This study aimed to quantify the effect of (i) different CT scanners; (ii) different CT protocols (with variations in slice thickness, FOV, and reconstruction kernel); and (iii) air between calibration phantom and patient on HU, BMD, and FE failure load.

We confirmed that differences between scanners in HU, BMD, and calculated failure loads can exist, even when a standard CT protocol is used and scanners are regularly calibrated according to manufacturer's specifications. A recent study by our group, using tissue characterization phantoms, showed that differences between scanners decreased after calibration using a calibration phantom to BMD.<sup>11</sup> The current study evaluated the inter-scanner effects using cadaveric femora; however, now, the use of a calibration phantom did not always correct for inter-scanner differences. Hence, using CT images from various scanners resulted in differences in BMD and subsequent failure loads. The study of Carpenter *et al.*<sup>13</sup> determined inter-scanner differences with the use of femora, and determined BMD in healthy subjects based on QCT images. They reported differences up to ~20% in cortical and ~40% in trabecular BMD, when patients were scanned on two different CT scanners. In their case, this led to an average of ~12% and maximally ~23% in subsequently calculated failure loads with the use of FE models that simulated single-leg stance. Percentage-wise, most of our inter-scanner differences were smaller than those of Carpenter *et al.*,<sup>13</sup> while the differences in both trabecular BMD and failure load between the Toshiba CT scanner and the other scanners were similar to their findings.

The effect of variations in slice thickness and FOV was small (<4% on average), whereas the effect of reconstruction kernel was larger (average of 16% in HU, 17% in BMD, and 9% in failure load). The current HU results were in correspondence with our previous phantom study, also showing small effects of changes in slice thickness and FOV, while reconstruction kernel did affect HU considerably.<sup>11</sup> In general, the variation in failure load between the femora was larger compared to the variations in HU and BMD. Possibly, this is due to the exponential functions to calculate material properties.<sup>19</sup> Additionally, the calculated failure load is the result of many numeric calculations that may increase the effect of small deviations in input data. Nevertheless, the exact algorithms behind different reconstruction kernels that calculate the HU from the X-ray projection data remain unknown. It is, therefore, hard to predict in what way the CT images will be affected by a certain kernel. On different CT scanners, different reconstruction kernels were chosen and each kernel had different effects on the outcomes. In most cases, the calibration phantom was not able to correct the effect of changes in reconstruction kernel. For example, on P1 relatively high effects on HU in the cortical ROI were smaller after calibration to BMD, but the small effect on the trabecular HU increased after calibration on the same scanner. The Toshiba scanner was an exception: the effect of reconstruction kernel on HU was quite large, but was much smaller after calibration to BMD, and as a result, the kernel had a relatively small effect on the calculated failure load. Other studies also found that different reconstruction kernels lead to differences in HU,<sup>9,14</sup> BMD,<sup>14</sup> and FE calculated vertebral stiffness.<sup>20</sup> Calibration with the use of a calibration phantom did not decrease this effect.<sup>14</sup>

As a third aim, we investigated the effect of an air gap between body model and calibration phantom. HU in the cortical and trabecular ROI decreased when the knees of the body model were lifted from the calibration phantom ( $\leq 3\%$  on average), which is probably due to changes of the position of the femur in the scanner gantry.<sup>8,11,21</sup> Since in most cases the effect of the air gap was larger after calibration ( $\leq 5\%$  on average), we assumed that the calibration phantom was somewhat affected by the air-tissue transition, automatically resulting in decreased simulated failure load on all CT scanners ( $\leq 8\%$  on average) except for GE. However, to what extent the simulations were affected by the air artifact remains unclear, as the change in position in the gantry also played a role. In our previous patient study,<sup>7</sup> we noted that air gaps induced a visibly larger artifact in the calibration phantom compared to the current results. Additionally, this artifact was only seen in scans made on a relatively old CT scanner (AcQSim CT, Philips Medical Systems, Eindhoven, The Netherlands). Possibly, newer scanners can better handle the tissue-air-phantom transition, although air gaps should be avoided if possible. Our study had some limitations. First of all, femora were thawed and refrozen multiple times, which might have led to some bone tissue damage. Possibly, this caused the change in bone (marrow) composition in the trabecular bone in one femoral head over time, leading to exclusion of this femur from the analysis of the trabecular ROI. Secondly, the position of the femora was not completely identical in each scanner. Femoral head placement in the acetabulum was comparable, but anteversion angles could deviate between scans. Nevertheless, we chose to use the body model to better resemble an actual patient's CT scan.<sup>16</sup> Additionally, the body model position could vary somewhat between the CT scanners, despite careful position in the isocenter of each gantry. However, the position variations were very small and, therefore, we do not expect any significant effects of the placement of the femora and body model. We used an automated algorithm for registrations, which error is anticipated to be less than a single voxel.<sup>22,23</sup> In addition, the registration between a coarser and finer scan can be a source of additional variation.

Ultimately, we aim to correct for CT scanner or protocol related variations. Previously, Keyak *et al.*<sup>19</sup> assumed that differences between CT scanners and protocols or other varying parameters could be corrected when using a calibration phantom. Additionally, Giambini *et al.*<sup>20</sup> stated that the research community should come up with a standard clinical CT protocol as input for FE models. However, the present study showed that calibration did not always suffice, suggesting that even the use of a standard protocol could not fully correct for inter-scanner differences. Although we only applied the material behavior as described by Keyak *et al.*,<sup>19</sup> we expect that the effects of differences in CT scanner or CT protocol would be rather comparable when using other non-linear relationships. Therefore, it would be better to develop an effective method for correcting such differences, for example by comparing different kernels beforehand and choosing the most similar kernels between CT scanners for patient scans. In our patient study, we aim to differentiate between high and low fracture risk patients based on failure loads calculated by CT-based FE models. With respect to the fracture risk predictions, the differences between CT scanners and settings would be critical for patients that have a failure load around the threshold that distinguishes high from low fracture risk patients. In those cases, a patient would switch from a high-fracture risk prediction to a low-fracture risk prediction, or the other way

around, when scanned on another CT scanner or with another kernel. Based on the results from the current study, we suggest applying a CT scanner- and setting-dependent level of uncertainty to the failure loads of patients' femora. Subsequently, the patients can be categorized in three groups: high fracture risk, possible high fracture risk, and low fracture risk.

Within the process of creating an FE model, there are many other variables that can result in variations in failure load. Such uncertainties are unwanted when giving patient-specific advice. Although we now have investigated the effect of CT scanners and protocols, other factors, such as the effect of loading conditions or lytic versus blastic lesions, should be explored as well in the future.

In conclusion, this study showed that quantitative analysis of CT images acquired with different CT scanners could induce changes in HU, BMD, and calculated failure load up to 17%. When using different CT settings, changes in slice thickness and FOV had small effects ( $\leq 4\%$  on average), but reconstruction kernels induced variations up to on average 9% in failure load. Additionally, air between patient and calibration phantom slightly decreased the HU, BMD and failure loads ( $\leq 8\%$  on average), and should therefore, if possible, be avoided. Finally, for using FE modeling as a clinical tool to predict fracture risk, we suggest applying a CT scanner- and setting-dependent level of uncertainty to the femoral failure load of patients, and categorizing them in three groups: high fracture risk, possible high fracture risk, and low fracture risk.

## Acknowledgements

This project was funded by the Dutch Cancer Society (KUN 2012–5591). We thank Leon Driessen and Richard van Swam for their help with the body model; Robert Kaatee, Wim Jansen, Esther Raaijmakers, and all radiotherapy technologists for their help with the CT scanners; and Gerjon Hannink for his help with the statistical analysis.

## References

1. Van der Linden YM, Dijkstra PD, Kroon HM, Lok JJ, Noordijk EM, Leer JW, Marijnen CA. 2004. Comparative analysis of risk factors for pathological fracture with femoral metastases. *J Bone Joint Surg Br* 86:566-573.
2. Van der Linden YM, Kroon HM, Dijkstra SP, Lok JJ, Noordijk EM, Leer JW, Marijnen CA. 2003. Simple radiographic parameter predicts fracturing in metastatic femoral bone lesions: results from a randomised trial. *Radiother Oncol* 69:21-31.
3. Derikx LC, van Aken JB, Janssen D, Snyers A, van der Linden YM, Verdonschot N, Tanck E. 2012. The assessment of the risk of fracture in femora with metastatic lesions: comparing case-specific finite element analyses with predictions by clinical experts. *J Bone Joint Surg Br* 94:1135-1142.
4. Keyak JH, Kaneko TS, Rossi SA, Pejic MR, Tehranzadeh J, Skinner HB. 2005. Predicting the strength of femoral shafts with and without metastatic lesions. *Clin Orthop* 439:161-170.
5. Tanck E, van Aken JB, van der Linden YM, Schreuder HW, Binkowski M, Huizenga H, Verdonschot N. 2009. Pathological fracture prediction in patients with metastatic lesions can be improved with quantitative computed tomography based computer models. *Bone* 45:777-783.
6. Goodheart JR, Cleary RJ, Damron TA, Mann KA. 2015. Simulating activities of daily living with finite element analysis improves fracture prediction for patients with metastatic femoral lesions. *J Orthop Res* 33:1226-1234.
7. Eggermont F, Derikx LC, Verdonschot N, Van der Geest ICM, De Jong MAA, Snyers A, Van der Linden YM, Tanck E. 2018. Can patient-specific finite element models better predict fractures in metastatic bone disease than experienced clinicians? Towards introducing computational modelling into daily clinical practice. *Bone Joint Res* 7:430-439.
8. Levi C, Gray JE, McCullough EC, Hattery RR. 1982. The unreliability of CT numbers as absolute values. *AJR Am J Roentgenol* 139:443-447.
9. Birnbaum BA, Hindman N, Lee J, Babb JS. 2007. Multi-detector row CT attenuation measurements: assessment of intra- and interscanner variability with an anthropomorphic body CT phantom. *Radiology* 242:109-119.
10. Sande EP, Martinsen AC, Hole EO, Olerud HM. 2010. Interphantom and interscanner variations for Hounsfield units--establishment of reference values for HU in a commercial QA phantom. *Phys Med Biol* 55:5123-5135.
11. Free J, Eggermont F, Derikx L, Van Leeuwen R, Van der Linden Y, Jansen W, Raaijmakers E, Tanck E, Kaatee R. 2018. The effect of different CT scanners, scan parameters and scanning setup on Hounsfield units and calibrated bone density: a phantom study. *Biomedical Physics & Engineering Express* 4:055013.
12. Dragomir-Daescu D, Salas C, Uthamaraj S, Rossman T. 2015. Quantitative computed tomography-based finite element analysis predictions of femoral strength and stiffness depend on computed tomography settings. *J Biomech* 48:153-161.
13. Carpenter RD, Saeed I, Bonaretti S, Schreck C, Keyak JH, Streeper T, Harris TB, Lang TF. 2014. 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* 36:1225-1232.
14. Giambini H, Dragomir-Daescu D, Huddleston PM, Camp JJ, An KN, Nassr A. 2015. The Effect of

- Quantitative Computed Tomography Acquisition Protocols on Bone Mineral Density Estimation. *J Biomech Eng* 137.
15. Boden SD, Goodenough DJ, Stockham CD, Jacobs E, Dina T, Allman RM. 1989. Precise measurement of vertebral bone density using computed tomography without the use of an external reference phantom. *J Digit Imaging* 2:31-38.
  16. Tanck E, Deenen JC, Huisman HJ, Kooloos JG, Huizenga H, Verdonchot N. 2010. An anatomically shaped lower body model for CT scanning of cadaver femurs. *Phys Med Biol* 55:N57-62.
  17. Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. 2010. elastix: A Toolbox for Intensity-Based Medical Image Registration. *IEEE Trans Med Imaging* 29:196-205.
  18. Shamonin DP, Bron EE, Lelieveldt BPF, Smits M, Klein S, Staring M, In AsDN. 2014. Fast parallel image registration on CPU and GPU for diagnostic classification of Alzheimer's disease. *Frontiers in Neuroinformatics* 7.
  19. Keyak JH, Kaneko TS, Tehranzadeh J, Skinner HB. 2005. Predicting proximal femoral strength using structural engineering models. *Clin Orthop*:219-228.
  20. Giambini H, Dragomir-Daescu D, Nassr A, Yaszemski MJ, Zhao C. 2016. Quantitative Computed Tomography Protocols Affect Material Mapping and Quantitative Computed Tomography-Based Finite-Element Analysis Predicted Stiffness. *J Biomech Eng* 138.
  21. Hunter TB, Pond GD, Medina O. 1983. Dependence of substance CT number on scanning technique and position within scanner. *Comput Radiol* 7:199-203.
  22. Eggermont F, Derikx LC, Verdonchot N, Hannink G, Kaatee R, Tanck E, van der Linden YM. 2017. Limited short-term effect of palliative radiation therapy on quantitative computed tomography-derived bone mineral density in femora with metastases. *Adv Radiat Oncol* 2:53-61.
  23. Knoop TH, Derikx LC, Verdonchot N, Slump CH. 2015. A novel framework for the temporal analysis of bone mineral density in metastatic lesions using CT images of the femur. *SPIE Medical Imaging: International Society for Optics and Photonics*; pp. 94143A-94143A-94111.

## Supplementary material

**Table S-1:** Outcomes of the standard scan on each CT scanner for each femur separately. Note that the standard deviations (SD) of the HU are mainly caused by the anatomy of the bone structure within the ROI.

	Femur	P1	P2	GE	To	
Cortical ROI	HU	#1	1043 (328)	1032 (358)	1004 (352)	1083 (381)
		#2	1181 (238)	1197 (238)	1158 (234)	1234 (264)
		#3	1095 (125)	1107 (116)	1065 (135)	1162 (157)
		#4	1125 (234)	1131 (228)	1102 (223)	1144 (252)
		#5	1175 (237)	1180 (224)	1134 (247)	1228 (266)
		#6	1290 (114)	1289 (111)	1253 (112)	1341 (130)
	BMD (mg/cm³)	#1	908	927	874	876
		#2	1028	1075	1009	999
		#3	963	993	932	944
		#4	989	1015	964	930
		#5	1032	1056	992	997
		#6	1133	1154	1097	1089
Trabecular ROI	HU	#1	103 (97)	99 (103)	108 (100)	98 (103)
		#2	318 (139)	311 (152)	315 (141)	317 (148)
		#3	308 (182)	307 (190)	314 (173)	311 (182)
		#4	278 (139)	267 (166)	280 (154)	278 (148)
		#5	130 (131)	126 (139)	140 (130)	132 (130)
	BMD (mg/cm³)	#1	88	91	91	76
		#2	276	281	272	254
		#3	271	277	272	249
		#4	245	241	242	222
		#5	114	115	119	104
Failure load (N)	#1	2309	2517	2590	2096	
	#2	7152	7500	7224	6429	
	#3	4778	5170	5315	4543	
	#4	6824	6012	6677	5385	
	#5	1272	1340	1499	1218	
	#6	4283	4140	4244	3815	



**Table S-2:** Output of the statistical linear mixed models (difference, 95% confidence interval and p-value) of the absolute differences between the CT scanners using the standard protocol (3 mm slices, FOV 480, standard kernel), for HU and BMD in the cortical and trabecular ROI, and simulated failure load (N).

Cortical HU				
	Scanners	Difference	95% CI	p-value
	P1-P2	-9.1	-15.1 – -3.1	0.003
	P1-GE	28.5	22.6 – 34.5	<0.001
	P1-To	-54.1	-60.1 – -48.1	<0.001
	P2-GE	37.6	31.6 – 43.6	<0.001
	P2-To	-45.0	-51 – -39	<0.001
	GE-To	-82.6	-88.6 – -76.6	<0.001
Cortical BMD (mg/cm³)				
	Scanners	Difference	95% CI	p-value
	P1-P2	-35.1	-41.7 – -28.5	<0.001
	P1-GE	22.9	16.3 – 29.5	<0.001
	P1-To	24.2	17.6 – 30.7	<0.001
	P2-GE	58.0	51.4 – 64.6	<0.001
	P2-To	59.3	52.7 – 65.8	<0.001
	GE-To	1.3	-5.3 – 7.8	0.7
Trabecular HU				
	Scanners	Difference	95% CI	p-value
	P1-P2	5.9	3.7 – 8.1	<0.001
	P1-GE	-3.3	-5.5 – -1	0.004
	P1-To	2.2	0 – 4.4	0.05
	P2-GE	-9.2	-11.4 – -6.9	<0.001
	P2-To	-3.7	-5.9 – -1.5	0.001
	GE-To	5.5	3.2 – 7.7	<0.001
Trabecular BMD (mg/cm³)				
	Scanners	Difference	95% CI	p-value
	P1-P2	-3.1	-7.7 – 1.6	0.2
	P1-GE	-3.4	-8 – 1.3	0.2
	P1-To	17.6	13 – 22.2	<0.001
	P2-GE	-0.3	-4.9 – 4.3	0.9
	P2-To	20.7	16 – 25.3	<0.001
	GE-To	21.0	16.3 – 25.6	<0.001
Failure load (N)				
	Scanners	Difference	95% CI	p-value
	P1-P2	-82.7	-266 – 100.6	0.4
	P1-GE	-79.0	-262.4 – 104.3	0.4
	P1-To	497.0	313.6 – 680.3	<0.001
	P2-GE	3.7	-179.7 – 187	1.0
	P2-To	579.7	396.3 – 763	<0.001
	GE-To	576.0	392.7 – 759.3	<0.001

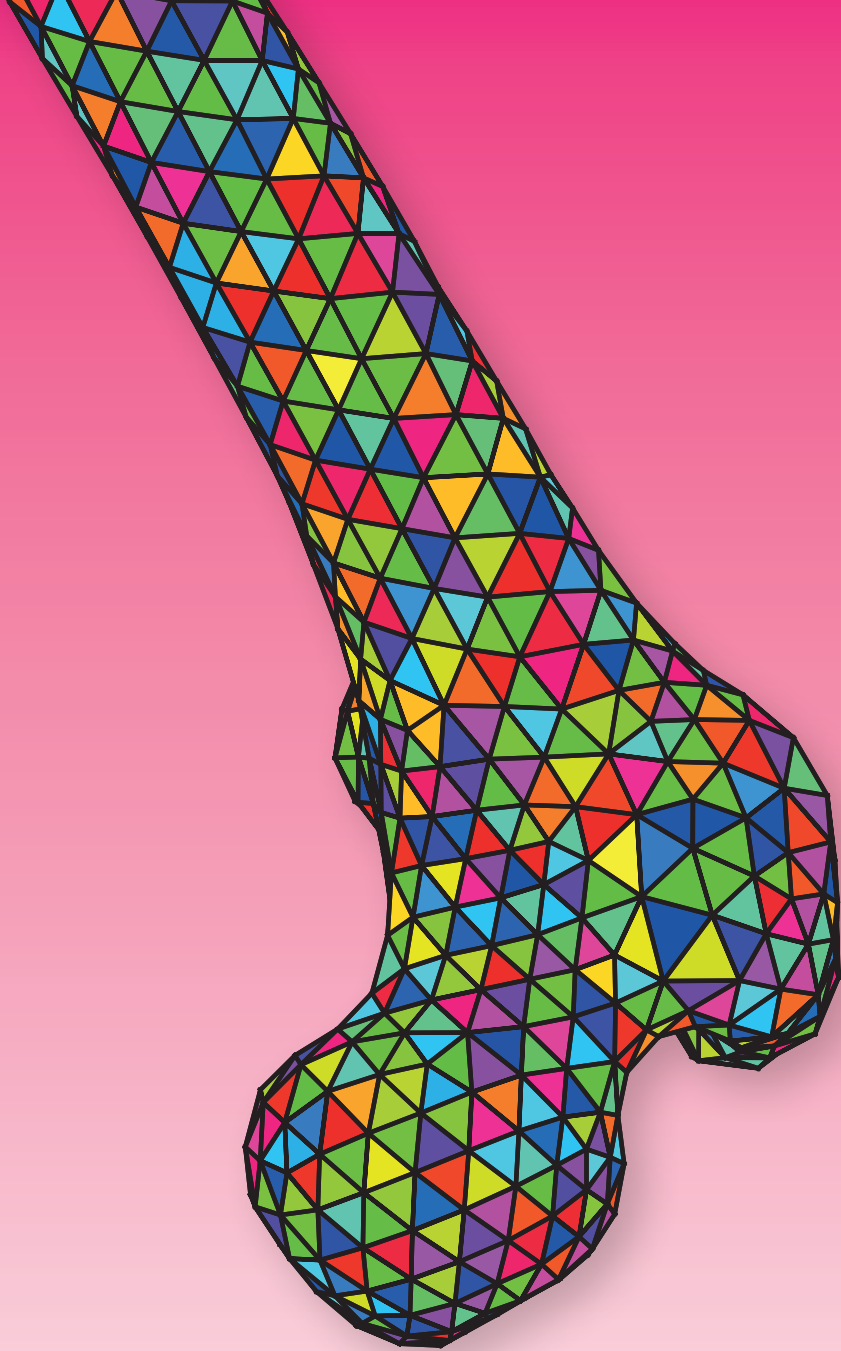
**Table S-3:** Output of the statistical linear mixed models (difference, 95% confidence interval and p-value) of the absolute effects of changes in Slice thickness, FOV and Reconstruction kernel, for HU and BMD in the cortical and trabecular ROI, and simulated failure load (N). \*The interaction between CT scanner and slice thickness or FOV was not significant and therefore, the effects of slice thickness and FOV hold for all of the CT scanners.

<b>Cortical HU</b>				
		Effect	95% CI	p-value
Slice thickness*		11.5	8.5 – 14.5	<0.001
FOV*		-10.8	-13.8 – -7.8	<0.001
Kernel	P1	102.9	96.9 – 108.9	<0.001
	P2	-39.7	-45.6 – -33.7	<0.001
	GE	15.6	9.6 – 21.6	<0.001
	To	133.2	127.2 – 139.2	<0.001
<b>Cortical BMD (mg/cm<sup>3</sup>)</b>				
		Effect	95% CI	p-value
Slice thickness*		11.8	8.5 – 15	<0.001
FOV*		-11.0	-14.3 – -7.7	<0.001
Kernel	P1	25.2	18.6 – 31.7	<0.001
	P2	183.2	176.6 – 189.8	<0.001
	GE	1.2	-5.4 – 7.7	0.7
	To	27.0	20.5 – 33.6	<0.001
<b>Trabecular HU</b>				
		Effect	95% CI	p-value
Slice thickness*		-1.2	-2.3 – -0.1	0.04
FOV*		-1.1	-2.2 – 0	0.05
Kernel	P1	2.5	0.2 – 4.7	0.03
	P2	-14.9	-17.1 – -12.6	<0.001
	GE	1.7	-0.6 – 3.9	0.1
	To	29.9	27.7 – 32.1	<0.001
<b>Trabecular BMD (mg/cm<sup>3</sup>)</b>				
		Effect	95% CI	p-value
Slice thickness*		-1.0	-3.3 – 1.3	0.4
FOV*		-0.7	-3 – 1.7	0.6
Kernel	P1	-14.5	-19.1 – -9.9	<0.001
	P2	14.6	10 – 19.2	<0.001
	GE	-2.5	-7.1 – 2.1	0.3
	To	-12.1	-16.7 – -7.5	<0.001
<b>Failure load (N)</b>				
		Effect	95% CI	p-value
Slice thickness*		100.5	8.9 – 192.2	0.03
FOV*		-61.8	-153.5 – 29.8	0.2
Kernel	P1	-312.0	-495.3 – -128.6	0.001
	P2	582.5	399.2 – 765.8	<0.001
	GE	-73.0	-256.4 – 110.3	0.4
	To	-179.8	-363.2 – 3.5	0.06

**Table S-4:** Output of the statistical linear mixed models (difference, 95% confidence interval and p-value) of the absolute effects of an air gap (0, 5, 10 cm) between calibration phantom and body model, for HU and BMD in the cortical and trabecular ROI, and simulated failure load (N). \*The interaction between CT scanner and air gap was not significant for cortical HU, trabecular HU and failure load and therefore, these effects of slice thickness and FOV hold for all of the CT scanners.

<b>Cortical HU</b>						
Scanner	0 cm - 5 cm Difference	95% CI	p-value	5 cm - 10 cm Difference	95% CI	p-value
All*	-14.2	-20.1 – -8.3	<0.001	-12.2	-18.1 – -6.3	<0.001
<b>Cortical BMD (mg/cm<sup>3</sup>)</b>						
Scanner	0 cm - 5 cm Difference	95% CI	p-value	5 cm - 10 cm Difference	95% CI	p-value
P1	-39.0	-49.3 – -28.7	<0.001	-15.7	-25.9 – -5.4	0.003
P2	-20.7	-31 – -10.5	<0.001	-21.1	-31.4 – -10.9	<0.001
GE	-21.9	-32.2 – -11.7	<0.001	-9.8	-20 – 0.5	0.06
To	-21.8	-32.1 – -11.5	<0.001	-16.7	-27 – -6.4	0.001
<b>Trabecular HU</b>						
Scanner	0 cm - 5 cm Difference	95% CI	p-value	5 cm - 10 cm Difference	95% CI	p-value
All*	-1.4	-3.2 – 0.4	0.1	-1.3	-3.1 – 0.5	0.2
<b>Trabecular BMD (mg/cm<sup>3</sup>)</b>						
Scanner	0 cm - 5 cm Difference	95% CI	p-value	5 cm - 10 cm Difference	95% CI	p-value
P1	-7.7	-12.1 – -3.3	0.001	-3.6	-8 – 0.7	0.1
P2	-9.3	-13.7 – -5	<0.001	-3.4	-7.8 – 0.9	0.1
GE	1.5	-2.9 – 5.9	0.5	-0.4	-4.8 – 4	0.9
To	-5.0	-9.3 – -0.6	0.03	-3.4	-7.8 – 1	0.1
<b>Failure load (N)</b>						
Scanner	0 cm - 5 cm Difference	95% CI	p-value	5 cm - 10 cm Difference	95% CI	p-value
All*	-231.3	-354.6 – -107.9	<0.001	-43.9	-167.3 – 79.5	0.5
<b>Failure load (N)</b>						
Scanner	0 cm - 5 cm Difference	95% CI	p-value	5 cm - 10 cm Difference	95% CI	p-value
All*	-231.3	-354.6 – -107.9	<0.001	-43.9	-167.3 – 79.5	0.5
<b>Failure load (N)</b>						
Scanner	0 cm - 5 cm Difference	95% CI	p-value	5 cm - 10 cm Difference	95% CI	p-value
All*	-231.3	-354.6 – -107.9	<0.001	-43.9	-167.3 – 79.5	0.5





# CHAPTER 7

**CALIBRATION WITH OR WITHOUT PHANTOM FOR  
FRACTURE RISK PREDICTION IN CANCER PATIENTS WITH  
FEMORAL BONE METASTASES USING CT-BASED FINITE  
ELEMENT MODELS**



---

Florieke Eggermont, Nico Verdonschot, Yvette van der Linden, Esther Tanck

PLoS ONE 2019; 14(7), e0220564.



## Introduction

Patients with advanced cancer and bone metastases have an increased risk of a pathological fracture. Occurrence of these fractures in the femur of the patient leads to immediate reduced mobility, pain and distress, and causes reduced quality of life. When patients present with a painful femoral metastasis, treatment plans are based on the fracture risk estimated by the clinical team: patients with a low fracture risk undergo conservative treatment such as radiotherapy, while patients with a high fracture risk are considered for prophylactic stabilization surgery to reduce the chance of fracturing.<sup>1,2</sup> In current clinical practice, fracture risk is estimated using CT scans and X-rays, but this appears to be difficult, leading to over and under treated patients.<sup>1</sup>

Finite element (FE) models have shown to be promising as a tool for fracture risk prediction.<sup>3-7</sup> Quantitative Computed Tomography (QCT) scans can be used to segment patient-specific bone geometries that function as input for the FE models. Additionally, Hounsfield units (HU) in the QCT scan can be converted to bone mineral densities (BMD) that are used to model element-specific bone material properties.<sup>3-7</sup> Currently, these conversions to BMD are usually done with either solid or liquid calibration phantoms that contain certain known concentrations of for example calcium hydroxyapatite ( $\text{CaCO}_3$  or  $\text{CaHA}$ ) or hydrogen dipotassium phosphate ( $\text{K}_2\text{HPO}_4$  or  $\text{KHP}$ ). These calibration phantoms are of reasonable size and need to be scanned along with the patient.

However, such separate calibration phantoms are not routinely available, quite expensive, and result in inability to use everyday clinical CT scans without a phantom for FE modeling. Since the FE patient databases are now dependent on CT scans including a calibration phantom made for scientific prospective studies, the databases are currently limited in size and clinical validation of these FE models moves slowly. If a phantomless calibration method is available to obtain BMD from HU, FE models can be generated retrospectively from clinical CT databases that lack calibration phantoms. In this manner, a large database could be built more easily for further validation of FE models for clinical fracture risk assessments in patients with femoral bone metastases. Additionally, a phantomless calibration method to use prospectively for each patient presenting with femoral bone metastases would be very helpful for clinical implementation of FE modeling as a fracture risk prediction tool.

Several methods have been developed for calibrating CT scans without a calibration phantom.<sup>8-15</sup> Some studies calibrate with the use of a calibration function obtained from a separate scan containing a calibration phantom,<sup>8</sup> determine calibration factors based on CT scans that contain a calibration phantom and apply this calibration factor to CT scans without calibration phantoms,<sup>9</sup> or calculate BMD using a regression model based on previous phantom calibration.<sup>15</sup> Another phantomless option is to use patient-specific internal calibration methods, which are based on HU of specific tissues, such as fat and muscle tissue<sup>10-13</sup> or external air and either aortic blood or visceral fat.<sup>14</sup> Studies comparing phantom calibration with phantomless calibrations showed that they yielded comparable results.<sup>8-15</sup> These studies used a single CT scanner or multiple CT scanners with a standardized protocol. Since it is known that changes in CT protocol can yield differences in HU,<sup>16-18</sup> one could expect an effect of CT scanner or protocol for certain phantomless calibration methods on BMD or FE outcomes.



Additionally, most studies only determined the effect of different calibration methods on vertebral trabecular BMD,<sup>8-12</sup> but not on FE outcomes. The before mentioned calibration methods function well for trabecular BMD determination, probably because they cover HU in the same range of trabecular bone.<sup>10-14</sup> However, when applying these calibration methods to cortical bone far outside the range of calibration values, this could lead to extrapolation errors. Since FE models of femurs contain both trabecular and cortical bone material properties, it should be determined whether comparable results are obtained when using phantomless calibrations. Although both calibration with and without a phantom are based on linear relationships between HU and BMD, the non-linear relationship between BMD and bone material properties<sup>19</sup> probably will affect the fracture risk as calculated by the FE models in a non-linear manner. A study by Lee *et al.* tested the effect of a phantomless calibration based on external air and visceral fat on femoral FE strength using research-quality clinical-resolution CT scans that had been analyzed in prior clinical drug trials. They found good correspondence between phantom and phantomless calibrations, with a mean difference between the calibrations of 30 N (0.8%).<sup>14</sup> Another study used a general calibration function to calibrate CT scans for femoral FE models and found errors of total strain energy of 0.91% in comparison with a phantom calibration.<sup>15</sup> Nevertheless, for FE modeling purposes, phantomless calibration methods have been studied limitedly. In addition, comparisons between several phantomless calibration methods for FE modeling have never been made.

The aim of this study was to develop a new calibration method that enables calibration of HU to BMD for finite element modeling of femurs with bone metastases without the use of a calibration phantom. The new phantomless calibration method should yield results similar to the phantom calibration. We evaluated two phantomless calibration methods: one based on HU of certain tissues within the CT scan and one non-patient-specific calibration function.

## Methods

### CT scans

Between August 2006 and September 2009,<sup>7,20</sup> and January 2015 and April 2017, patients with cancer and bone metastases in the femur that were treated with radiotherapy in one of four radiotherapy institutes in the Netherlands (Radboud university medical center, Nijmegen; Leiden University Medical Center, Leiden; Radiotherapeutic Institute Friesland, Leeuwarden; Bernard Verbeeten Institute, Tilburg) were asked to participate in a prospective cohort study that investigated if FE models were able to predict whether patients would or would not fracture their femur within six months. Hence, two cohorts were included both generated with an institutional review board approved research protocol (NL12568.099.06 and 2013/305). Institutes were instructed to generate QCT images of the patients using a standardized protocol, with the following settings: 120 kVp, 220 or variable mA, slice thickness 3 mm, pitch 1.5, spiral and standard reconstruction, field of view (FOV) 480 mm, in-plane resolution 0.9375 mm. In a few cases, the standardized protocol was accidentally violated, resulting in CT scans

with a different FOV or reconstruction kernel. Since patients were included from four institutes, evidently four different CT scanners were used. These CT scanners comprised two Philips Brilliance Big Bore (Philips-1 and Philips-2, Philips Medical Systems, Eindhoven, The Netherlands) scanners, one GE Optima CT580 (GE Healthcare, Milwaukee, WI) and one Toshiba Aquilion/LB (Toshiba Medical Systems, Tokyo, Japan).

Patients with predominantly blastic femoral metastasis were excluded from the current study, as in a previous study,<sup>7</sup> we found that the bone strength of such femurs was overestimated, probably due to unrealistically strong material properties in the FE model because of the high degree of mineralization in blastic lesions. Also, patients were excluded if they had a hip or knee prosthesis, the femur was incompletely scanned or the calibration phantom was not correctly placed, or body weight was absent from the clinical research files. This resulted in inclusion of 57 patients, with 67 femurs that were affected with bone metastases (Philips-1: 20 patients, 27 femurs; Philips-2: 16 patients, 18 femurs; GE: 8 patients, 8 femurs; Toshiba: 13 patients, 14 femurs).

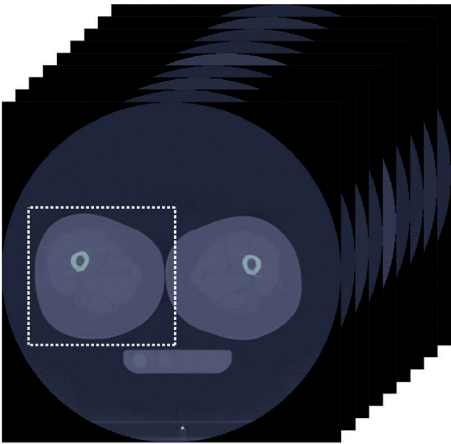
### Phantom calibration

The patients were scanned on top of a solid calibration phantom (Image Analysis, Columbia, KY), that contained four known CaHA concentrations (0, 50, 100 and 200 mg/cm<sup>3</sup>). The known densities in this phantom were used to calibrate HU to CaHA density, which is a measure of BMD. A mean diaphyseal calibration was applied by determining the HU in the four rods over nine diaphyseal slices and correlate them with the known CaHA concentrations of the calibration phantom.<sup>7</sup> The selection of the diaphyseal slices was protocolized by starting with the slice containing no buttox or genitals and selecting the 8 consecutive slices. To correct for inter-scanner differences,<sup>16,17,21</sup> the calibration curves were corrected toward the Philips-1 scanner, on which the FE model was validated,<sup>3,6,22</sup> by cross-calibration. For this, we used phantom scans (Gammex 467 phantom, RMI Gammex, Middleton, WI, USA)<sup>16</sup> to determine the linear correlation function between aberrant CT scanner and the Philips-1 scanner ( and , for Philips-2 and Toshiba, respectively. GE was not corrected). Next, we applied this function to the HU within the calibration phantom and subsequently determined the corrected calibration function. In a similar manner, the CT scans acquired with a different reconstruction kernel were corrected ( for Toshiba). CT scans with a different FOV were not corrected, as we have shown before that the effect on HU was negligible.<sup>17</sup>

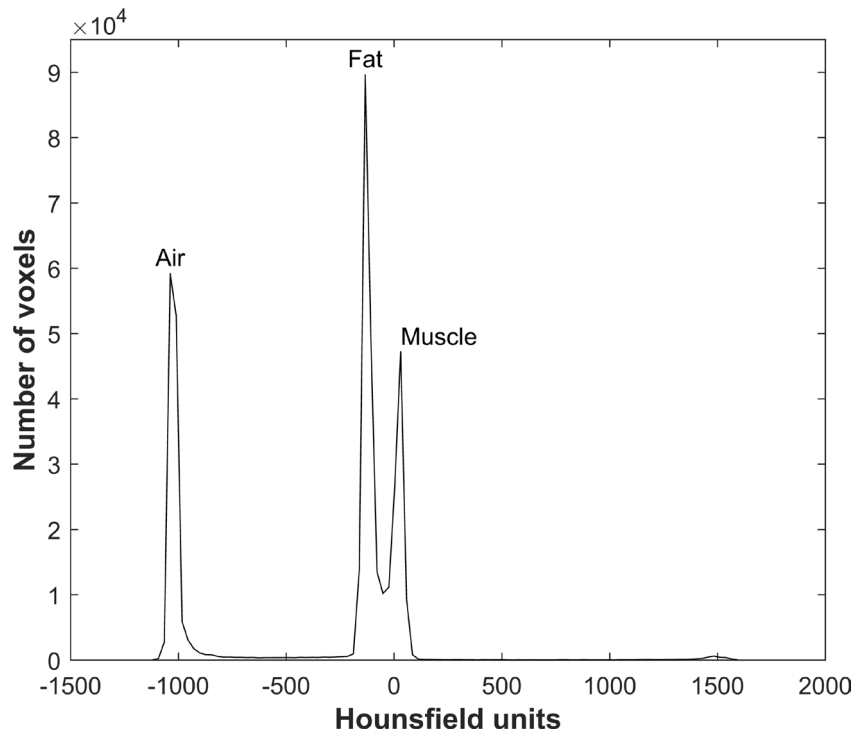
### Phantomless calibration

#### *Air-fat-muscle calibration*

We developed a phantomless calibration based on HU of certain tissues (air, fat, muscle and cortical tissue) within the CT scan. We first determined the accuracy of several calibrations with different combinations of air, fat, muscle and cortical tissue in a pilot study (see Supplementary Material). The combination of air, fat and muscle yielded the highest correlation with respect to the phantom calibration, and was therefore selected as the first phantomless calibration method.



**Figure 1:** An example of the region of interest (white dashed box) over nine diaphyseal slices that was used for the phantomless air-fat-muscle calibration



**Figure 2:** An example of a histogram of the Hounsfield units within the region of interest, used to extract the peaks for air, fat and muscle. An additional relatively small peak is visible around 1500 HU, indicating the cortical bone of the femur.

For the air-fat-muscle calibration, the same nine diaphyseal slices as used for the phantom calibration were selected, using the same protocol for slice selection. The succeeding steps of the air-fat-muscle calibration were completely automated. On the nine selected slices, a square region of interest was defined including the tissue of the right leg and some surrounding air ( $\pm 1$  cm on each side of the leg; Figure 1). Next, a combined histogram of all HU in the volume of interest (i.e. nine slices together) was created to extract the peaks for air, fat and muscle tissue (Figure 2). The mode of the HU around the histogram peak ( $\pm 50$  HU) was calculated, to determine the exact peak in HU for each of the tissues. By using the mode, the method was least susceptible to outliers. Subsequently, the determined HU peaks were linearly fitted to the reference “BMD” values for each patient to obtain the air-fat-muscle calibration function. These values were obtained by phantom calibrating the HU peaks of air, fat and muscle of a randomized subgroup comprising 10 patients scanned on the Philips-1 scanner. Subsequently, we averaged and rounded them, resulting in reference “BMD” values of -840, -80 and 30 for air, fat and muscle, respectively. The linear fits between HU and “BMD” were very good with an average  $R^2$  of  $1.000 \pm 0.000$  (slope =  $1.194 \pm 0.016$ , intercept =  $2.232 \pm 12.042$ ). No scanner- or kernel-specific correction was applied for the air-fat-muscle calibration method.

### ***Non-patient-specific calibration***

Additionally, we determined a non-patient-specific calibration function to convert HU to BMD by averaging all calibration functions of all 26 patients scanned on the Philips-1 scanner (6 patients were later excluded due to abovementioned reasons). We only used the 26 patients scanned on this particular CT scanner because of inter-scanner differences we found in previous studies.<sup>7,16,17</sup> The used patients were scanned on the CT scanner on which the FE model was validated.<sup>3,6,22</sup> This calibration function () was then applied to each of the 57 included patients. No scanner- or kernel-specific correction was applied for the non-patient-specific calibration method.

### **FE models**

FE models were created for each included femur according to a previously described protocol.<sup>3,7</sup> These FE models have previously been validated in an experimental setting, by comparing experimental bone strength of cadaveric femurs with simulated lytic lesion with predicted bone strength by FE models.<sup>3,6,22</sup> These FE models have also shown to be able to predict fracture risk in a clinical setting.<sup>7</sup> In summary, the femoral geometry was obtained from the CT scans (Mimics 14.0, Materialise, Leuven, Belgium) and was converted to a solid mesh (Patran 2011, MSC Software Corporation, Santa Ana, CA, USA). Non-linear isotropic bone material properties<sup>19</sup> were calculated sequentially in three ways from the BMD that were obtained with 1) use of the calibration phantom, 2) the phantomless air-fat-muscle calibration method and 3) the non-patient-specific calibration method. For accurate comparison, only the material properties were varied, whereas the other aspects of the FE model remained unchanged. The FE model was positioned in a stance configuration by aligning the femoral head center with the knee joint center. The model was distally fixated by two bundles of high-stiffness springs and via a cup on the femoral head a displacement-driven load was applied. MSC.MARC (v2013.1, MSC Software

Corporation, Santa Ana, CA, USA) was used for the FE simulations. We used Keyak's material model to describe the post-failure behaviour, starting with an initial perfectly plastic phase, followed by a strain softening phase and an indefinite perfectly plastic phase.<sup>19</sup> Incremental displacement and contact normal forces were registered and it was assumed that fracture occurred when maximum total reaction force was reached, which was defined as the failure load. In total, 201 FE models were made (67 femurs  $\times$  3 calibration methods).

### **Statistical analysis**

The phantom calibration method was used as gold standard, as the FE model was validated with the use of this calibration method.<sup>3,6,22</sup> The mean failure loads and standard deviations (SD) for each of the calibration methods were determined to enable interpretation of the results that follow from the statistical analysis. A linear mixed model was used to determine the differences in failure load between the different calibration methods. Femur was nested within CT scanner, because patients were only scanned on one CT scanner each. Femur, nested in CT scanner, was added as random intercept to disregard the variability between femurs and CT scanners.<sup>17</sup> In this way, the model analyzes the effects of calibration methods, instead of differences between femurs. Although we initially did not intend to investigate this, it was additionally tested whether adding protocol violations (other FOV or reconstruction kernel) as a fixed factor resulted in a better model fit based on likelihood-ratio tests. Adding FOV as a fixed factor did not significantly improve the likelihood-ratio and was therefore not added, whereas reconstruction kernel did and was added to the model as fixed factor. The interaction between calibration method and kernel was added to the model, since it was significant and increased the model's fit based on a likelihood-ratio test. The level of significance was defined at  $p < 0.05$ . Additionally, we made the comparisons between the different calibration methods visible by creating correlation plots and Bland-Altman plots.

## **Results**

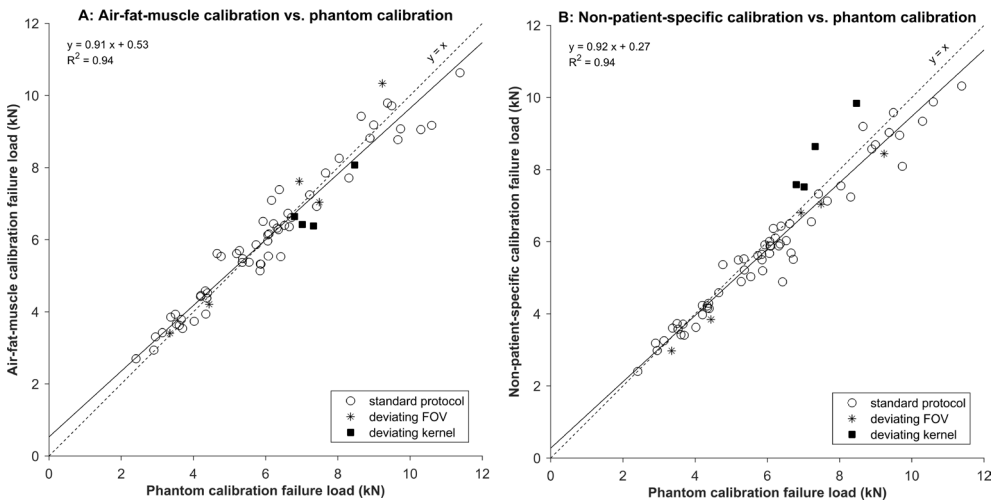
### **Patients and CT scans**

Sixty-seven femurs in 57 patients affected with bone metastases were included. Despite the protocolization of CT scanning, five femurs of five different patients scanned on two different CT scanners were scanned with an aberrant FOV (between 509 mm and 652 mm, instead of 480 mm). We found in a previous study that changes in FOV had little effect on failure loads.<sup>16,17</sup> Additionally, on one of the CT scanners, four femurs of four different patients were scanned with a different reconstruction kernel (detail kernel for scanning of the head, instead of the standard bone kernel). As mentioned before, CT scans acquired with a different kernel were corrected.

### Differences between calibrations

The mean failure loads were 6.17 kN (SD 2.11 kN), 6.16 kN (SD 1.99 kN) and 5.95 kN (SD 2.01 kN) for the phantom calibration, air-fat-muscle calibration and non-patient-specific calibration, respectively.

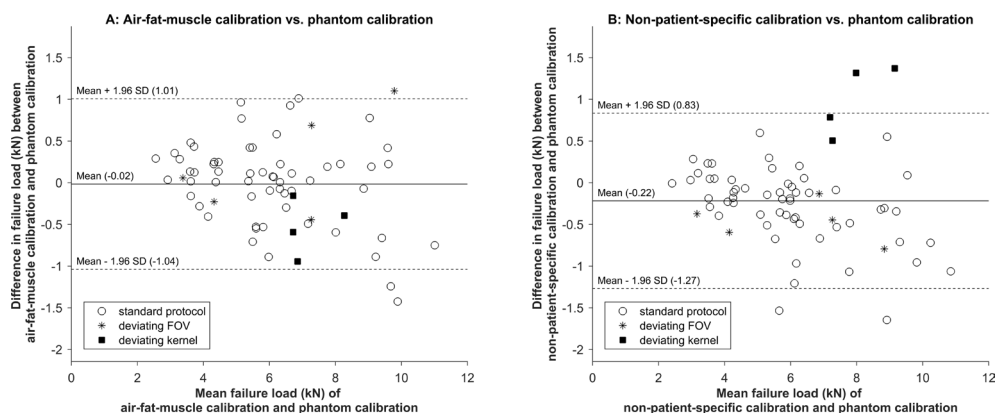
The correlations between the phantom calibration and other calibrations were very high ( $R^2 = 0.94$  for air-fat-muscle calibration and  $R^2 = 0.94$  for non-patient-specific calibration, Figure 3). There were no significant differences in FE failure loads based on the phantom calibration and air-fat-muscle calibration (0.02 kN, 95% confidence interval (CI) -0.10 – 0.13 kN,  $p = 0.8$ ), whereas the difference in failure loads between the phantom calibration and the non-patient-specific calibration was significant (-0.29 kN, 95% CI -0.41 – -0.18 kN,  $p < 0.001$ ). Similarly, the Bland-Altman plots showed a slightly higher agreement between phantom and air-fat-muscle calibration (-0.02, 95% CI -1.04 – 1.01) compared to the agreement between phantom and non-patient-specific calibration (-0.22, 95% CI -1.27 – 0.83, Figure 4).



**Figure 3:** Correlations between phantom and air-fat-muscle calibration (A) and between phantom and non-patient-specific calibration (B).

Changing reconstruction kernel had no significant effect on the phantom and air-fat-muscle calibration (1.31 kN, 95% CI -0.72 – 3.34 kN,  $p = 0.2$  and 0.77 kN, 95% CI -1.26 – 2.8 kN,  $p = 0.5$ , respectively), whereas changing reconstruction kernel resulted in significantly higher failure loads when using the non-patient-specific calibration (2.60 kN, 95% CI 0.57 – 4.63 kN,  $p = 0.01$ ).

Failure locations of the individual femurs were similar for the different calibration methods.



**Figure 4:** Bland-Altman plots for phantom versus air-fat-muscle calibration (A) and phantom versus non-patient-specific calibration (B).

## Discussion

Previously, we developed a patient-specific FE model that can be used in clinical practice to differentiate between high or low fracture risk in advanced cancer patients with femoral bone metastases.<sup>7</sup> This FE model has been based on CT scans that are calibrated using a separate calibration phantom that is scanned together with the patient. The aim of this study was to develop and evaluate two phantomless calibration methods for FE modeling of femurs of patients with cancer and bone metastases.

We developed the air-fat-muscle calibration and the non-patient-specific calibration. We found strong correlations between the phantom calibration and both phantomless calibration methods. In addition, there was no significant difference in failure load between the phantom calibration and air-fat-muscle calibration, whereas the difference between phantom and non-patient-specific calibration was significant. Although the difference between the phantom and phantomless calibrations may seem small, it can be critical for patients that have failure loads around the threshold distinguishing patients with a low fracture risk from patients with a high fracture risk. Additionally, although this study was not designed to investigate this, it should be mentioned that the non-patient-specific calibration worked well for protocolized CT scans, but seemed to have trouble correcting for changes in reconstruction kernel. In those cases, the air-fat-muscle calibration was more accurate. Changes in FOV did not lead to differences in accuracy of the calibration methods, probably because changes in FOV have little effect on HU.<sup>16,17</sup> When the same non-patient-specific calibration function is used for different CT protocols, this method lacks accuracy. This is inconvenient, since one would need to have a calibration function for each time a new CT scanner or protocol is being used. Therefore, a patient-specific calibration method, such as the air-fat-muscle calibration, would be more useful and robust.

Lee *et al.* tested a non-patient-specific phantomless calibration function on femoral FE models, and found it to be reliable as a replacement for phantom calibration.<sup>15</sup> However, they used one CT scanner and well protocolized CT scans, and therefore it has not yet been investigated whether non-patient-specific calibration methods are also useable for CT scans obtained on different CT scanners and with different settings. Additionally, another study investigated phantomless calibration for FE purposes and found comparable results between the phantom and phantomless calibrations.<sup>14</sup> They tested their phantomless air-fat method in 40 patients scanned on 24 different CT scanners, but only included scans that were made according to a standard scan protocol. Other studies used combinations of air and fat<sup>14</sup> or fat and muscle<sup>10-13</sup> for their phantomless calibrations, but none used the combination of air, fat and muscle, like we did. In a pilot study (see Supplementary Material), we first determined the accuracy of several calibrations with different combinations of air, fat, muscle and cortical tissue, but we found that the combination of air, fat and muscle yielded the highest correlation with respect to the phantom calibration. Addition of cortical tissue did not improve the correlation, which can be explained by the large variation in cortical density between patients.

In a number of other studies small ROIs were placed within certain tissues by hand to obtain the reference HU for the calibration,<sup>11-13</sup> which can be susceptible to inter-observer errors. We limited the variation between possible observers by automating all calibration methods. The only step that required manual input was the selection of the CT slices used for the calibrations. However, slice selection was strictly protocolized as well.

Each calibration method has its advantages and disadvantages. Calibration based on a calibration phantom that is scanned along with the patient has the benefit that the phantom is similarly affected as the FE modeled bone itself by patient- and scan-specific characteristics or artifacts, such as beam hardening, and might therefore be able to correct for such scan-specific characteristics or artifacts. The air-fat-muscle calibration method may be even better in capturing such potential artifacts, since the tissues on which the calibration is based are closer to the FE modeled bone compared to a calibration phantom. The non-patient-specific calibration cannot correct for any patient-specific CT artifacts.

Another important advantage of a calibration phantom is that the calcium densities in this phantom are precisely known. However, these usually only comprise low calcium densities, since high densities in the calibration phantom can lead to unnecessary artifacts in patient CT scans. As a result, the low calcium densities in the calibration phantom have to be extrapolated to higher (cortical) bone-equivalent densities, which is susceptible to errors. That also applies to the air-fat-muscle calibration. Furthermore, we have seen that calibration phantoms can be affected by shadow artifacts caused by air gaps between patient and calibration phantom, leading to errors in the calibration.<sup>7,16,17</sup> Such shadow artifacts are not relevant for the air-fat-muscle calibration. However, possible patient-specific variations in fat or muscle composition can affect the air-fat-muscle calibration, mainly when a patient is suffering from pathologies that are known to affect the attenuation values on CT.<sup>23</sup> We assume that variations will usually be small, since we used the mode instead of mean or median of the HU peaks. Taking air as reference should be without problems, as the radiodensity of air is also used for calibration of attenuation coefficients to HU. The main disadvantage of the patient-specific calibration



involves the fact that HU can vary between different CT scanners,<sup>16,17,24-26</sup> and therefore it might be better to use scanner-specific calibration functions, as mentioned before. However, this requires generating new calibration functions each time a new scanner is used, which is quite labor-intensive. Moreover, the major downsides of using separate calibration phantoms in clinical practice are the fact that patients have to lie on top of a separate mattress, and for the departments the expensive price and the logistical challenges it brings to scan each patient using the phantom. Additionally, when using the phantom calibration, there is a need for a scanner- and kernel-specific correction, for which extra CT scans of a tissue characterizing phantom have to be made.<sup>16,17</sup> This requires CT scanning and analyzing of many extra CT scans, mainly when one would want corrections for all different reconstruction kernels, which is hardly workable. Both the air-fat-muscle as the non-patient-specific calibration do not require additional logistics or costs. Also, the air-fat-muscle calibration seems to be less affected by changes in CT scanner or protocols, possibly because it uses reference tissues that are closer to the FE modeled bone in the isocenter. It is known that CT scans are affected by scanner and settings in a non-uniform manner, with a different effect near the isocenter of the FOV in comparison to the edges of the FOV, where the calibration phantom is placed. On the contrary, using the same non-patient-specific calibration function on all CT scanners and for all CT protocols, it will not be able to supply any form of correction. As a result, air-fat-muscle calibration seemed to be preferable over non-patient-specific calibration, as the air-fat-muscle calibration seemed to be better in handling deviations to the scan protocol comparable to the phantom calibration.

It should be mentioned that there was no gold standard while evaluating the different calibration methods. As the FE model has been validated while making use of a phantom calibration,<sup>3</sup> we chose this method to be the gold standard. This validation was done by creating FE models of cadaveric femurs, which were experimentally loaded until failure, and correlating the predicted failure load with the experimental failure load.<sup>3</sup> Although it is impossible to achieve this, it would be better to know the real failure loads of the patients' femurs. Then it would be possible to investigate which of all calibration methods is the best in approaching the true failure loads.

In conclusion, phantomless calibration of CT scans using the air-fat-muscle calibration method is preferable over the non-patient-specific calibration method. The phantomless calibration method will stimulate the prospective use of the FE model as a fracture risk prediction tool for each patient that presents with femoral bone metastases with clinical implementation as ultimate goal. Additionally, with the use of the phantomless calibration method, FE models of retrospective CT scans without calibration phantoms can be generated and a large database can be built that can be used for the validation of FE models for application to fracture risk assessment in patients with femoral bone metastases.

## Acknowledgements

This work was supported by the Dutch Cancer Society (KUN 2012-5591), the Dutch Science Foundation NWO-STW (NPG.06778), and Fonds NutsOhra (1102-071).

## References

1. Van der Linden YM, Dijkstra PD, Kroon HM, Lok JJ, Noordijk EM, Leer JW, Marijnen CA. 2004. Comparative analysis of risk factors for pathological fracture with femoral metastases. *J Bone Joint Surg Br* 86:566-573.
2. Van der Linden YM, Kroon HM, Dijkstra SP, Lok JJ, Noordijk EM, Leer JW, Marijnen CA. 2003. Simple radiographic parameter predicts fracturing in metastatic femoral bone lesions: results from a randomised trial. *Radiother Oncol* 69:21-31.
3. Derikx LC, van Aken JB, Janssen D, Snyers A, van der Linden YM, Verdonchot N, Tanck E. 2012. The assessment of the risk of fracture in femora with metastatic lesions: comparing case-specific finite element analyses with predictions by clinical experts. *J Bone Joint Surg Br* 94:1135-1142.
4. Goodheart JR, Cleary RJ, Damron TA, Mann KA. 2015. Simulating activities of daily living with finite element analysis improves fracture prediction for patients with metastatic femoral lesions. *J Orthop Res* 33:1226-1234.
5. Keyak JH, Kaneko TS, Rossi SA, Pejic MR, Tehranzadeh J, Skinner HB. 2005. Predicting the strength of femoral shafts with and without metastatic lesions. *Clin Orthop* 439:161-170.
6. Tanck E, van Aken JB, van der Linden YM, Schreuder HW, Binkowski M, Huizenga H, Verdonchot N. 2009. Pathological fracture prediction in patients with metastatic lesions can be improved with quantitative computed tomography based computer models. *Bone* 45:777-783.
7. Eggermont F, Derikx LC, Verdonchot N, Van der Geest ICM, De Jong MAA, Snyers A, Van der Linden YM, Tanck E. 2018. Can patient-specific finite element models better predict fractures in metastatic bone disease than experienced clinicians? Towards introducing computational modelling into daily clinical practice. *Bone Joint Res* 7:430-439.
8. Habashy AH, Yan X, Brown JK, Xiong X, Kaste SC. 2011. Estimation of bone mineral density in children from diagnostic CT images: a comparison of methods with and without an internal calibration standard. *Bone* 48:1087-1094.
9. Budoff MJ, Malpeso JM, Zeb I, Gao YL, Li D, Choi TY, Dailing CA, Mao SS. 2013. Measurement of phantomless thoracic bone mineral density on coronary artery calcium CT scans acquired with various CT scanner models. *Radiology* 267:830-836.
10. Boden SD, Goodenough DJ, Stockham CD, Jacobs E, Dina T, Allman RM. 1989. Precise measurement of vertebral bone density using computed tomography without the use of an external reference phantom. *J Digit Imaging* 2:31-38.
11. Mueller DK, Kutscherenko A, Bartel H, Vlassenbroek A, Ourednicek P, Erckenbrecht J. 2011. Phantom-less QCT BMD system as screening tool for osteoporosis without additional radiation. *Eur J Radiol* 79:375-381.
12. Weaver AA, Beavers KM, Hightower RC, Lynch SK, Miller AN, Stitzel JD. 2015. Lumbar Bone Mineral Density Phantomless Computed Tomography Measurements and Correlation with Age and Fracture Incidence. *Traffic injury prevention* 16 Suppl 2:S153-160.
13. Boomsma MF, Slouwerhof I, van Dalen JA, Edens MA, Mueller D, Milles J, Maas M. 2015. Use of internal references for assessing CT density measurements of the pelvis as replacement for use of an external phantom. *Skeletal Radiol* 44:1597-1602.
14. Lee DC, Hoffmann PF, Kopperdahl DL, Keaveny TM. 2017. Phantomless calibration of CT scans for

- measurement of BMD and bone strength-Inter-operator reanalysis precision. *Bone* 103:325-333.
15. Lee YH, Kim JJ, Jang IG. 2019. Patient-Specific Phantomless Estimation of Bone Mineral Density and Its Effects on Finite Element Analysis Results: A Feasibility Study. *Computational and mathematical methods in medicine* 2019:4102410.
  16. Free J, Eggermont F, Derikx L, Van Leeuwen R, Van der Linden Y, Jansen W, Raaijmakers E, Tanck E, Kaatee R. 2018. The effect of different CT scanners, scan parameters and scanning setup on Hounsfield units and calibrated bone density: a phantom study. *Biomedical Physics & Engineering Express* 4:055013.
  17. Eggermont F, Derikx LC, Free J, Van Leeuwen R, Van der Linden YM, Verdonshot N, Tanck E. 2018. Effect of Different CT Scanners and Settings on Femoral Failure Loads Calculated by Finite Element Models. *J Orthop Res* 36:2288-2295.
  18. Giambini H, Dragomir-Daescu D, Huddleston PM, Camp JJ, An KN, Nassr A. 2015. The Effect of Quantitative Computed Tomography Acquisition Protocols on Bone Mineral Density Estimation. *J Biomech Eng* 137.
  19. Keyak JH, Kaneko TS, Tehranzadeh J, Skinner HB. 2005. Predicting proximal femoral strength using structural engineering models. *Clin Orthop*:219-228.
  20. Eggermont F, Derikx LC, Verdonshot N, Hannink G, Kaatee R, Tanck E, van der Linden YM. 2017. Limited short-term effect of palliative radiation therapy on quantitative computed tomography-derived bone mineral density in femora with metastases. *Adv Radiat Oncol* 2:53-61.
  21. Carpenter RD, Saeed I, Bonaretti S, Schreck C, Keyak JH, Streeper T, Harris TB, Lang TF. 2014. 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* 36:1225-1232.
  22. Derikx LC, Vis R, Meinders T, Verdonshot N, Tanck E. 2011. Implementation of asymmetric yielding in case-specific finite element models improves the prediction of femoral fractures. *Comput Methods Biomech Biomed Engin* 14:183-193.
  23. Aubrey J, Esfandiari N, Baracos VE, Buteau FA, Frenette J, Putman CT, Mazurak VC. 2014. Measurement of skeletal muscle radiation attenuation and basis of its biological variation. *Acta physiologica* 210:489-497.
  24. Birnbaum BA, Hindman N, Lee J, Babb JS. 2007. Multi-detector row CT attenuation measurements: assessment of intra- and interscanner variability with an anthropomorphic body CT phantom. *Radiology* 242:109-119.
  25. Levi C, Gray JE, McCullough EC, Hattery RR. 1982. The unreliability of CT numbers as absolute values. *AJR Am J Roentgenol* 139:443-447.
  26. Sande EP, Martinsen AC, Hole EO, Olerud HM. 2010. Interphantom and interscanner variations for Hounsfield units--establishment of reference values for HU in a commercial QA phantom. *Phys Med Biol* 55:5123-5135.

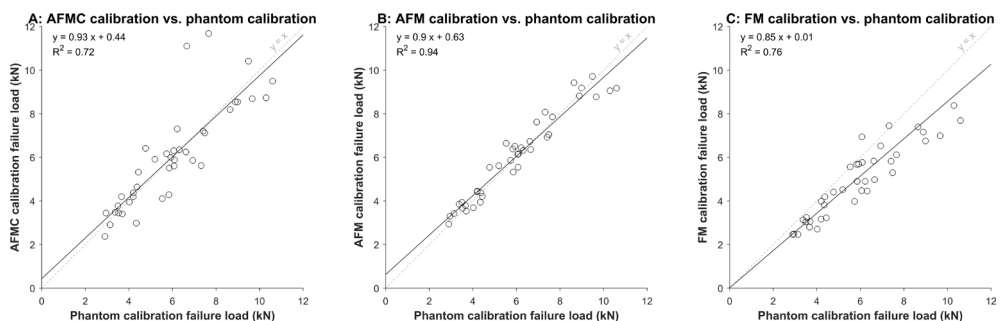
## Supplementary material

### Pilot study to determine the most accurate phantomless calibration method

For the drafting of the phantomless calibration, we first performed an extensive literature search to determine what methods were already being used. We found that calibration based on combinations of fat and muscle tissue [1-4] or external air and either aortic blood or visceral fat [5] were used for vertebral trabecular BMD measurements.

Subsequently, peaks for air, fat, muscle and cortical tissue were subtracted from the patients' CT scans. Nine diaphyseal slices were selected, starting with the slice containing no buttox or genitals and selecting the 8 consecutive slices. On the nine selected slices, a square region of interest was defined including the tissue of the right leg and some surrounding air ( $\pm 1$  cm on each side of the leg). Next, a histogram of the HU in this region was created to extract the peaks for air, fat and muscle tissue. The mode of the HU around the histogram peak ( $\pm 50$  HU) was calculated, to determine the exact peak in HU for each of the tissues. The determined HU peaks were linearly correlated to fixed "BMD" values of -840, -80, 30 and 1210 for air, fat, muscle and cortical tissue, respectively. The fixed "BMD" values were obtained by phantom calibrating the HU peaks of air, fat, muscle and cortical bone of a randomized subgroup comprising 10 patients scanned on the Philips-1 scanner, and subsequently averaging and rounding them.

For this pilot, we decided to continue with air-fat-muscle (AFM, because this method yielded the highest correlation with phantom calibration), fat-muscle (FM, because this method is used for BMD measurements in most studies) and air-fat-muscle-cortex (AFMC, because this method includes the largest range of HU and would avoid extrapolation of the calibration to the higher densities in cortical bone). For a subselection of patients ( $n = 40$ ), FE models were created using these phantomless calibration methods to determine the correlation with phantom calibration in terms of failure load (see Figure below). Since the AFM calibration resulted in the largest correlation ( $R^2 = 0.94$ , see Figure B), we chose to use this method as phantomless calibration.

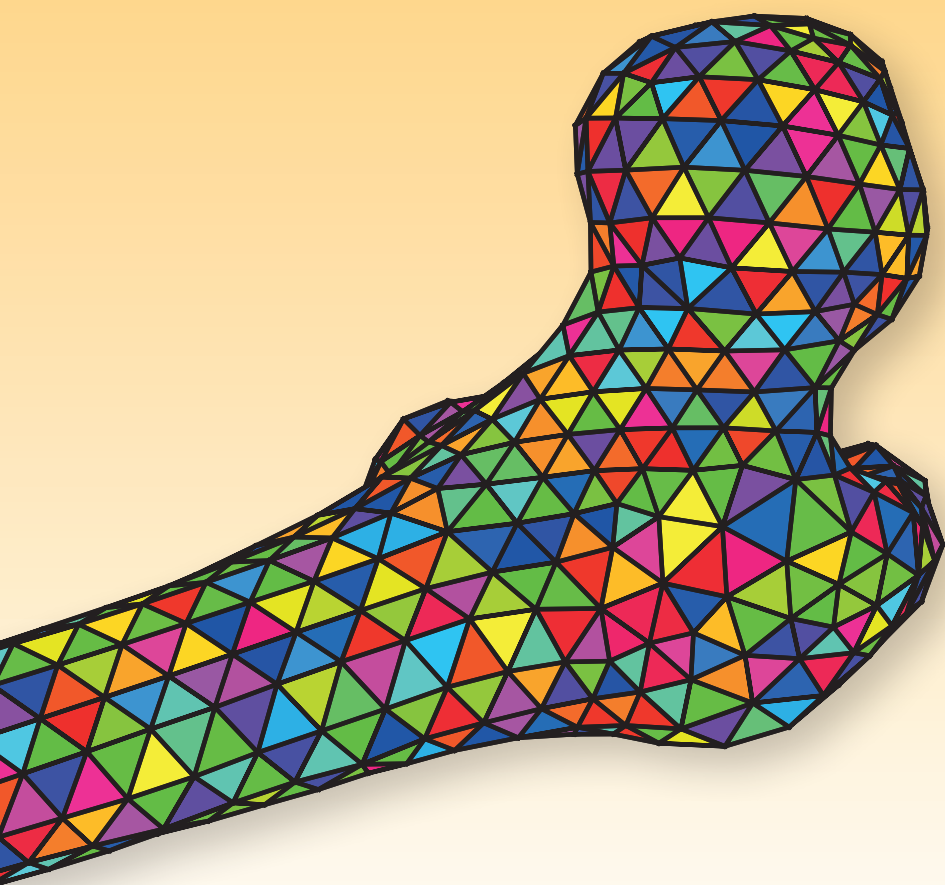


**Figure:** Correlations between phantom and air-fat-muscle-cortex (AFMC, figure A), air-fat-muscle (AFM, figure B) and fat-muscle (FM, figure C) calibrations. BW = body weight.

## References

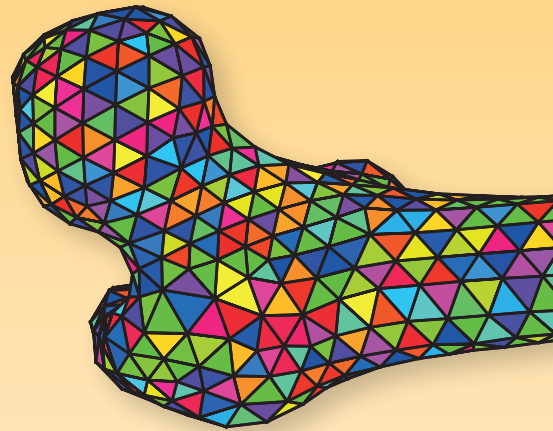
1. Boden SD, Goodenough DJ, Stockham CD, Jacobs E, Dina T, Allman RM. Precise measurement of vertebral bone density using computed tomography without the use of an external reference phantom. *J Digit Imaging*. 1989;2(1):31-8.
2. Mueller DK, Kutscherenko A, Bartel H, Vlassenbroek A, Ourednicek P, Erckenbrecht J. Phantom-less QCT BMD system as screening tool for osteoporosis without additional radiation. *Eur J Radiol*. 2011;79(3):375-81.
3. Weaver AA, Beavers KM, Hightower RC, Lynch SK, Miller AN, Stitzel JD. Lumbar Bone Mineral Density Phantomless Computed Tomography Measurements and Correlation with Age and Fracture Incidence. *Traffic injury prevention*. 2015;16 Suppl 2:S153-60.
4. Boomsma MF, Slouwerhof I, van Dalen JA, Edens MA, Mueller D, Milles J, *et al*. Use of internal references for assessing CT density measurements of the pelvis as replacement for use of an external phantom. *Skeletal Radiol*. 2015;44(11):1597-602.
5. Lee DC, Hoffmann PF, Kopperdahl DL, Keaveny TM. Phantomless calibration of CT scans for measurement of BMD and bone strength-Inter-operator reanalysis precision. *Bone*. 2017;103:325-33.





# CHAPTER 8

## **PATIENT-SPECIFIC FINITE ELEMENT COMPUTER MODELS IMPROVE FRACTURE RISK ASSESSMENTS IN CANCER PATIENTS WITH FEMORAL BONE METASTASES COMPARED TO CLINICAL GUIDELINES**



---

Florieke Eggermont, Gerco van der Wal, Paulien Westhoff, Arjonne Laar, Marianne de Jong,  
Tom Rozema, Herman Kroon, Onarisa Ayu, Loes Derikx, Sander Dijkstra, Nico Verdonshot,  
Yvette van der Linden, Esther Tanck

Bone, in press





## Introduction

Patients with bone metastases carry a risk of pathological fractures.<sup>1-3</sup> If a pathological fracture occurs in a weight-bearing bone such as the femur, this leads to an immediate decrease in the patient's mobility and self-care, and as a result in a reduced quality of life and possibly shortened survival.<sup>4,5</sup> Treatment of bone metastases in intact femurs is, therefore, based on the expected fracture risk.<sup>6,7</sup> Patients with an expected low fracture risk are treated with radiotherapy, usually a single fraction (SF) of 8 Gy, to relieve pain, whereas patients with an expected high fracture risk are considered firstly for preventive stabilizing surgery. In case of an expected high fracture risk, but the patient refuses surgery, radiotherapy in multiple fractions (MF) will be given, with the goal to prevent a pathological fracture by inducing remineralization.<sup>6-8</sup>

Currently, fracture risk assessment is based on available imaging such as conventional radiographs and CT scans, on which lesion characteristics like size<sup>7,9,10</sup> and radiographic appearance<sup>7,9-11</sup> are measured. Mirels et al.<sup>11</sup> developed a scoring system that is widely used for fracture risk assessment. This score combines ratings of pain, lesion type, size and location, and runs from 4 to 12. Generally, a patient should be considered for surgery if the Mirels' score is 9 or higher.<sup>11</sup> However, it is known that this score is very conservative and results in large numbers of overtreatment (positive predictive value (PPV) of 14%, negative predictive value (NPV) of 100%). As a consequence, patients who would never have developed a fracture during their remaining lifetime undergo surgery.<sup>7</sup>

Fracture risk can be assessed by measuring axial cortical involvement of the metastatic lesion on conventional radiographs,<sup>6</sup> which has shown to be more accurate compared to Mirels' scoring system.<sup>7</sup> Recently, the accuracy of the 30 mm threshold of axial cortical involvement was validated using 100 patients with 110 femoral bone metastases.<sup>12</sup> The negative predictive value (NPV) of the 30 mm threshold was high (96-97%), indicating that the 30 mm threshold was very accurate for ruling out pathological fractures. However, the positive predictive value (PPV) was limited (20-23%), which means that only one of four or five patients who were identified as high risk indeed fractured their femur during follow-up, indicating substantial overtreatment.<sup>6,12</sup>

Since actual pathological fractures result in higher morbidity and mortality<sup>4,5</sup> and are associated with longer hospital stays and higher costs<sup>13</sup> compared to prophylactic surgery of impending lesions, surgical overtreatment is generally accepted. However, unnecessary invasive treatments should be prevented as much as possible since this results in additional costs, hospitalisation and a certain risk of complications, especially in cancer patients with often poor general clinical condition and limited life expectancy. Hence, there is still room for improvement and a need to develop a more accurate fracture risk assessment tool.

A patient-specific finite element (FE) computer model based on quantitative CT scans (QCT) is a promising tool for fracture risk assessment.<sup>14-20</sup> In a recent cohort study,<sup>21</sup> we showed that the fracture risk assessments of the FE models were superior to those of experienced clinicians that assessed fracture risk in a test set-up on digitally reconstructed radiographs (DRRs). A limitation of that study was the poor visibility of the metastases on the DRRs. Therefore, the aim of the current study was

to compare fracture risk assessments by FE computer models with fracture risk assessments based on axial cortical involvement on diagnostic radiographs as described in current clinical guidelines for cancer patients with femoral bone metastases.

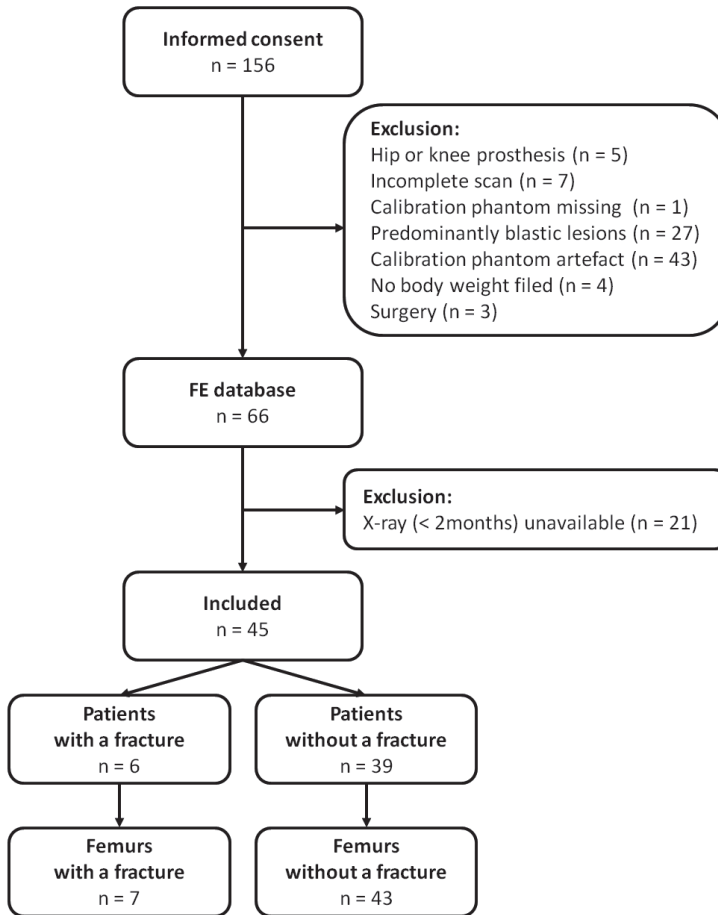
## Methods and Materials

### Patients

Two multicentre prospective cohort studies were performed between August 2006 and September 2009,<sup>21,22</sup> and between January 2015 and April 2017, with the aim to investigate fracture risk assessment utilizing FE models in patients with femoral bone metastases. Specific inclusion and exclusion criteria have been discussed elsewhere.<sup>21,22</sup> In summary, patients with advanced cancer and referred for radiotherapy of bone metastases in the femur were asked to participate in four radiotherapy institutes in the Netherlands (Radboud university medical center, Nijmegen; Leiden University Medical Center, Leiden; Radiotherapeutic Institute Friesland, Leeuwarden; Bernard Verbeeten Institute, Tilburg). Ethical approval was obtained from all participating centres. Patients were treated with radiotherapy according to the current clinical guidelines.<sup>6,7</sup> Lesions with an axial cortical involvement  $\leq 30$  mm were treated with 8 Gy SF. Lesions with an axial cortical involvement  $> 30$  mm were considered for prophylactic stabilizing surgery. However, if the patient's condition was too poor and surgery was undesirable or impossible, the patient received MF radiotherapy (e.g. 5 or 6 fractions of 4 Gy) to induce remineralisation of the bone.<sup>23</sup> If a patient was too ill to travel to the radiotherapy department for multiple fractions, it was accepted to deviate from the treatment guidelines and apply 8 Gy SF. Patients who were referred for surgery were not included in this study. In total, 156 patients gave informed consent. Patients were followed for six months or until a fracture occurred or until death, whichever occurred first.

Baseline characteristics (sex, age, primary tumour, radiotherapy schedule, pain (on a scale from 0 to 10), and Karnofsky performance status (KPS, on a scale from 0 to 100)<sup>24</sup>) were recorded prior to radiotherapy. QCT scans used for radiotherapy planning were made at baseline using a standardized protocol (120 kVp, 220 or variable mA, slice thickness 3 mm, pitch 1.5 or  $<1$ , spiral and standard reconstruction, field of view (FOV) 480 mm, in-plane resolution 0.9375 mm). In thirteen patients, it was not possible to create FE models due to a hip or knee prosthesis ( $n=5$ ), an incompletely scanned femur ( $n=7$ ), or a missing calibration phantom ( $n=1$ ) (see Figure 1 for the flow chart of patient inclusion for the current study). In a previous study it was shown that bone strength of femurs with osteoblastic lesions was overestimated<sup>21</sup>, probably due to the fact that the empirically established FE material model is not valid for the highly mineralized (pathological) bone tissue in osteoblastic lesions. As a result, patients with predominantly osteoblastic appearance were excluded ( $n=27$ ) from the analysis. Also, patients were excluded if the calibration was affected by an air artefact ( $n=43$ ),<sup>21</sup> causing a shading artifact on the upper half of the calibration phantom and therefore resulted in unreliable calibration functions. Additionally, four patients were excluded because their body weight was missing

from the clinical research files and three patients were excluded because they underwent preventive stabilizing surgery shortly after inclusion. In total, 66 patients (76 affected femurs) were included in the FE database.

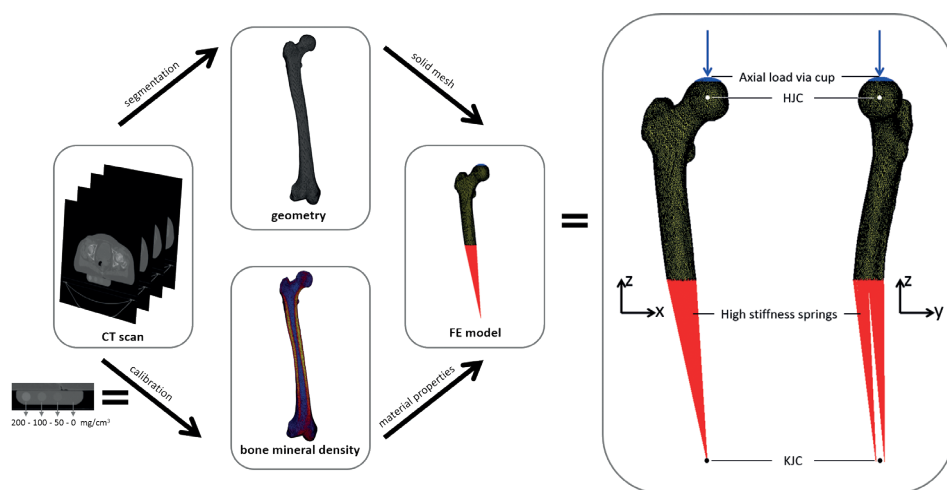


**Figure 1:** Flowchart of patient and femur inclusion

For the current study, we included the patients from the FE database with conventional anteroposterior (AP) and/or lateral radiographs available within a two month period prior to radiotherapy.<sup>12</sup> Patients were excluded from the current study if no radiographs were available (n=21).<sup>12</sup> This resulted in inclusion of 45 patients, with 50 affected femurs (Figure 1). Twelve of these femurs were part of the study group published before.<sup>21</sup>

## FE models

Patient-specific femoral FE models were generated as described previously.<sup>16,21</sup> In short, for each irradiated femur, the three-dimensional geometry was segmented from the CT scan (Mimics 11.0 and 14.0, Materialise, Leuven, Belgium), and subsequently converted into a solid mesh of tetrahedral elements (Patran 2005r2 and 2011, MSC Software Corporation, Santa Ana, CA, USA). Additionally, the CT scans were calibrated with the use of a solid calibration phantom containing known calcium equivalent densities (Image Analysis, Columbia, KY, USA) that was scanned along with the patient. With the use of this calibration, Hounsfield units were converted to calcium equivalent values, which were used to calculate non-linear isotropic material behaviour for each tetrahedral element based on the material model of Keyak et al.<sup>25</sup> To correct for inter-scanner differences,<sup>26-28</sup> cross-calibration using phantom scans (Gammex 467 phantom, RMI Gammex, Middleton, WI, USA)<sup>27</sup> was performed per scanner.



**Figure 2:** The workflow of generating the FE model. The CT scan is used to obtain the geometry, which is converted into a solid mesh. Additionally, the CT is calibrated using the calibration phantom to obtain bone mineral densities, which are used to calculate non-linear isotropic material behaviour for each tetrahedral element. The FE model is distally fixed at the knee joint centre (KJC) using two high-stiffness springs. Load was applied by displacing a cup on the femoral head in line with the hip joint centre (HJC) in axial direction.



**Figure 3:** Example of a measurement of the axial cortical involvement of the metastases

The FE models of the proximal femur were distally fixed at the knee joint centre using two high-stiffness springs (200.000.000 N/m) and loaded by incrementally displacing a cup on the femoral head in axial direction (Figure 2, MSC.MARC 2007r1 and 2013.1, MSC Software Corporation, Santa Ana, CA, USA). During the FE simulation, incremental displacement and contact normal forces were recorded. Failure load of the femur was defined as the maximum total reaction force.<sup>16,21,25</sup> Failure loads were normalized for body weight (BW). The previously determined critical threshold of failure load of  $7.5 \times \text{BW}$  was applied to distinguish low from high fracture risk femurs: patients with a failure load of  $7.5 \times \text{BW}$  or lower were identified as having high fracture risk, whereas patients with a failure load higher than  $7.5 \times \text{BW}$  were classified as low fracture risk.<sup>21</sup>

### **Clinical guidelines: 30mm threshold of axial cortical involvement**

Three experienced assessors (radiation oncologist, orthopaedic surgeon and skeletal radiologist) individually measured the axial cortical involvement on the available radiographs as described in earlier studies (Figure 3).<sup>6,12</sup> According to current clinical guidelines, the femur was defined at high risk of fracture if the axial cortical involvement of the metastasis was over 30 mm.<sup>6</sup> In case of disagreement, consensus was reached through discussion between the three assessors.

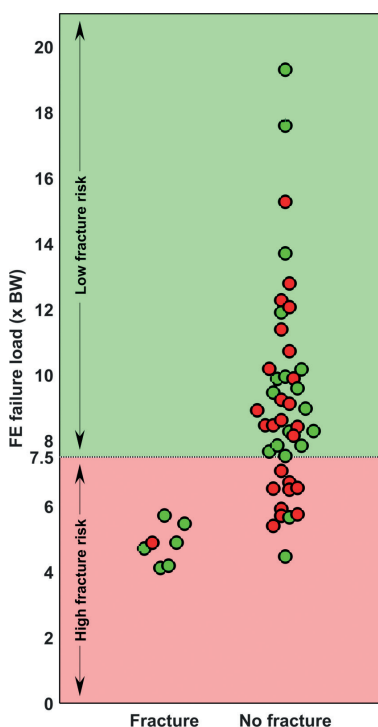
### **Statistical analysis**

Baseline characteristics were compared between the groups of patients who did and did not develop a pathological fracture during follow-up using Mann-Whitney U (age, pain, KPS), Fisher exact (sex,

radiotherapy schedule, affected femur) and Pearson  $\chi^2$  (primary tumour) tests. FE predictions were compared to clinical assessments by means of diagnostic accuracy values: sensitivity, specificity, PPV and NPV. We compared sensitivities and specificities between FE and clinical assessments using McNemar's test.<sup>29</sup>

## Results

In total, 45 patients with 50 affected femurs were included in this study. Six patients developed seven fractures (14%). Median time to fracture was 8 weeks (range 1-18). Thirteen patients died during the 6 month follow-up (33%). There were no significant differences in baseline characteristics between patients who did and did not develop a femoral fracture (Table 1). Five out of the seven fractured femurs had been treated with 8 Gy SF. Examples of CT images of a few patients can be found in Supplementary Figure 1.



**Figure 4:** Femoral failure load corrected for bodyweight for the femurs that did or did not fracture during follow-up. The threshold at 7.5 x BW is based on a previous study<sup>21</sup> and is used to differentiate between high and low fracture risk. According to the FE model, 18 femurs had a high fracture risk, of which 7 actually fractured during follow-up. Green dots indicate correct clinical assessments and red dots indicate incorrect assessments based on the 30 mm threshold of axial cortical involvement according to the clinical guidelines

According to the FE model, all seven fractures had a failure load below 7.5 x BW, and were accordingly correctly assessed as high risk, resulting in a sensitivity of 100% (Figure 4, Table 2). Based on the 30 mm threshold, all but one of the fractures were correctly identified (sensitivity of 86%, difference of 14%, 95% confidence interval of -23% to 51%). Of the 43 non-fractured femurs, the FE model correctly predicted 32 femurs as low risk (specificity of 74%), whereas 18 non-fractured femurs were accurately assessed as low risk (specificity of 42%, statistical significant difference of 32%, 95% confidence interval of 14% to 47%) by using the 30 mm threshold. The NPV of the FE model was 100%, demonstrating that none of the femurs with a low fracture risk as calculated by the FE model fractured. The NPV of the axial cortical involvement was slightly lower (95%). The PPV of the FE model was 39%, indicating that 39% of the femurs with a high fracture risk assessment in reality fractured. For the axial cortical involvement, the PPV was 19%.

There was no correlation between the femoral failure load corrected for bodyweight and the measured axial cortical involvement measured on conventional radiographs (Figure 5).

**Table 1:** Patient characteristics of the 45 included patients of whom 6 developed a pathological fracture during follow-up

	Patients with a femoral fracture (n=6)	Patients without a femoral fracture (n=39)	p-value
<b>Sex</b>			0.7
Male	2 (33%)	19 (49%)	
Female	4 (67%)	20 (51%)	
<b>Age at inclusion</b>			0.9
Mean (SD), years	64.2 (6.2)	63.5 (10.9)	
<b>Primary tumour</b>			0.5
Breast	2 (33%)	11 (28%)	
Prostate	0 (0%)	4 (10%)	
Lung	2 (33%)	12 (31%)	
Multiple myeloma	2 (33%)	5 (13%)	
Other	0 (0%)	7 (18%)	
<b>Radiotherapy schedule</b>			0.4
Single fraction (1 x 8 Gy)	5 (83%)	24 (62%)	
Multiple fractions (5-6 x 4 Gy)	1 (17%)	15 (38%)	
<b>Pain score</b>			0.4
Mean (SD)	7.5 (2.5)	5 (2.5)	
<b>Karnofsky performance status<sup>24</sup></b>			0.6
Mean (SD)	68.3 (14.7)	79.3 (10.1)	
<b>Affected femur</b>			1
Unilateral	5 (83%)	34 (87%)	
Bilateral	1 (17%)	5 (13%)	

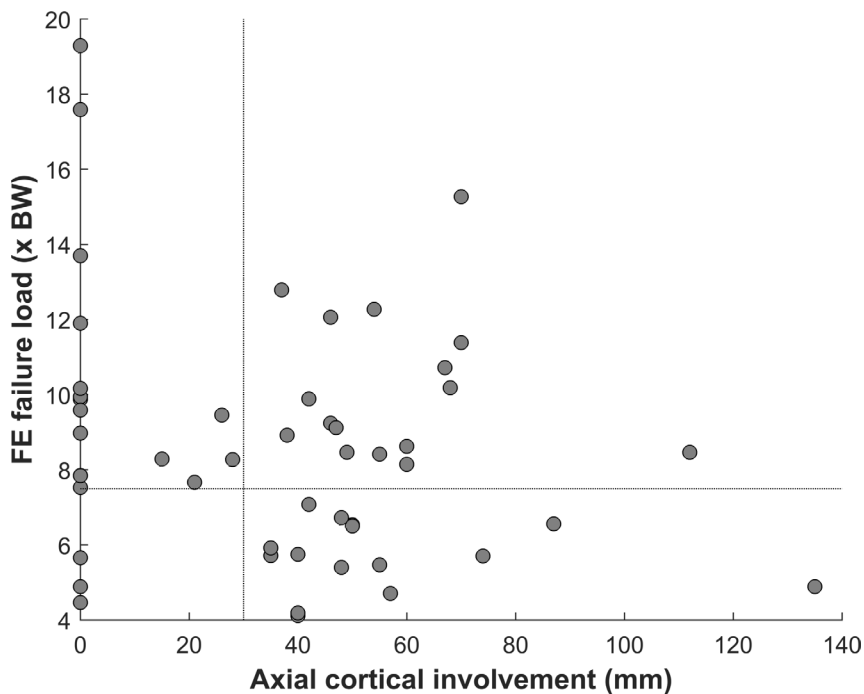
Pain was assessed at baseline using a pain score ranging from 0 (= no pain) to 10 (= worst imaginable pain). Karnofsky performance status<sup>24</sup> ranging from 0 (= lowest performance) to 100 (= highest performance).



**Table 2:** Diagnostic accuracy values of the FE model and the clinical assessments based on 30 mm threshold for axial cortical involvement of the femoral metastases on radiographs.

	Femurs that fractured during follow-up (n=7)	Femurs that did not fracture during follow-up (n=43)	Sensitivity	Specificity	PPV	NPV
<b>Axial cortical involvement<sup>6</sup></b>						
>30 mm (high risk)	6	25	86%	42%	19%	95%
≤30 mm (low risk)	1	18				
<b>FE model<sup>21</sup></b>						
≤7.5 (high risk)	7	11	100%	74%	39%	100%
>7.5 (low risk)	0	32				

PPV = positive predictive value, NPV = negative predictive value.



**Figure 5:** The correlation between the femoral failure load corrected for bodyweight and the measured axial cortical involvement measured on conventional radiographs. The dotted lines depict the thresholds for differentiating between high and low fracture risk.

## Discussion

For assessment of expected fracture risk in patients with femoral bone metastases, this study compared FE computer models with axial cortical involvement on conventional radiographs as described in current clinical guidelines. FE models were better at assessing fracture risk in comparison to the clinical guidelines: the fracture risk of more femurs, either high or low, was correctly assessed. Clinically, fracture risk is estimated based on radiographs or CT scans, and FE models are currently not used. Clinical guidelines, such as the 30 mm threshold of axial cortical involvement, have been constructed to align the method for fracture risk assessment between different medical specialists. However, previous studies have shown that substantial numbers of patients are under- and overtreated based on such clinical guidelines.<sup>6,7,12</sup> Although promising, fracture risk assessment by the FE model still resulted in a fair number of false positives (11), although there were even more when using axial cortical involvement (25). Remarkably, five out of the seven femurs that developed a fracture during follow-up were treated with SF. Since we have no information about the initial fracture risk assessment, we can only speculate on why this is the case. These patients were either incorrectly assessed as low fracture risk at inclusion, or had insufficient clinical condition to undergo MF radiotherapy. However, the latter reason seems doubtful, as patients had to be in quite good clinical condition ( $KPS^{24} \geq 60$ ) to be included in the study. Therefore, this shows that the fracture risks were probably not always assessed according to the axial cortical involvement as stated in the clinical guidelines, which underlines the need for a better, standardized and more reliable fracture risk assessment tool.

Previously, we performed a comparable study,<sup>21</sup> in which we also generated FE models and compared the FE with clinical fracture risk assessments based on the axial cortical involvement measured on DRRs instead of conventional radiographs. Such DRRs are reconstructed radiographs based on the CT images with a rather coarse resolution ( $0.9375 \times 0.9375 \times 3$  mm). Hence, the DRR image quality was rather poor in comparison to conventional radiographs. In that study, we included 47 femurs of which nine fractured (twelve of these femurs are part of the current study too) and found that the FE model had higher sensitivity, i. e. was more accurate in identifying patients with a high fracture risk, compared to clinical assessments, whereas specificity was lower for the FE models than for the clinical assessments on DRRs.<sup>21</sup> With respect to our previous study, the sensitivity and NPV of the FE model increased (89% for the previous study vs. 100% for the current study, and 97% vs. 100%, respectively), whereas specificity and PPV were slightly lower (79% vs. 74% and 50% vs. 39%, respectively). This can be explained by the fact that part of the CT scans used in the previous study were affected by air artefacts, in addition to possible inter-scanner differences,<sup>26-28</sup> and were, therefore, analyzed for each institute separately. In the current FE database, the CT scans with air artefacts were excluded and inter-scanner differences were corrected for, resulting in more accurate FE fracture predictions. Furthermore, the sensitivity and NPV of the clinical assessments in the previous study were lower (between 0% and 22% vs. 86% for the current study, and between 80% and 84% vs. 95%, respectively), whereas mainly specificity was much higher (between 84% and 97% vs. 42%) in contrast to that of the current study. PPV varied largely in the previous study between the different clinicians (0% to

50%). The diagnostic accuracy values of the current study are closer to diagnostic accuracy of previous studies using the 30 mm threshold on conventional radiographs.<sup>6,12</sup> These deviating results of our previous study can be explained by the use of DRRs instead of radiographs for the clinical assessments on which the lesions were often not well visible, and, consequently, resulted in more femurs assessed as low risk. In the current study, conventional radiographs were used to measure the axial cortical involvement, and, therefore, these clinical assessments were more valid.

Another study by Goodheart et al.<sup>30</sup> compared fracture risk estimation by FE models with Mirels' scoring system. Mirels' scoring system, based on ratings of pain, lesion type, size and location,<sup>11</sup> is, however, known to overestimate fracture risk, leading to large numbers of overtreatment,<sup>31</sup> which can be reduced when applying the 30 mm threshold.<sup>7</sup> Goodheart et al.<sup>30</sup> showed that the FE model and Mirels' scoring system had similar sensitivity, whereas specificity was higher for the FE model compared to Mirels'. They concluded that FE models can improve fracture prediction over clinical assessments based on Mirels' scoring system.<sup>30</sup> Previously, Van der Linden et al.<sup>7</sup> and recently, Van der Wal et al.<sup>12</sup> showed that fracture risk assessment based on axial cortical involvement of the metastases was more accurate in comparison with Mirels' scoring system. In the current study, we showed that FE models can further improve these fracture risk assessments.

Other biomechanical methods, such as computed tomography rigidity analysis (CTRA)<sup>32,33</sup> and high resolution MRI-based FE models<sup>34</sup> also show promising results to be used to assess fracture risk. However, to our knowledge, the MRI-based FE models have not yet been tested in patients with actual metastatic lesions. Also, both CTRA as MRI-based FE models are not yet clinically being implemented. This study had some limitations. Firstly, conventional radiographs were not available for all 66 patients (representing 76 femurs with bone metastases) in the FE database, resulting in exclusion of 21 patients (26 femurs). Secondly, patients with large lesions and an expected high fracture risk who were surgically treated were not included. Additionally, we only included patients already referred for palliative radiotherapy, so patients without symptoms were not included in these analyses. Consequently, we cannot conclude on whether the FE model would prevent any unnecessary surgeries in those patient groups. Thirdly, some patients died shortly after inclusion or became immobile and, consequently, did not develop femoral fractures during follow-up. They might have developed fractures if they had lived longer or had engaged in activities which would increase load bearing onto their legs.

FE models are probably better at assessing fracture risk in comparison to simple measurements on conventional radiographs due to the fact that they consider for example location and 3D geometry of the lesion, general bone quality, or geometry of the bone, whereas measuring the lesion in the cortex on a two dimensional radiograph does not take these factors into account. To this date, the FE model has only been used in research settings, but we are currently working towards clinical implementation. To enable implementation, we have developed a phantomless calibration method,<sup>35</sup> which facilitates widespread use of QCT and FE by avoiding the requirement of a calibration phantom during the CT scan session. For this phantomless calibration, CT densities of air, fat and muscle tissue are used, which results in FE failure loads comparable to those calculated using the conventional phantom calibration.<sup>35</sup> A limitation of the FE models is that they are currently not applicable for patients with

predominantly osteoblastic lesions. This is due to the fact that the material model used is apparently not valid for the high CT density of such lesions and, hence, results in aberrant material properties.<sup>21</sup> None of the patients affected with predominantly osteoblastic lesions developed a fracture during follow-up. In the near future, if patients with femoral bone metastases visit their medical specialist, an FE fracture risk assessment can be ordered, just like ordering a lab-test. The treating physician can then discuss the results of the FE model with the patient, taking into account clinical factors such as the patient's clinical condition, life expectancy, and activity level.

In conclusion, this study showed that patient-specific FE models improve femoral fracture risk assessments in comparison to measuring axial cortical involvement on conventional radiographs as described in the current clinical guidelines on bone metastases. The FE models could prevent unnecessary surgical procedures and, therefore, improve quality of life of those patients. Therefore, clinical implementation of the FE models is supported.

## Acknowledgements

The authors thank Leon Driessen and Max Bakker for their help in generating the patient-specific FE models. This work was supported by the Dutch Science Foundation NWO-STW (NPG.06778), Fonds NutsOhra (1102-071), Furlong Research Charitable Foundation, the Dutch Cancer Society (KUN 2012-5591), and the Dutch Cancer Society/Alpe d'HuZes (UL2013-6286).

## References

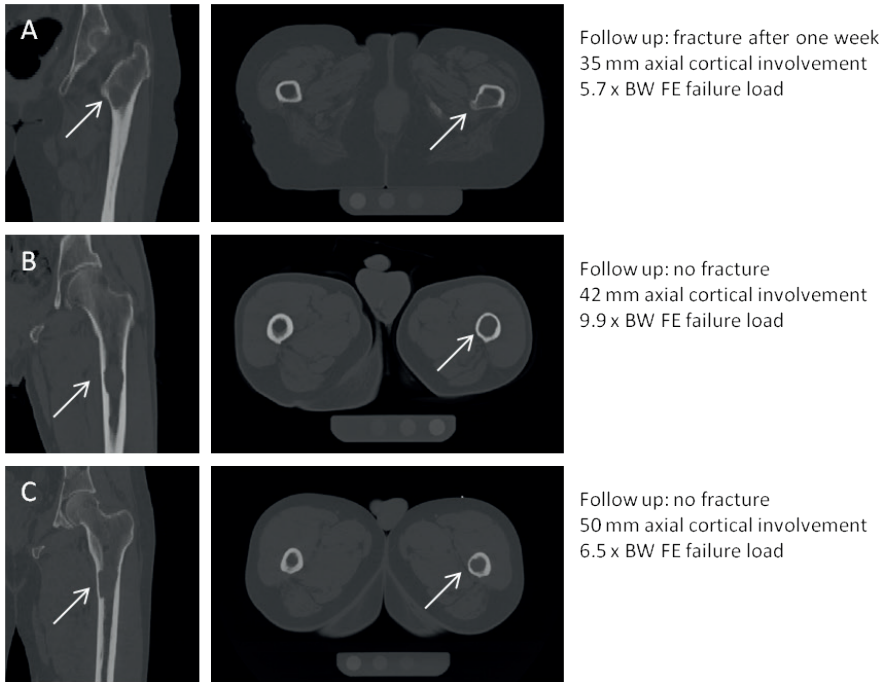
1. Coleman RE. 1997. Skeletal complications of malignancy. *Cancer* 80:1588-1594.
2. Coleman RE. 2006. Clinical features of metastatic bone disease and risk of skeletal morbidity. *Clin Cancer Res* 12:6243s-6249s.
3. Mundy GR. 1997. Mechanisms of bone metastasis. *Cancer* 80:1546-1556.
4. Ratasvuori M, Wedin R, Keller J, Nottrott M, Zaikova O, Bergh P, Kalen A, Nilsson J, Jonsson H, Laitinen M. 2013. Insight opinion to surgically treated metastatic bone disease: Scandinavian Sarcoma Group Skeletal Metastasis Registry report of 1195 operated skeletal metastasis. *Surg Oncol* 22:132-138.
5. Mavrogenis AF, Pala E, Romagnoli C, Romantini M, Calabro T, Ruggieri P. 2012. Survival analysis of patients with femoral metastases. *J Surg Oncol* 105:135-141.
6. Van der Linden YM, Kroon HM, Dijkstra SP, Lok JJ, Noordijk EM, Leer JW, Marijnen CA. 2003. Simple radiographic parameter predicts fracturing in metastatic femoral bone lesions: results from a randomised trial. *Radiother Oncol* 69:21-31.
7. Van der Linden YM, Dijkstra PD, Kroon HM, Lok JJ, Noordijk EM, Leer JW, Marijnen CA. 2004. Comparative analysis of risk factors for pathological fracture with femoral metastases. *J Bone Joint Surg Br* 86:566-573.
8. Koswig S, Budach V. 1999. Remineralisation und Schmerzlinderung von Knochenmetastasen nach unterschiedlich fraktionierter Strahlentherapie (10 mal 3 Gy vs. 1 mal 8 Gy). Eine prospektive Studie. *Strahlenther Onkol* 175:500-508.
9. Snell W, Beals RK. 1964. Femoral Metastases and Fractures from Breast Cancer. *Surg Gynecol Obstet* 119:22-24.
10. Beals RK, Lawton GD, Snell WE. 1971. Prophylactic internal fixation of the femur in metastatic breast cancer. *Cancer* 28:1350-1354.
11. Mirels H. 1989. Metastatic disease in long bones. A proposed scoring system for diagnosing impending pathologic fractures. *Clin Orthop* 256-264.
12. Van der Wal CWPG, Eggermont F, Kroon HM, Ayu O, Fiocco M, Slot A, Snyers A, Rozema T, Verdonschot N, Dijkstra PDS, Tanck E, Van der Linden YM. Radiotherapy & Oncology, in press. Validation of a simple radiographic parameter to predict fracturing in metastatic femoral bone lesions.
13. Blank AT, Lerman DM, Patel NM, Rapp TB. 2016. Is Prophylactic Intervention More Cost-effective Than the Treatment of Pathologic Fractures in Metastatic Bone Disease? *Clin Orthop Relat Res* 474:1563-1570.
14. Yosibash Z, Mayo RP, Dahan G, Trabelsi N, Amir G, Milgrom C. 2014. Predicting the stiffness and strength of human femurs with real metastatic tumors. *Bone* 69:180-190.
15. Keyak JH, Kaneko TS, Rossi SA, Pejicic MR, Tehranzadeh J, Skinner HB. 2005. Predicting the strength of femoral shafts with and without metastatic lesions. *Clin Orthop* 439:161-170.
16. Derikx LC, van Aken JB, Janssen D, Snyers A, van der Linden YM, Verdonschot N, Tanck E. 2012. The assessment of the risk of fracture in femora with metastatic lesions: comparing case-specific finite element analyses with predictions by clinical experts. *J Bone Joint Surg Br* 94:1135-1142.
17. Schileo E, Taddei F, Cristofolini L, Viceconti M. 2008. Subject-specific finite element models implementing a maximum principal strain criterion are able to estimate failure risk and

- fracture location on human femurs tested in vitro. *J Biomech* 41:356-367.
18. Orwoll ES, Marshall LM, Nielson CM, Cummings SR, Lapidus J, Cauley JA, Ensrud K, Lane N, Hoffmann PR, Kopperdahl DL, Keaveny TM, Osteoporotic Fractures in Men Study G. 2009. Finite element analysis of the proximal femur and hip fracture risk in older men. *J Bone Miner Res* 24:475-483.
  19. Keyak JH, Sigurdsson S, Karlsdottir GS, Oskarsdottir D, Sigmarsdottir A, Kornak J, Harris TB, Sigurdsson G, Jonsson BY, Siggeirsdottir K, Eiriksdottir G, Gudnason V, Lang TF. 2013. Effect of finite element model loading condition on fracture risk assessment in men and women: the AGES-Reykjavik study. *Bone* 57:18-29.
  20. Benca E, Synek A, Amini M, Kainberger F, Hirtler L, Windhager R, Mayr W, Pahr DH. 2019. QCT-based finite element prediction of pathologic fractures in proximal femora with metastatic lesions. *Sci Rep* 9:10305.
  21. Eggermont F, Derikx LC, Verdonschot N, Van der Geest ICM, De Jong MAA, Snyers A, Van der Linden YM, Tanck E. 2018. Can patient-specific finite element models better predict fractures in metastatic bone disease than experienced clinicians? Towards introducing computational modelling into daily clinical practice. *Bone Joint Res* 7:430-439.
  22. Eggermont F, Derikx LC, Verdonschot N, Hannink G, Kaatee R, Tanck E, van der Linden YM. 2017. Limited short-term effect of palliative radiation therapy on quantitative computed tomography-derived bone mineral density in femora with metastases. *Adv Radiat Oncol* 2:53-61.
  23. Koswig S, Budach V. 1999. [Remineralization and pain relief in bone metastases after after different radiotherapy fractions (10 times 3 Gy vs. 1 time 8 Gy). A prospective study]. *Strahlenther Onkol* 175:500-508.
  24. Karnofsky D, Burchenal J. 1949. The clinical evaluation of chemotherapeutic agents in cancer. In: MacLeod C editor. *Evaluation of Chemotherapeutic Agents*. New York: Columbia University Press; pp. 191-205.
  25. Keyak JH, Kaneko TS, Tehranzadeh J, Skinner HB. 2005. Predicting proximal femoral strength using structural engineering models. *Clin Orthop*:219-228.
  26. Carpenter RD, Saeed I, Bonaretti S, Schreck C, Keyak JH, Streeper T, Harris TB, Lang TF. 2014. 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* 36:1225-1232.
  27. Free J, Eggermont F, Derikx L, Van Leeuwen R, Van der Linden Y, Jansen W, Raaijmakers E, Tanck E, Kaatee R. 2018. The effect of different CT scanners, scan parameters and scanning setup on Hounsfield units and calibrated bone density: a phantom study. *Biomedical Physics & Engineering Express* 4.
  28. Eggermont F, Derikx LC, Free J, Van Leeuwen R, Van der Linden YM, Verdonschot N, Tanck E. 2018. Effect of Different CT Scanners and Settings on Femoral Failure Loads Calculated by Finite Element Models *J Orthop Res*.
  29. Newcombe RG. 2001. Simultaneous comparison of sensitivity and specificity of two tests in the paired design: a straightforward graphical approach. *Stat Med* 20:907-915.
  30. Goodheart JR, Cleary RJ, Damron TA, Mann KA. 2015. Simulating activities of daily living with finite element analysis improves fracture prediction for patients with metastatic femoral lesions. *J Orthop Res* 33:1226-1234.

31. Damron TA, Morgan H, Prakash D, Grant W, Aronowitz J, Heiner J. 2003. Critical evaluation of Mirels' rating system for impending pathologic fractures. *Clin Orthop Relat Res*:S201-207.
32. Anez-Bustillos L, Derikx LC, Verdonchot N, Calderon N, Zurakowski D, Snyder BD, Nazarian A, Tanck E. 2014. Finite element analysis and CT-based structural rigidity analysis to assess failure load in bones with simulated lytic defects. *Bone* 58:160-167.
33. Nazarian A, Entezari V, Zurakowski D, Calderon N, Hipp JA, Villa-Camacho JC, Lin PP, Cheung FH, Aboulafia AJ, Turcotte R, Anderson ME, Gebhardt MC, Cheng EY, Terek RM, Yaszemski M, Damron TA, Snyder BD. 2015. Treatment Planning and Fracture Prediction in Patients with Skeletal Metastasis with CT-Based Rigidity Analysis. *Clin Cancer Res* 21:2514-2519.
34. Rajapakse CS, Gupta N, Evans M, Alizai H, Shukurova M, Hong AL, Cruickshank NJ, Tejwani N, Egol K, Honig S, Chang G. 2019. Influence of bone lesion location on femoral bone strength assessed by MRI-based finite-element modeling. *Bone* 122:209-217.
35. Eggermont F, Verdonchot N, Van der Linden Y, Tanck E. 2019. Calibration with or without phantom for fracture risk prediction in cancer patients with femoral bone metastases using CT-based finite element models. *Plos One* 14:e0220564.

## Supplementary material

**Supplementary Figure 1:** Example CT images of A) a patient that developed a fracture, and B) and C) patients that did not develop a fracture during follow-up







# CHAPTER 9

## GENERAL DISCUSSION AND FUTURE PERSPECTIVES





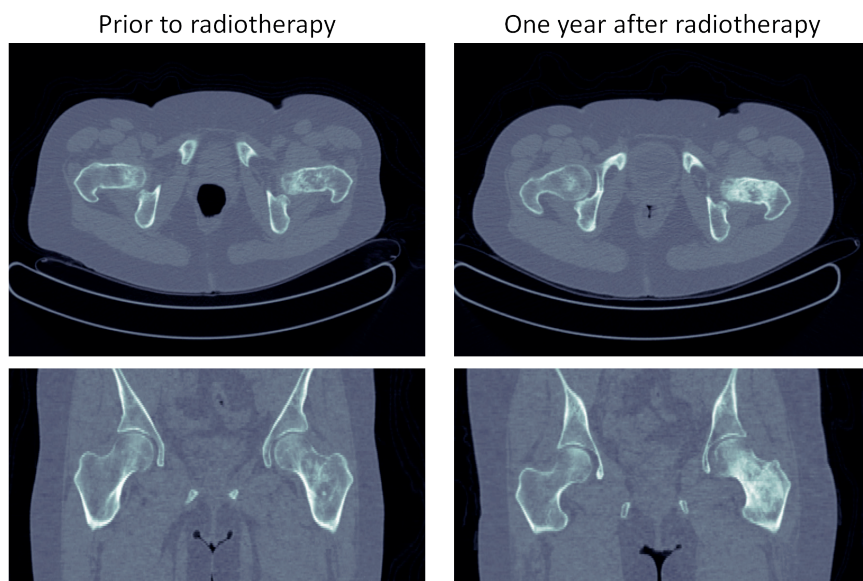
Patients with advanced cancer and femoral bone metastases have an increased risk of fracturing their femur as a result of reduced load-bearing capacities caused by the metastatic disease. Such pathological fractures lead to immediate reduced mobility, pain and distress, and by that, a reduced quality of life. Moreover, pathological fractures may lead to a cascade of events such as deteriorating condition, bedridden and consequentially a shortened remaining survival. The goal is therefore to prevent pathological fracturing. The choice of treatment of metastatic bone lesions without evident fracture, i.e. before pathological fracturing, is based on the fracture risk estimated by the multidisciplinary clinical team: patients with an expected high fracture risk are considered for prophylactic stabilizing surgery to reduce the chance of fracturing, whereas patients with an expected low fracture risk or insufficient clinical condition to undergo surgery, or patients who refuse surgery, are treated with conservative treatment such as radiotherapy, to diminish the pain and, if possible, induce remineralization to restore bone quality and as a result reduce the fracture risk. Fracture risk assessment is subjective and dependent on the treating clinician, and therefore, so is the current treatment of femoral bone metastases. The current Dutch clinical guideline states that patients with femoral lesions with over 30 mm axial cortical involvement as measured on conventional radiographs have a 23% risk of impending fracturing, and should be surgically stabilized.<sup>1,2</sup> This easy-to-use method for fracture risk assessment is scarcely being used in other countries than the Netherlands. In **Chapter 2**, we evaluated this 30 mm threshold on its ability to distinguish between high and low fracture risk patients. We found that clinical assessments were better when based on 30 mm axial cortical involvement in comparison to fracture risk estimations based on expert opinion, and should therefore be used in all clinical decision making, until a more reliable fracture risk assessment method is available. However, there were differences between observers in their assessments on whether or not a lesion was larger than 30 mm, and there was room for improvement in fracture risk prediction. Therefore, this study also demonstrated that the assessment of fracture risk using measurements on conventional radiographs or CT scans remains challenging, and as such, there is a necessity for a more accurate fracture risk assessment tool for patients with bone metastases.

Our previous work and this thesis showed that utilisation of patient-specific finite element (FE) computer models is a promising tool for predicting fracture risk. Over the past years, at the Orthopaedic Research Laboratory (ORL), such an FE model has been developed and validated *in vitro* using cadaveric femurs with artificial lesions created by drilling defects in the cortex, which were loaded under axial compression to measure strength.<sup>3,4</sup> However, this experimental test set-up is evidently a simplified representation of the physiological circumstances. *In vivo*, metastases consist of lytic or blastic tissue, and probably behave mechanically differently compared to defects drilled in healthy bone. Additionally, fractures do not only occur under axial loading conditions. Furthermore, CT images that function as input to the FE models differ between *in vitro* and *in vivo* situations: *in vitro*, femurs are scanned in a water basin, whereas under *in vivo* circumstances the scanned femur is affected by beam hardening due to bony structures and soft tissue around the femurs. Therefore, the FE model needs to be validated *in vivo* in patients, before it can be used in clinical practice in the future. The goal of this thesis was

to further develop and validate the patient-specific FE model to predict fracture risk in patients with cancer and bone metastases, and to enhance the applicability for future clinical implementation.

## Remineralization of metastatic bone lesions

We performed a multicentre prospective patient cohort study (CT femur study, funded by the Dutch Science Foundation NWO-STW (NPG.06778)) with the aim to *in vivo* validate the FE model (see appendix for the flowchart of the patients in all studies). In three radiotherapy institutes in the Netherlands, patients who were treated with single (SF) or multiple (MF) fraction radiotherapy (RT) for pain were included. Patients underwent CT scans prior to, and 1, 4 and 10 weeks after RT, and filled in questionnaires on physical activity, pain and quality of life. With the use of the CT scans, we were able to determine the short-term effect of RT on bone mineral density (BMD) in proximal femurs as well as within the lytic, blastic and mixed metastatic lesions (**Chapter 3**). We found that, in contradiction to previous studies,<sup>5-9</sup> BMD of lytic lesions did not significantly increase after RT, whereas BMD in both mixed and blastic lesions increased. Additionally, there was no difference between SF and MF RT in terms of remineralization. Correspondingly, a recent systematic review showed that there is currently no evidence for any effect of RT on bone quality or fracture risk.<sup>10</sup> Thus, using RT to achieve short-term remineralization seems to be doubtful. However, clinical experience is that remineralization does occur (Figure 1).



**Figure 1:** Clinical example of a patient from the Netherlands Cancer Institute – Antoni van Leeuwenhoek Amsterdam showing remineralization of bone metastases in the right femur one year after radiotherapy (5 x 7 Gy).

Possibly we would have found higher response rates if the follow-up period had extended 12 weeks<sup>5,9</sup> or total doses were higher.<sup>5</sup> Additionally, recently, a German study investigated the effect of several multiple fraction RT schedules, between 20 and 40 Gy in total, on remineralization in terms of changes in stability of spinal lytic metastases after 3 and 6 months.<sup>11</sup> They retrospectively analyzed HU in manually selected regions from a total of 826 patients at 3 months and 787 patients at 6 months, and reported whether the stability of the vertebral metastases changed after MF RT. They concluded that significantly more vertebrae were classified as stable after 3 and 6 months of MF RT, indicating that there was a remineralizing effect of RT. The same research group also performed two randomized trials using 55 and 60 patients with spinal bone metastases to determine differences in local responses between stereotactic body RT (SBRT, 24 Gy) and three-dimensional conformal RT (3DCRT, 10 x 3 Gy)<sup>12</sup> and between intensity-modulated RT (IMRT, 10 x 3 Gy) and 3DCRT (10 x 3 Gy),<sup>13</sup> respectively. In both studies, they showed that bone density had significantly increased 3 and 6 months after all different types of RT. The latter studies did not take external beam RT into account, but did show that other kinds of RT can result in bone density increases. It should be mentioned that in the abovementioned studies, again, the bone density was measured manually, and thus accuracy can be questioned. Additionally, vertebral metastases, which are known to be more responsive to RT than femoral metastases,<sup>5,14</sup> were studied, and overall, higher total radiation doses were used. Furthermore, many patients also received systemic treatment such as chemo- or immunotherapy or used bisphosphonates, which can have an effect on bone remodelling as well. Therefore, it remains difficult to draw hard conclusions on the remineralization ability of RT on femoral metastases. Nevertheless, RT should definitely not be eliminated as a treatment strategy for patients with bone metastases, since there is enough evidence that RT is effective for pain treatment.<sup>15-18</sup> More research is necessary to determine the separate effects of RT and systemic therapies in terms of remineralization. Additionally, it should be investigated whether or not RT improves bone strength of metastatic femurs.

## FE modelling for fracture risk prediction in patients with bone metastases

The primary aim of the abovementioned patient study was to evaluate the FE models on their ability to predict fracture risk in patients with bone metastases. One of the problems with *in vivo* validation is the lack of a gold standard. *In vitro*, one can obtain strength of a femur by experimentally breaking it. *In vivo*, this is obviously not possible. Therefore, we use fractures that occur during follow-up as the ground truth for validation, and, although this might not be the ideal criterion, it is clinically relevant. Since it is difficult to ask patients in the palliative phase of their disease to participate in prospective research,<sup>19</sup> and the incidence of fractures in patients with femoral bone metastases referred for palliative radiotherapy is about 10%, a multicentre approach is required to obtain a fair number of patients to validate the FE model based on fractured and non-fractured femurs. In **Chapter 4**, we created FE models of 39 patients with predominantly lytic lesions, who were followed through their hospital records for six months to determine whether or not they fractured their femur. We analyzed

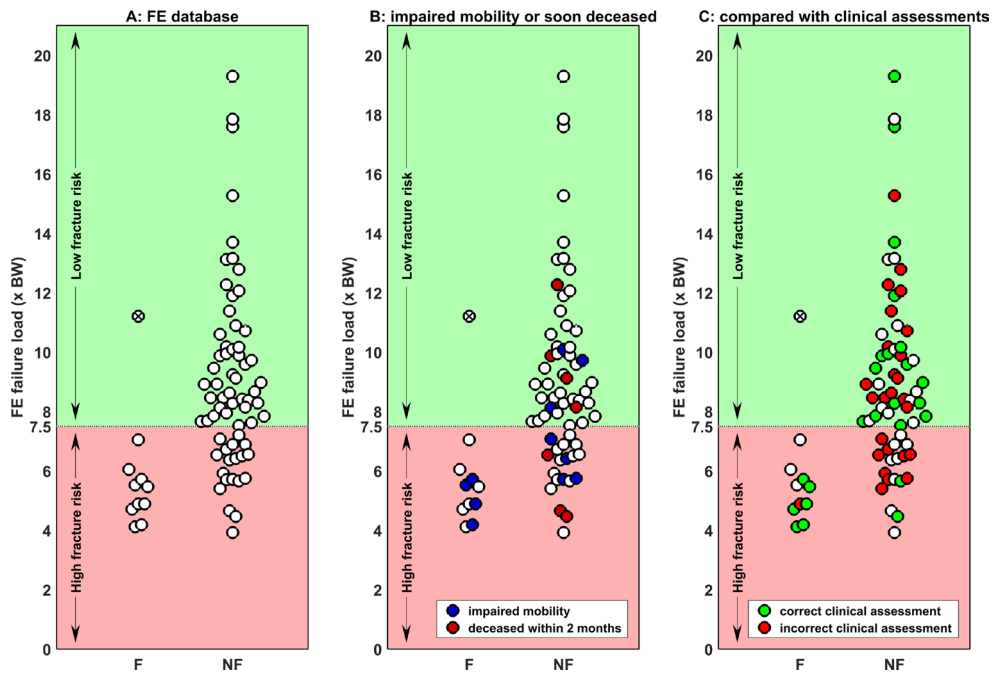
the patients separately for the three institutes due to possible inter-scanner differences, which was at that time brought to our attention via a recent study by Carpenter *et al.*<sup>20</sup> They found differences of up to 2500 N (~23%) when simulating a single-leg stance induced fracture based on CT scans of a healthy subject scanned on two different CT scanners. We additionally asked experienced clinicians to assess the fracture risk based on digitally reconstructed radiographs (DRRs) as a test set-up, since conventional radiographs were not available for all patients. To compare their assessments with the FE model, we defined an institute-specific critical FE failure load classifying a patient to a high or a low fracture risk. We found that the FE model was more accurate than experienced clinicians at identifying patients with a high fracture risk, but slightly less accurate in identifying patients with a low fracture risk.

As we encountered inter-scanner differences in Chapter 4, we investigated the effect of different CT scanners and settings used in the KWF study on a tissue characterizing (Gammex) phantom in **Chapter 5**, and in a more physiological environment by scanning cadaveric femurs in an anatomical body model<sup>21</sup> in **Chapter 6**. We found that substantial differences in Hounsfield units (HU) and BMD existed between different CT scanners, even when using a standardized CT protocol. These inter-scanner differences were also found in the FE models that were generated in Chapter 6. Additionally, changes in reconstruction kernel had a large effect on HU, BMD and FE failure load. Chapters 5 and 6 confirmed the need to correct CT scans obtained from different CT scanners or with different reconstruction kernels before pooling FE results. Therefore, we corrected the CT scans of the KWF femur study by cross-calibrating using Gammex phantom scans obtained in Chapter 5 resulting in improved results.

To create an FE database, we combined the femurs affected with predominantly lytic lesions from the CT femur study ( $n = 16$ ) and KWF femur study ( $n = 60$ , see the patient flowchart in the appendix). During the 6 month follow-up, 11 femurs fractured (Figure 2A). One fractured femur from the KWF study had a failure load corrected for BW over 7.5, and was therefore incorrectly assessed as low fracture risk by the FE model. This patient's femur was affected with non-Hodgkin lymphoma instead of bone metastases caused by a solid tumour, and in fact, not eligible for the study. The patient's femur was irradiated with 30 Gy in 10 fractions and he reported significantly less pain about 2 to 3 months later indicating regression of the lesion. However, six months after RT pain increased, and although the patient was able to walk with a cane, the radiograph showed a neck fracture, probably an insufficiency fracture due to the effect of RT on the lymphoma localisation. The patient was operated on, and the pathology showed no residual tumour. Since this patient was affected with lymphoma instead of bone metastases caused by solid tumours, we decided to exclude him from the final FE database.

As a result, at the time of writing, our FE database consists of a total of 75 femurs, including 10 fractures. Most femurs fractured within 3 months after inclusion, with one outlier that fractured after 5 months. The critical failure load established in Chapter 4 was applied to this patient group, and led to 100% sensitivity of predicting risk of fracture, and a specificity of 71% (Table 1). It should be mentioned that amongst the false positives (the patients that were estimated to have a high fracture risk, but did not fracture their femur within follow-up,  $n=19$ ), three patients died within two months after the CT scan, and five patients had impaired mobility (in a wheelchair or confined to bed, concluded

from questionnaires on pain (BPI<sup>22</sup>), the level of activity and quality of life (parts of LAPAQ<sup>23</sup>, SF-36<sup>24</sup> and WOMAC<sup>25</sup>) within the CT femur study, and on health status (EQ-5D-3L<sup>26,27</sup>) within the KWF femur study, that were filled in during follow-up, Figure 2B). These patients possibly would have fractured their femur when they had lived longer or were more active, although we can evidently never prove that. Positive predictive value was calculated to be 34%, indicating that for one in three femurs with a high estimated fracture risk, prophylactic surgery is indeed necessary. The FE model proved to be very helpful for determining patients that will not fracture their femur, as represented by the negative predictive value of 100%.



**Figure 2:** A) Femoral failure load for all femurs ( $n = 76$ ) in the FE dataset that did (F) or did not (NF) sustain a femoral fracture during follow-up, corrected for body weight (BW). B) Patients with impaired mobility (in a wheelchair or confined to bed, based on questionnaires filled in by the patients during follow-up) are indicated in blue and patients that were deceased within 2 months after the CT scan are indicated in red. C) Failure loads are compared with clinical assessments based on the 30 mm threshold for axial cortical involvement as described by the current clinical guidelines. Correct clinical assessments are indicated in green and incorrect assessments in red. The white dots indicate missing clinical assessments, due to unavailability of pre-treatment conventional radiographs. The crossed dot indicates the patient with the non-Hodgkin lymphoma that was excluded from the final FE database.

In **Chapter 8**, we compared the fracture risk predictions of the FE model with those of the 30 mm threshold for axial cortical involvement measured on conventional radiographs as described in the current clinical guidelines in a subset of 45 patients with 50 affected femurs. Results showed that the



FE model had significantly higher specificity and PPV, and sensitivity and NPV were slightly higher compared to axial cortical involvement measurements performed by experienced medical specialists. If we assume that for the remaining 25 femurs in the FE database, the clinical assessments based on axial cortical involvement would result in the correct fracture risk, the diagnostic accuracy of the FE model would remain superior to the 30 mm threshold for axial cortical involvement (Figure 2C).

**Table 1:** Diagnostic accuracy values of the FE model of all femurs in the FE dataset (n = 75)

	Fractured femurs (n=10)	Not fractured femurs (n=65)	Sensitivity	Specificity	PPV	NPV
<b>FE failure load (x BW)</b>						
≤7.5 (high risk)	10	19	100%	71%	34%	100%
>7.5 (low risk)	0	46				

Other research groups have also put effort in trying to predict fractures in femurs affected with bone metastases with the use of FE modelling. Goodheart *et al.*<sup>28</sup> investigated FE models with different loading conditions on their ability to predict fractures in patients with femoral metastases. Patients were categorized in three groups: five patients fractured their femur within 4 months, 28 patients did not fracture their femur, and 11 patients underwent stabilizing surgery. Mirels' scores were obtained for each of the patients. The Mirels' score is a clinical score that results from a combination of ratings of pain, and lesion type, size and location and ranges between 4 and 12. Generally, if a patient has a Mirels' score of 9 or higher, he is flagged as a high fracture risk patient.<sup>29</sup> Goodheart *et al.* created FE models using three different loading cases: axial compression, level walking, and aggressive stair ascent. For the latter two loading cases, the loads on the head were in a different direction (not entirely axial) and abductor loads were added. FE models simulating level walking showed superior results compared to both axial compression and aggressive stair ascent, with a sensitivity of 80% and a specificity of 85.7%. Our FE model with simple axial loading seems to perform similarly compared to their more sophisticated loading condition simulating level walking. However, it is difficult to compare the results, since they are based on different patient groups. Goodheart *et al.*<sup>28</sup> subsequently combined the FE predictions with Mirels' score. In this case, both the FE model and the Mirels' score had to indicate a fracture for the concerning femur to be classified as high fracture risk. It is generally known that the Mirels' score usually overestimates the fracture risk. In the study by Van der Linden *et al.*,<sup>1</sup> the Mirels' score had a sensitivity of 100%, but a specificity of only 13%, indicating that many patients with a Mirels' score over 9 would not fracture their femur if not preventively stabilized. This is also shown by the 80% sensitivity but 43% specificity and 20% PPV in the study by Goodheart *et al.*<sup>28</sup> However, all four femurs that were incorrectly predicted to have a high fracture risk by the FE model had a Mirels' score below 9, which resulted in 100% specificity, whereas sensitivity remained unchanged. Therefore, they concluded that combining clinical scoring methods with FE results can

even further improve fracture risk predictions. It is difficult to speculate why the Mirels' score was below 9 for the four femurs that were wrongly assessed by their FE model, because they did not report the corresponding femur- and lesion-specific characteristics.

It should be mentioned that Goodheart *et al.*<sup>28</sup> included six blastic lesions in their non-fracture group, but from their study it remains unclear what the estimated strength of these femurs was. In this thesis, we excluded all femurs with predominantly blastic lesions, as we found that FE models of femurs with blastic lesions predicted unrealistically strong bones. As a result, it can be questioned if the relationship between bone density and strength of normal bone is applicable to blastic lesions as well, since these lesions are very radiodense on CT and will therefore have very strong mechanical properties in the FE model, whereas it has been shown that the bone structure within blastic lesions is affected, which may weaken the bone.<sup>30-32</sup>

In another study by Sternheim *et al.*,<sup>33</sup> forty-nine patients with bone metastases in the femur were retrospectively included from a database of patients referred to the orthopaedic surgeon. Five of them had pathological fractures already at referral. Of the other forty-four patients that were referred for stabilizing surgery, eleven declined surgery. Additionally, the authors had a "disease-free" group, existing of twelve femurs of the abovementioned patients that did not show any signs of metastases on the CT scans. They used the strain fold ratio as a threshold to distinguish between high and low fracture risk based on the five fractured femurs and the "disease-free" femurs. This strain fold ratio was defined as the ratio between the absolute maximum principal strain in the metastasis and the typical median strain in the same anatomical region of the "disease-free" femurs. They applied their threshold to the patients that declined surgery, of which none fractured their femur within 5 months, and found a specificity of 63%. Additionally, applying the threshold to the patients that underwent surgery resulted in a specificity of 39%, meaning that 39% of the surgeries may have been unnecessary. However, their study can be criticised for several reasons. First of all, the disease-free femurs on which the threshold was determined may not have been disease-free, as these are contralateral femurs of affected femurs. Additionally, a major limitation lies in the CT scans, which were acquired on many different CT scanners and with different scan and reconstruction protocols, which lacked calibration phantoms. Therefore their results should be interpreted with care.

## Future improvements necessary for clinical implementation

Although the FE model results are very promising, there are some issues that require attention before the FE models can be fully implemented in clinical practice. The first problem is related to the modelling of blastic tissue. We are able to calculate strength of femurs with predominantly lytic lesions. However, patients affected with predominantly blastic lesions were excluded from the FE database, since the femoral bone strength within the blastic lesions was overestimated, probably due to the high degree of mineralization. This resulted in unrealistically strong material properties in the FE model, as bone in blastic lesions lacks structural integrity<sup>30-32</sup> and has shown to have a lower young's modulus,<sup>34</sup> and

therefore generally a decreased structural bone strength. To make the FE models also applicable for patients with blastic lesions, the mechanical material properties of the blastic modelled tissue should be adapted. Although some effort has been put into determining mechanical properties of metastatic bone in the past,<sup>34,35</sup> there has not yet been consensus in this research area. Hipp *et al.*<sup>34</sup> determined apparent densities and mechanical properties of 134 cubic normal, lytic or blastic bone specimens taken from lumbar and thoracic vertebrae from two donors with bone metastases. They found that blastic lesions had significantly higher apparent density compared to normal and lytic specimens, but the young's modulus was lower for both lytic and blastic specimens compared to normal specimens. These results imply that an adapted material model should be used for metastatic tissue. On the other hand, another study by Kaneko *et al.*<sup>35</sup> concluded that relationships between ash density and young's modulus as well as strength in cubic specimens of distal femoral trabecular bone were comparable between metastatic and healthy bone. They used a total of 56 cubic specimens obtained from two femurs with blastic lesions, one femur with mixed lesions and one femur with lytic lesions, as well as eight femurs with no metastases of which four patients died from cancer and four died from another cause than cancer. However, the authors point out that the donors of the different groups vary a lot in age and gender, which could explain the fact that they did not find a difference between metastatic and healthy mechanical properties. Additionally, both above-described studies used a limited number of donors as well as a limited number of specimens. Therefore, to make the FE model suitable for blastic lesions as well, additional experiments should be performed to determine the mechanical material properties of blastic bone tissue in relation with CT densities. Another option would be to adapt the material properties of the blastic regions within the FE models and perform sensitivity analyses to find the optimal blastic material properties resulting in the most accurate fracture risk predictions by the FE model. However, the latter method would require a relatively large amount of fractures that occur in blastic femurs, which is lacking in the current database.

The next challenge lies in detecting the voxels affected by the blastic lesions accurately. Blastic lesions are often of diffuse or permeative character, which makes it difficult to indicate the exact boundaries of such lesions.<sup>36-38</sup> It would be ideal if detection of metastases could be done automatically, for example with the use of deep learning, a machine learning method. Deep learning has shown to be very good at recognizing structures in high-dimensional data, and has been used for many applications such as speech recognition, but also detection, segmentation and recognition of objects and regions in images.<sup>39</sup> The latter purpose makes it potentially useful in the field of (bio)medical sciences. Application of deep learning has been investigated for, for instance, detection of early stage prostate cancer and lymph nodes on histopathological images,<sup>40</sup> breast cancer on mammograms<sup>41</sup> or lung nodules on CT images.<sup>42</sup> Convolutional neural networks (ConvNets) are a specific type of deep learning, which are complex algorithms existing of many mathematical building blocks that are able to recognize certain patterns in, for example, CT images, and can possibly be applicable for detection of metastatic bone lesions. The network will need to be trained using a dataset with CT scans on which each voxel has been classified as metastatic or healthy. This brings us to the main difficulty: metastases must be segmented accurately and reliably, which is known to be very difficult.<sup>36-38</sup> Roth *et al.*<sup>43</sup> have

put some effort in designing ConvNets for the detection of blastic vertebral bone metastases, and managed to obtain an area under the Free-response Receiver Operating Characteristic (FROC) curve of 0.834, which is quite good, keeping in mind that an area of 1 represents a perfect test and an area of 0.5 represents a worthless test. However, there is still room for improvement. Additionally, they only focussed on blastic lesions, while lytic and mixed lesions were not studied. Hence, more work needs to be put into designing and validating a deep learning network that can accurately identify bone metastases. Such a deep learning network can be of great value for the FE models, as well as in clinical practice, for example for radiotherapy planning.

Another matter that requires attention before clinical implementation of the FE model is the limited size of the current database and the need for further validated scientific evidence. Although we have shown in prospective studies that the FE model tends to be more accurate in identification of patients that will and will not fracture their femur within a few months in comparison to clinical guideline, the strength of the evidence is good, but not excellent. A larger patient dataset on which we can validate the threshold would be very important to improve the level of evidence. For this, retrospective databases with CT scans of patients with femoral bone metastases may be used. Essential is that body weight and information on pathological fractures within a few months after CT are known. Currently, the CT scans that are used as input to the FE model require a calibration phantom that is scanned along with the patient to calculate patient-specific BMD from HU. However, this calibration phantom comes with some downsides, such as patients having to lie on top of a separate mattress which can be uncomfortable; it is logistically challenging to scan each patient using the phantom; and the calibration phantom is expensive (about €5000 per phantom). Furthermore, the calibration phantom seems to be sensitive to air between phantom and patient, resulting in inaccurate calibration in such cases. For all these reasons, in **Chapter 7**, we developed a phantomless calibration method based on HU peaks for air, fat and muscle tissue within the CT scan that enables generating FE models based on CT scans without a calibration phantom. It appeared that FE failure loads corrected for body weight were not statistically different when obtained with the air-fat-muscle or phantom calibration. With the use of the phantomless air-fat-muscle calibration method, there is no need for a calibration phantom incorporated in the CT scans, which makes it more likely to find suitable retrospective CT databases. However, prior to application of the phantomless calibration, it needs to be investigated whether the air-fat-muscle calibration yields similar results to the phantom calibration when using other CT scanners or protocols than used in Chapter 7. Also, clinical implementation of the FE model as tool for fracture risk assessment will become easier with the air-fat-muscle calibration.

When enlarging the FE database, it would be very useful to know more about the patients' activity levels. We now validate the model based on fractured and non-fractured femurs but do not take the load bearing as a result of activity level into account. Possibly, some patients do not fracture their femur since they avoid activities which would increase load bearing onto their legs, whereas some patients fracture their femur due to a much more active life style. Nowadays, there are many devices that can measure activity, for example accelerometers or pedometers in smart phones. Such measurements could be of great value for future studies on fracture risk predictions.

Another problem is that generating a fracture risk assessment based on an FE model remains quite time-consuming. Although the process used to take more than a day, presently, the generation of one FE model still takes a few hours. Segmentation and the FE simulation take up most of the time and are steps that are difficult to fully automate in a reliable and robust manner. One of the solutions for a faster fracture risk prediction is to develop a statistical model that can calculate the bone strength based on several characteristics of the bone geometry, density and lesion. For that, the previously mentioned automatic (lesion) detection based on deep learning can be valuable. In that case, the FE model would be no longer necessary. However, to develop and validate such a statistical model is challenging and requires a very large database, which will take a few years before available.

## Clinical implementation

At present, we feel confident to initiate clinical implementation of the developed workflow, since the FE computer models have shown their potential to aid in fracture risk prediction for patients with cancer and predominantly lytic bone metastases. We have received funding from Betaalbaar Beter (a cooperation between the Radboudumc and health insurance company VGZ to aid in implementation of innovations that improve and reduce costs of health care) to initiate implementation in the Radboudumc in 2019. Additionally, the Innovatiefonds Zorgverzekeraars is funding the additional implementation in other radiotherapy institutes via the Landelijk Platform Palliatieve Radiotherapie (LPPR) in 2019 and 2020. All 21 Dutch radiotherapy departments cooperate closely through this LPPR that is part of the Dutch Society of Radiotherapy and Oncology (In Dutch: Nederlandse Vereniging Radiotherapie en Oncologie (NVRO)).

National implementation of our new method is quite challenging. As software is categorized under medical devices, it means that CE (Conformité Européene) certification is needed when used in clinical practice. There are obvious differences in terms of rules and regulations between medical devices used for research purposes and those used for clinical applications. Evidently, one needs to have solid evidence that the tool is reliable and safe in both cases, which has to be approved by an accredited Medical Research Ethics Committees (MREC). In case the medical device is used outside the Radboudumc it should be reported to the Health and Youth Care Inspectorate (Inspectie Gezondheidszorg en Jeugd, IGJ). Additionally, CE certification indicating that the product complies with the applicable EU regulations is obligatory before implementation of medical devices. This CE certification is only necessary for medical devices that are on the market. As long as the FE model is implemented in a research setting, no CE certification is required. As a result, we can build stronger evidence to acquire the CE certification in a later stage, when we feel confident to start distributing the FE model nationally or even internationally.

The current plan is that clinicians can order a fracture risk prediction based on a patient-specific FE model (then called a “Bone Strength score” or “BOS score”), just like they would order a lab-test, comparable to, for example, gene expression profiling for breast or lung cancer. For this, they will send

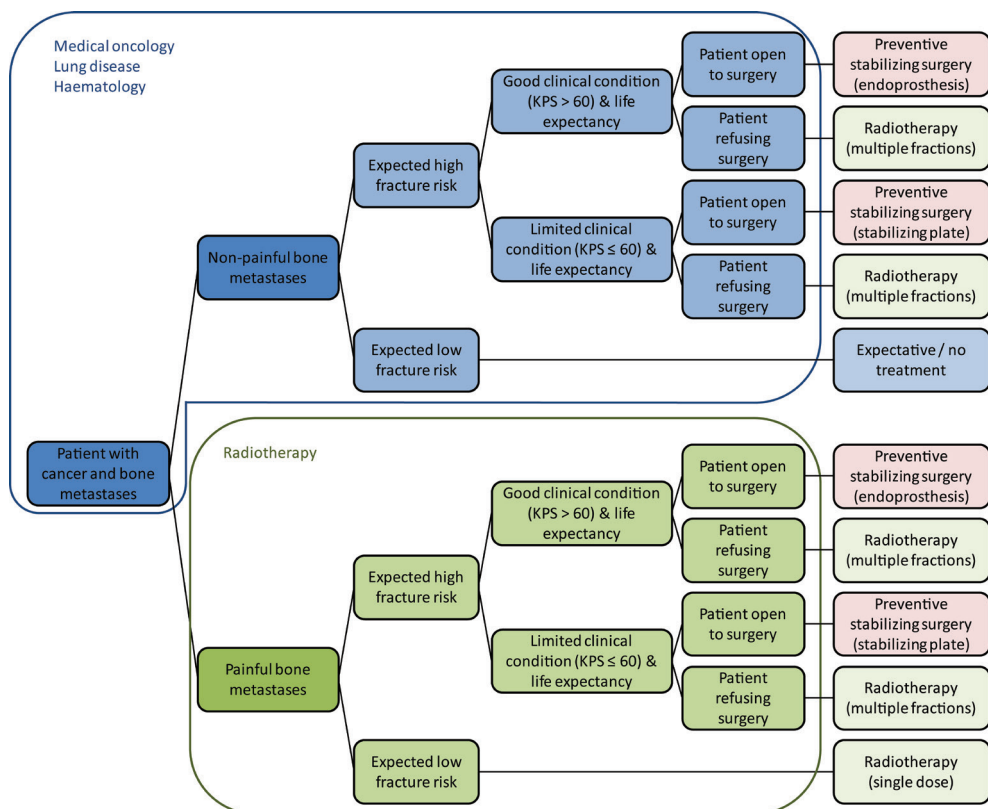
the CT scan to the ORL and we will then create the FE model and formulate a recommendation using the calculated fracture risk of the patient (see Concept of BOS score Report in the supplementary material). We target for a response time of three days, also taking into account the fact that there may be multiple requests pending or that there might occur some unexpected errors that need attention and delay the process. Medical specialists have indicated that a delivery time of three days is acceptable. Subsequently, the BOS score can be used in the decision making process for determining the best treatment for the patient. The BOS score will be a complementary tool to the clinical guidelines, as the clinician also takes patient-related clinical health factors, such as life expectancy, performance status, activity level, and pain, into account. Moreover, the BOS score can be used in the shared decision making, a process in which clinician and patient discuss the treatment options to take also the patient's wishes into account, by supporting the fracture risk assessment with an objective biomechanical model. During the implementation phase, the FE database will be expanded as patients of whom BOS scores are created will be followed over time to obtain fracture data.

After successful pilot implementation in the first radiotherapy institutes, further deployment needs to be achieved. In the Netherlands, national clinical guidelines are constructed to align local protocols and backup individual treatment decisions. When necessary these guidelines are updated. Recognition of the BOS score by the LPPR is a prerequisite to implement the use of the BOS score in clinical practice. The Dutch radiotherapy departments intend to implement the BOS score, spread the study results and add the use of the BOS score to the existing treatment protocols and guidelines, in case of successful application. Most patients with femoral metastases, who are initially seen at departments such as medical oncology, are referred to radiotherapy departments to undergo RT for pain as local treatment (Figure 3). As a result, most fracture risk predictions are done by radiation oncologists. Hence, if the radiation oncologists adopt the BOS score, then medical oncologists and surgeons will probably also start using the BOS score for fracture risk assessments.

## Concluding remarks

FE modelling is a highly potential tool for fracture risk assessment in patients with cancer and bone metastases. Currently, we are investigating the possibility of implementing the FE computer models in daily clinical practice. We have made progress in improving the FE model to facilitate clinical implementation, such as inter-scanner corrections and phantomless calibration methods. However, certain issues remain that need to be overcome before FE modelling can be widespread implemented in clinical practice. The FE model is not yet useable for blastic lesions, and the reliability of the FE predictions can be further increased by extending the patient FE database. This database extension will follow during clinical implementation, for which FE models will be created for many patients. The FE model to assess fracture risk assessment for cancer patients with lytic femoral bone lesions can support the decision making process in choosing the most suited local treatment option, and

subsequently improve quality of life by preventing pathological fracturing, and, decreasing surgical over and under treatment.



**Figure 3:** Flowchart for patients with cancer and bone metastases. Patients will initially come and see e.g. a medical oncologist, pulmonologist or haematologist, depending on their primary tumour. In case of painful bone metastases, the patient will be referred to a radiation oncologist. At the radiotherapy departments as well as the medical oncology, lung and haematology departments, fracture risk is assessed. Based on the fracture risk assessment (using either the FE model or an assessment of the conventional radiograph based on axial cortical involvement) as well as clinical condition (using Karnofsky Performance Status, KPS<sup>44</sup>), life expectancy (using for example the OPTIModel prediction model for survival<sup>45</sup>) and the patients' wishes, patients are referred to the orthopaedic or general surgeon, will be treated with radiotherapy or will not receive any specific treatment for their bone metastases until the patient's situation changes (e.g. disease progression).

## References

1. Van der Linden YM, Dijkstra PD, Kroon HM, Lok JJ, Noordijk EM, Leer JW, Marijnen CA. 2004. Comparative analysis of risk factors for pathological fracture with femoral metastases. *J Bone Joint Surg Br* 86:566-573.
2. Van der Linden YM, Kroon HM, Dijkstra SP, Lok JJ, Noordijk EM, Leer JW, Marijnen CA. 2003. Simple radiographic parameter predicts fracturing in metastatic femoral bone lesions: results from a randomised trial. *Radiother Oncol* 69:21-31.
3. Derikx LC, van Aken JB, Janssen D, Snyers A, van der Linden YM, Verdonschot N, Tanck E. 2012. The assessment of the risk of fracture in femora with metastatic lesions: comparing case-specific finite element analyses with predictions by clinical experts. *J Bone Joint Surg Br* 94:1135-1142.
4. Tanck E, van Aken JB, van der Linden YM, Schreuder HW, Binkowski M, Huizenga H, Verdonschot N. 2009. Pathological fracture prediction in patients with metastatic lesions can be improved with quantitative computed tomography based computer models. *Bone* 45:777-783.
5. Koswig S, Budach V. 1999. Remineralisation und Schmerzlinderung von Knochenmetastasen nach unterschiedlich fraktionierter Strahlentherapie (10 mal 3 Gy vs. 1 mal 8 Gy). Eine prospektive Studie. *Strahlenther Onkol* 175:500-508.
6. Chow E, Holden L, Rubenstein J, Christakis M, Sixel K, Vidmar M, Finkelstein J, Hayter C, Loblaw A, Wong R, Szumacher E, Danjoux C. 2004. Computed tomography (CT) evaluation of breast cancer patients with osteolytic bone metastases undergoing palliative radiotherapy—a feasibility study. *Radiother Oncol* 70:291-294.
7. Reinbold WD, Wannenmacher M, Hodapp N, Adler CP. 1989. Osteodensitometry of vertebral metastases after radiotherapy using quantitative computed tomography. *Skeletal Radiol* 18:517-521.
8. Foerster R, Eisele C, Bruckner T, Bostel T, Schlamp I, Wolf R, Debus J, Rief H. 2015. Bone density as a marker for local response to radiotherapy of spinal bone metastases in women with breast cancer: a retrospective analysis. *Radiat Oncol* 10:368.
9. Harada H, Katagiri H, Kamata M, Yoshioka Y, Asakura H, Hashimoto T, Furutani K, Takahashi M, Sakahara H, Nishimura T. 2010. Radiological response and clinical outcome in patients with femoral bone metastases after radiotherapy. *J Radiat Res* 51:131-136.
10. Groenen KH, Pouw MH, Hannink G, Hosman AJ, van der Linden YM, Verdonschot N, Tanck E. 2016. The effect of radiotherapy, and radiotherapy combined with bisphosphonates or RANK ligand inhibitors on bone quality in bone metastases. A systematic review. *Radiother Oncol* 119:194-201.
11. Sprave T, Hees K, Bruckner T, Foerster R, Bostel T, Schlamp I, El Shafie R, Nicolay NH, Debus J, Rief H. 2018. The influence of fractionated radiotherapy on the stability of spinal bone metastases: a retrospective analysis from 1047 cases. *Radiat Oncol* 13:134.
12. Sprave T, Verma V, Forster R, Schlamp I, Hees K, Bruckner T, Bostel T, El Shafie RA, Welzel T, Nicolay NH, Debus J, Rief H. 2018. Local response and pathologic fractures following stereotactic body radiotherapy versus three-dimensional conformal radiotherapy for spinal metastases - a randomized controlled trial. *BMC cancer* 18:859.
13. Sprave T, Verma V, Forster R, Schlamp I, Hees K, Bruckner T, Bostel T, El Shafie RA, Welzel T, Nicolay NH, Debus J, Rief H. 2018. Bone density and pain response following intensity-modulated



- radiotherapy versus three-dimensional conformal radiotherapy for vertebral metastases - secondary results of a randomized trial. *Radiat Oncol* 13:212.
14. Rieden K, Adolph J, Lellig U, zum Winkel K. 1989. [The radiotherapeutic effect on bone metastases in relation to the frequency of metastases, sites of metastases and histology of the primary tumor]. *Strahlenther Onkol* 165:380-385.
  15. Chow E, Zeng L, Salvo N, Dennis K, Tsao M, Lutz S. 2012. Update on the systematic review of palliative radiotherapy trials for bone metastases. *Clin Oncol* 24:112-124.
  16. Lutz S, Berk L, Chang E, Chow E, Hahn C, Hoskin P, Howell D, Konski A, Kachnic L, Lo S, Sahgal A, Silverman L, von Gunten C, Mendel E, Vassil A, Bruner DW, Hartsell W, American Society for Radiation O. 2011. Palliative radiotherapy for bone metastases: an ASTRO evidence-based guideline. *Int J Radiat Oncol Biol Phys* 79:965-976.
  17. Bedard G, Hoskin P, Chow E. 2014. Overall response rates to radiation therapy for patients with painful uncomplicated bone metastases undergoing initial treatment and retreatment. *Radiother Oncol* 112:125-127.
  18. van der Linden Y, Roos D, Lutz S, Fairchild A. 2009. International variations in radiotherapy fractionation for bone metastases: geographic borders define practice patterns? *Clin Oncol* 21:655-658.
  19. ZonMw. April 2014. Succesvol includeren in de palliatieve zorg.
  20. Carpenter RD, Saeed I, Bonaretti S, Schreck C, Keyak JH, Streeper T, Harris TB, Lang TF. 2014. 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* 36:1225-1232.
  21. Tanck E, Deenen JC, Huisman HJ, Kooloos JG, Huizenga H, Verdonchot N. 2010. An anatomically shaped lower body model for CT scanning of cadaver femurs. *Phys Med Biol* 55:N57-62.
  22. Fairbank JC, Couper J, Davies JB, O'Brien JP. 1980. The Oswestry low back pain disability questionnaire. *Physiotherapy* 66:271-273.
  23. Stel VS, Smit JH, Pluijm SM, Visser M, Deeg DJ, Lips P. 2004. Comparison of the LASA Physical Activity Questionnaire with a 7-day diary and pedometer. *J Clin Epidemiol* 57:252-258.
  24. Ware JE, Jr., Sherbourne CD. 1992. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 30:473-483.
  25. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. 1988. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol* 15:1833-1840.
  26. Williams A. 1990. Euroqol - a New Facility for the Measurement of Health-Related Quality-of-Life. *Health Policy* 16:199-208.
  27. Brooks R. 1996. EuroQol: The current state of play. *Health Policy* 37:53-72.
  28. Goodheart JR, Cleary RJ, Damron TA, Mann KA. 2015. Simulating activities of daily living with finite element analysis improves fracture prediction for patients with metastatic femoral lesions. *J Orthop Res* 33:1226-1234.
  29. Mirels H. 1989. Metastatic disease in long bones. A proposed scoring system for diagnosing impending pathologic fractures. *Clin Orthop*:256-264.
  30. Healey JH, Brown HK. 2000. Complications of bone metastases: surgical management. *Cancer* 88:2940-2951.

31. Sone T, Tamada T, Jo Y, Miyoshi H, Fukunaga M. 2004. Analysis of three-dimensional microarchitecture and degree of mineralization in bone metastases from prostate cancer using synchrotron microcomputed tomography. *Bone* 35:432-438.
32. Saylor PJ, Smith MR. 2010. Bone health and prostate cancer. *Prostate cancer and prostatic diseases* 13:20-27.
33. Sternheim A, Giladi O, Gortzak Y, Drexler M, Salai M, Trabelsi N, Milgrom C, Yosibash Z. 2018. Pathological fracture risk assessment in patients with femoral metastases using CT-based finite element methods. A retrospective clinical study. *Bone* 110:215-220.
34. Hipp JA, Rosenberg AE, Hayes WC. 1992. Mechanical properties of trabecular bone within and adjacent to osseous metastases. *J Bone Miner Res* 7:1165-1171.
35. Kaneko TS, Bell JS, Pejicic MR, Tehranzadeh J, Keyak JH. 2004. Mechanical properties, density and quantitative CT scan data of trabecular bone with and without metastases. *J Biomech* 37:523-530.
36. Dijkstra PD, Oudkerk M, Wiggers T. 1997. Prediction of pathological subtrochanteric fractures due to metastatic lesions. *Arch Orthop Trauma Surg* 116:221-224.
37. Keene JS, Sellinger DS, McBeath AA, Engber WD. 1986. Metastatic breast cancer in the femur. A search for the lesion at risk of fracture. *Clin Orthop Relat Res*:282-288.
38. Tatar Z, Soubrier M, Dillies AF, Verrelle P, Boisgard S, Lapeyre M. 2014. Assessment of the risk factors for impending fractures following radiotherapy for long bone metastases using CT scan-based virtual simulation: a retrospective study. *Radiat Oncol* 9.
39. LeCun Y, Bengio Y, Hinton G. 2015. Deep learning. *Nature* 521:436-444.
40. Litjens G, Sanchez CI, Timofeeva N, Hermesen M, Nagtegaal I, Kovacs I, Hulsbergen-van de Kaa C, Bult P, van Ginneken B, van der Laak J. 2016. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Sci Rep-Uk* 6.
41. Ribli D, Horvath A, Unger Z, Pollner P, Csabai I. 2018. Detecting and classifying lesions in mammograms with Deep Learning. *Sci Rep-Uk* 8.
42. Setio AAA, Ciompi F, Litjens G, Gerke P, Jacobs C, van Riel SJ, Wille MMW, Naqibullah M, Sanchez CI, van Ginneken B. 2016. Pulmonary Nodule Detection in CT Images: False Positive Reduction Using Multi-View Convolutional Networks. *IEEE Trans Med Imaging* 35:1160-1169.
43. Roth HR, Lu L, Liu J, Yao J, Seff A, Cherry K, Kim L, Summers RM. 2016. Improving Computer-Aided Detection Using Convolutional Neural Networks and Random View Aggregation. *IEEE Trans Med Imaging* 35:1170-1181.
44. Karnofsky D, Burchenal J. 1949. The clinical evaluation of chemotherapeutic agents in cancer. In: MacLeod C editor. *Evaluation of Chemotherapeutic Agents*. New York: Columbia University Press; pp. 191-205.
45. Willeumier JJ, van der Linden YM, van der Wal CWPG, Jutte PC, van der Velden JM, Smolle MA, van der Zwaal P, Koper P, Bakri L, de Pree I, Leithner A, Fiocco M, Dijkstra PDS. 2018. An Easy-to-Use Prognostic Model for Survival Estimation for Patients with Symptomatic Long Bone Metastases. *Journal of Bone and Joint Surgery-American Volume* 100:196-204.

## Supplementary material: Concept of BOS score Report

BOS score

**Radboudumc**  
university medical center

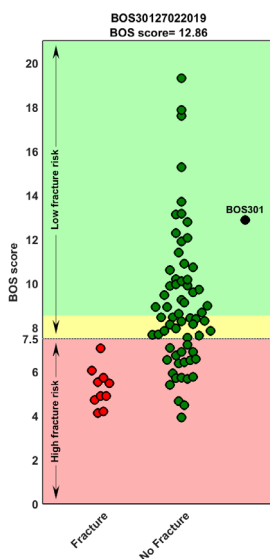
### BOS score Report – BOS30127022019 for fracture risk prediction in femoral bone metastases

In the figure below, you can find the calculated BOS score of patient BOS30127022019 relative to the patients in the BOS database.

As you can see, patient BOS30127022019 has a BOS score of 12.86, which is higher than the threshold of 7.5, indicating a low fracture risk.

#### Conclusion:

Patient BOS30127022019 has a low fracture risk.



*Additional results (weakest location of the bone and the value of the BOS score) can be found on the next page*

#### Terms and conditions

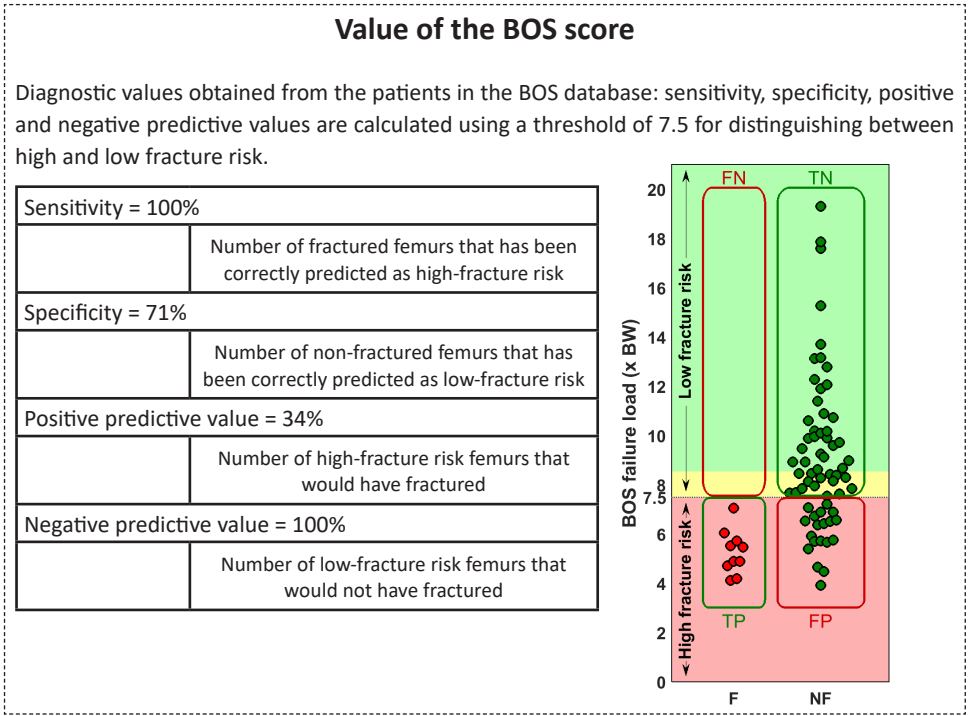
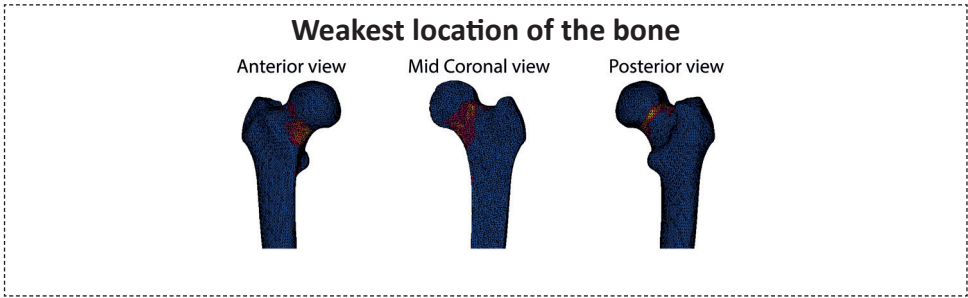
No information provided in this report will give any guarantees. It is explicitly the responsibility of the physician to use and interpret the outcomes from this report correctly. Health care providers should always also exercise their own independent clinical judgement when using the BOS score in conjunction with patient care.

27 February 2019  
page 1 of 2

**Contact**  
orthopaedicresearchlab@radboudumc.nl  
024 36 13366

BOS score

### Additional results – BOS30127022019

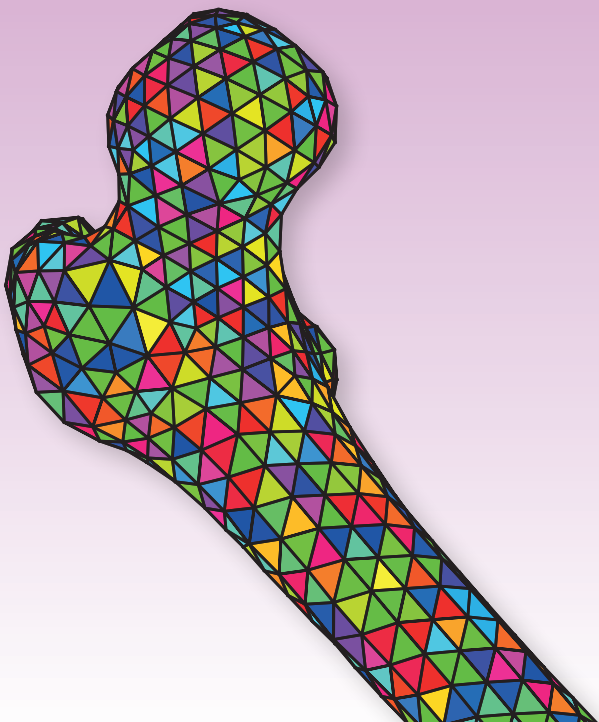


**Terms and conditions**

No information provided in this report will give any guarantees. It is explicitly the responsibility of the physician to use and interpret the outcomes from this report correctly. Health care providers should always also exercise their own independent clinical judgement when using the BOS score in conjunction with patient care.

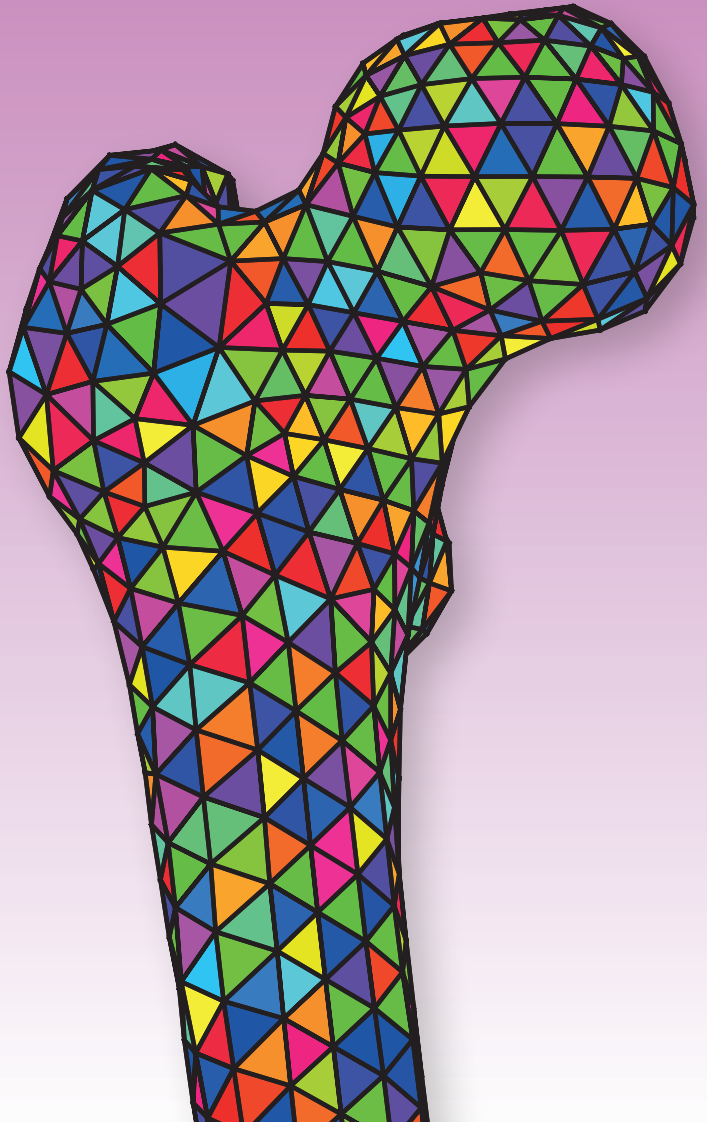
27 February 2019  
page 2 of 2

**Contact**  
orthopaedicresearchlab@radboudumc.nl  
024 36 13366



# CHAPTER 10

## SUMMARY





Patients with advanced cancer often develop bone metastases, and in approximately ten percent, these lesions occur in the femur. Femoral metastases may cause pain and can lead to pathological fractures, which severely affect the quality of life. When the femur fractures, patients instantly lose their mobility and necessary surgical treatment of such fractures is often complex with postoperative complications. As a result, their survival is decreased. Local treatment of patients with femoral metastases is therefore based on the expected fracture risk: patients with an expected low fracture risk are treated conservatively with non-invasive radiotherapy (RT) to decrease pain, whereas patients with an expected high fracture risk are considered for stabilizing surgery to prevent a fracture from occurring. In case a patient has a high fracture risk, but insufficient clinical condition to undergo surgery, or refuses surgery, radiotherapy is used to hopefully restore bone strength through remineralization. Therefore, accurate fracture risk prediction is important. However, the current clinical guidelines for fracture risk prediction are limited in sensitivity and specificity. As a result, it is difficult for clinicians to distinguish between high and low fracture risk lesions, leading to considerable numbers of under and over treatment. Patient-specific finite element (FE) computer models have the potential to improve fracture risk assessments. These FE models calculate bone strength, based on patient-specific anatomy and bone quality, obtained from quantitative CT scans, and have shown to be promising in experimental settings. The goal of this thesis was to further develop and validate the patient-specific FE model to predict fracture risk in patients with cancer and bone metastases. For this purpose, we performed two consecutive patient studies: the CT femur study (2006-2009, funded by the Dutch Science Foundation NWO-STW (NPG.06778)) and KWF femur study (2015-2017, funded by the Dutch Cancer Society (KUN 2012-5591)) (see appendix: flowchart of the patients used in all studies of this thesis), and prepared the workflow for large-scale clinical implementation. This chapter summarizes the key findings of this thesis.

To underline the need for a better fracture risk prediction tool, we initiated in **Chapter 2** with the evaluation of the current Dutch clinical guideline that states that femurs with an axial cortical involvement of more than 30 mm measured on conventional radiographs are at risk of pathological fracture and should be prophylactically stabilized. For this, 110 lesions of 100 patients were included from both the CT femur study and KWF femur study, of which 14 femurs fractured. Fracture risk was assessed by three observers (a radiation oncologist, a radiologist and an orthopaedic surgeon) separately by relying on their expert opinion and by measuring the lesion dimensions on conventional radiographs acquired in the two months prior to radiotherapy treatment. If observers had to estimate fracture risk based on expert opinion, sensitivity was 43%, specificity was 78%, and positive (PPV) and negative predictive values (NPV) were 22% and 90%, respectively. When the 30 mm threshold was applied, sensitivity was 86%, specificity 50%, PPV 20% and NPV 96%. It should be mentioned that observers did not always agree on whether or not a lesion was larger than 30 mm (Spearman correlations ranging between 0.44 and 0.57). We concluded that the 30 mm threshold for axial cortical involvement should be applied in all clinical decision making, until a more reliable fracture risk assessment method is available. Nevertheless, a more accurate tool for fracture risk assessment



in patients with cancer and femoral bone metastases is needed, since the assessment of fracture risk using measurements on conventional radiographs or CT scans remains challenging.

To determine the ability of FE models on differentiating between patients that would and would not fracture their femur, we used the CT femur patient study. In this prospective cohort study, we included 66 patients that were treated with single (SF,  $1 \times 8$  Gy) or multiple (MF, 5 or  $6 \times 4$  Gy) fraction radiotherapy for painful femoral bone metastases in three radiotherapy institutes in the Netherlands, and were followed for six months after inclusion. Patients underwent CT scans at baseline and at 1, 4 and 10 weeks after radiotherapy. Patients were scanned on top of a calibration phantom, which enabled us to convert Hounsfield units (HU) within the CT scan to bone mineral density (BMD). Firstly, we used the CT femur patient study to investigate the effect of SF and MF RT on BMD in the radiation field and within metastatic lesions, for lytic, blastic, and mixed lesions separately (**Chapter 3**). For this, we included the 42 patients who underwent more than one CT scan. We determined mean BMD on each CT scan for each proximal femur and in greater detail for a region of interest that contained the metastatic lesion. Ten weeks after palliative radiotherapy in patients with femoral metastatic lesions, a limited increase in BMD was seen with no beneficial effect of MF over SF RT. BMD in lytic lesions was unchanged but slightly increased in mixed and blastic lesions. Whether the latter implied progression of the disease or remineralization was unclear, and the subsequent clinical effect on femoral bone strength has to be investigated in the future.

In **Chapter 4**, the fracture risk assessment ability of the FE models was studied using the CT femur study. In that study, seven patients fractured a total of nine femurs during the follow-up of six months. We created FE models of 39 patients (47 femurs) with predominantly lytic metastases and calculated failure loads. We determined whether the FE models were able to distinguish between patients that did and did not fracture their femur during follow-up. Due to inter-scanner differences, patients were analyzed separately for the three institutes. We found that the median failure load was significantly lower for patients who sustained a fracture than for patients with no fractures in one of the institutes. In the two other institutes, the number of patients with a fracture was too low to make a clear distinction. Fracture locations were well predicted by the FE model when compared with post-fracture radiographs. The FE model was more accurate in identifying patients with a high fracture risk compared with experienced clinicians that assessed fracture risk on digitally reconstructed radiographs (DRRs), with a sensitivity of 89% for the FE models versus 0% to 33% for the clinical assessments. Specificity was 79% for the FE models versus 84% to 95% for clinical assessments. The clinicians indicated that the DRRs were limited in image quality in comparison to the conventional radiographs they use for the fracture risk assessments in daily clinical practice. We concluded that the FE models can be a valuable tool to improve clinical fracture risk predictions in metastatic bone disease, but future work in a larger patient population should confirm the higher predictive power of FE models compared with current clinical guidelines.

For the KWF femur study, for which 90 patients with femoral bone metastases treated with radiotherapy in four Dutch radiotherapy institutes were included, we quantified the differences between CT scanners and CT protocols, since this was one of the main problems we encountered in Chapter 4. In **Chapter 5**,

we therefore investigated the effect of different CT scanners, reconstruction protocols, scan positions, and air gaps on HU and/or BMD obtained with a calibration phantom, using tissue characterizing phantoms with known densities. By using phantoms, the set up was very reproducible on the different CT scanners. These phantoms, containing bone-like materials, were scanned on the four CT scanners. We found considerable differences between CT scanners in HU, which decreased after calibration to BMD. Additionally, we concluded that different reconstruction kernels and changes in position within the field of view (FOV) as well as air gaps should be avoided. In a subsequent study (**Chapter 6**), we aimed to quantify these effects on HU, BMD, and FE failure load of cadaveric femurs. For this purpose, we included six cadaveric femurs that were placed in an anatomical body model to mimic the lower body of a patient and scanned on top of a calibration phantom on the four CT scanners. Scans were made with the standardized protocol of the patient study, as well as with changes in slice thickness, FOV and reconstruction kernel. An additional scan was made with an air gap between the body model and calibration phantom. HU and calibrated BMD were determined in cortical and trabecular regions of interest, and FE models were constructed to calculate failure load. Again, we showed that there were differences between CT scanners up to an average difference of 17% in FE failure load. Furthermore, we found considerable effects of different reconstruction kernels with a largest average effect of 9% in FE failure load. Air between the body model and calibration phantom only had a slight effect. Chapters 5 and 6 showed that one should be aware of potential differences in CT values and subsequent FE outcomes when performing multicentre studies using different CT scanners or CT settings, mainly with various reconstruction kernels, which requires correction before pooling FE results.

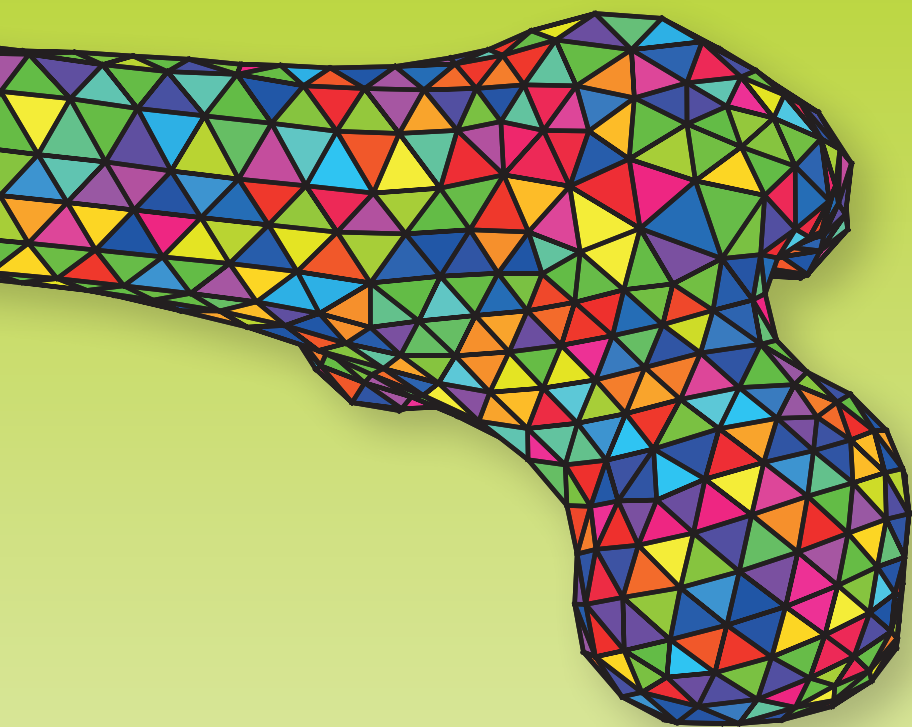
To enhance FE modelling for patients with bone metastases using everyday clinical CT scans without calibration phantoms, we investigated the possibility of phantomless calibration in **Chapter 7**. From both patient studies, we included 67 femurs and constructed FE models based on the phantom calibration and on two phantomless calibration methods: the “air-fat-muscle” and the “non-patient-specific” calibration. For air-fat-muscle calibration, a histogram of the HU in a standardized region of interest including the tissue of the patient’s right leg and some surrounding air was created to subtract peaks for air, fat and muscle tissue. These CT peaks were linearly fitted to fixed “BMD” values of the corresponding tissues to obtain a calibration function. For non-patient-specific calibration, the average phantom calibration function of 26 patients from one of the four institutes was applied to all patients. FE failure loads were determined and compared between phantom and phantomless calibration methods. We found that failure loads based on both phantomless calibration methods were highly correlated to those based on the phantom calibration. Although differences in failure loads between phantom and both air-fat-muscle and non-patient-specific calibrations were small, we advised to use the air-fat-muscle calibration method, as non-patient-specific calibration resulted in deviating failure loads in case of CT scanning with a different reconstruction kernel. With this air-fat-muscle calibration, clinical implementation of the FE model as a tool for fracture risk assessment will be easier, since FE models can be made for each patient that presents with femoral bone metastases, even if there is no calibration phantom available.

In **Chapter 8**, we compared the fracture risk prediction ability of the FE models with the current Dutch clinical guideline that describes that fracture risk should be assessed using the 30 mm threshold for axial cortical involvement measured on plain radiographs. For this, we combined the clinical assessments from Chapter 2 with the FE database we built from a subset of Chapter 4 (CT femur study) and the KWF femur study, and compared diagnostic accuracy (sensitivity, specificity, and positive and negative predictive values) of both methods for 50 femurs (45 patients). We found that the FE models were better at predicting fracture risk in comparison to the clinical guideline: fracture risk was correctly assessed for more femurs (sensitivity of 100% vs. 86%, specificity of 74% vs. 42%, PPV of 39% vs. 19%, and NPV of 100% vs. 95%, for the FE model vs. clinical guideline, respectively).

Finally, we reflected on the work described in this thesis and discussed the future challenges to further improve the FE models and implement the FE models into clinical practice in **Chapter 9**. As the FE models are able to aid in fracture risk assessment, we intend to commence with implementation of the FE models in clinical practice. The idea is that, rather than using this tool by themselves, a clinician would order an FE model calculation (then called “BOne Strength score” or “BOS score”), just like he or she would order a lab-test, comparable to, for example, gene expression profiling for breast or lung cancer. A patient-specific FE model will be constructed and an advice will be returned to the clinician about the expected fracture risk of the patient. Subsequently, the clinician can include this information in his consultation with the patient to choose the most optimal intervention. In this way, the FE model prediction has its value in shared decision making; the process in which clinician and patient discuss the treatment options to take the patient’s wishes into account. A quantitative assessment of the fracture risk will aid the clinician and patient to make the best judgement of the treatment path.

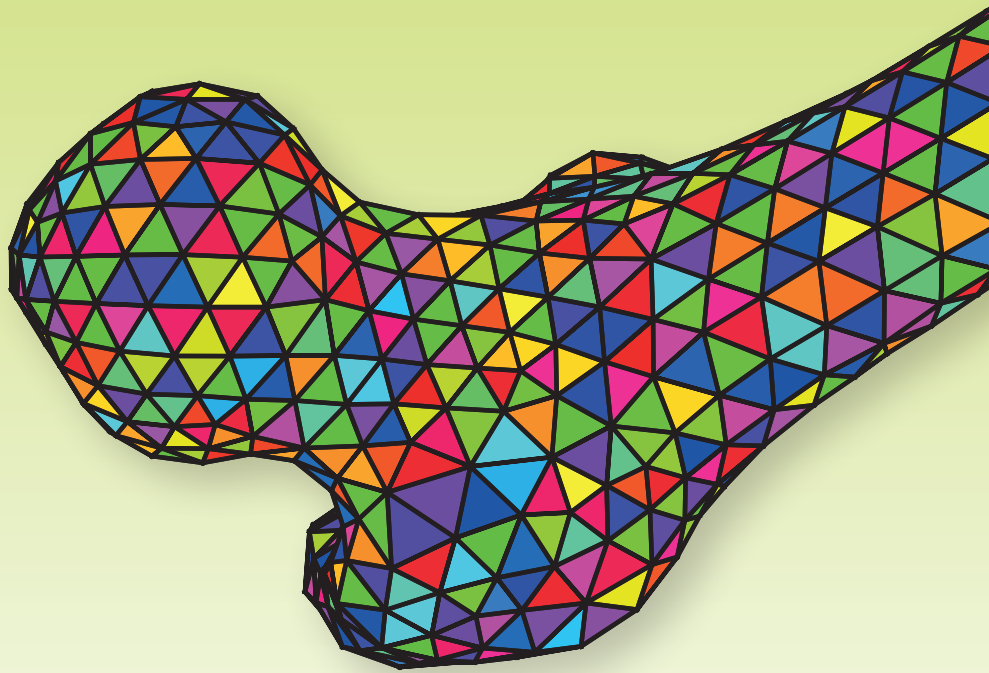
Thus with this thesis and the work that lies ahead, we intend to decrease the occurrence of pathological fracturing in patients with advanced cancer and femoral bone metastases and lower overtreatment by preventing unnecessary surgical procedures. With that, we hope to contribute to the quality of life of these patients. In the meantime, we intend to enlarge the FE database to further increase the reliability of the FE computer model as a fracture risk prediction tool. This can be done prospectively during implementation using the patients for which the FE fracture risk predictions are requested, or retrospectively by using clinical CT databases.





# CHAPTER 11

SAMENVATTING





Patiënten met uitgezaaide kanker krijgen vaak uitzaaiingen in het bot (botmetastasen). Deze botmetastasen bevinden zich in ongeveer tien procent van de gevallen in het dijbeen, ook wel het femur genoemd. Botmetastasen in het femur kunnen pijn veroorzaken en leiden soms tot zogenaamde pathologische fractures, waardoor lopen vaak acuut niet meer mogelijk is. Zo'n heftige gebeurtenis beïnvloedt de kwaliteit van leven van patiënten fors. Door een dergelijke fractuur in het femur is de patiënt niet meer mobiel en zal hij een operatie moeten ondergaan om het been te stabiliseren. Een dergelijke spoedoperatie is vaak complex en kan zorgen voor postoperatieve complicaties met erna ook een verhoogd risico op overlijden. Locale behandeling van patiënten met botmetastasen in het intacte femur is afhankelijk van het fractuurrisico: patiënten met een laag fractuurrisico worden conservatief behandeld met niet-invasieve radiotherapie om pijn te bestrijden, terwijl patiënten met een hoog fractuurrisico in aanmerking komen voor een preventieve stabiliserende operatie om een pathologische fractuur te voorkomen. In het geval dat een patiënt een hoog fractuurrisico heeft, maar zijn klinische conditie niet goed genoeg is om een dergelijke operatie te ondergaan of hij de operatie weigert, wordt hij behandeld met radiotherapie om zo hopelijk de botsterkte te verbeteren doordat in de maanden na radiotherapie mogelijk remineralisatie optreedt. Al met al is het dus belangrijk om het fractuurrisico goed in te schatten. Inschatting gebeurt momenteel door de lengte van de axiale corticale aantasting in het femur te bepalen. Echter is de sensitiviteit en specificiteit van de huidige klinische richtlijnen voor het voorspellen van fractuurrisico beperkt. Daardoor is het voor artsen moeilijk om onderscheid te maken tussen laesies die een hoog- en laag fractuurrisico veroorzaken, waardoor er een behoorlijk aantal patiënten wordt onderbehandeld (radiotherapie voor pijn en dan alsnog een pathologische fractuur) en overbehandeld (preventieve operatie terwijl dat eigenlijk niet nodig is).

Patiëntspecifieke eindige-elementenmodellen (EE-modellen) zijn computermodellen en kunnen mogelijk de fractuurrisicovoorspellingen verbeteren. Deze EE-modellen berekenen botsterkte op basis van de patiëntspecifieke anatomie en botkwaliteit verkregen uit kwantitatieve CT-scans, en hebben eerder al in een experimentele setting laten zien dat ze van toegevoegde waarde te kunnen zijn. Het doel van dit proefschrift was om het patiëntspecifieke EE-model voor fractuurrisicovoorspelling bij patiënten met kanker en botmetastasen verder te ontwikkelen en te valideren. Hiervoor hebben we twee achtereenvolgende patiëntenstudies uitgevoerd: de CT-femurstudie (2006-2009, gesubsidieerd door de Nederlandse Organisatie voor Wetenschappelijk Onderzoek - Stichting voor de Technische Wetenschappen NWO-STW (NPG.06778)) en de KWF-femurstudie (2015-2017, gesubsidieerd door KWF Kankerbestrijding (KUN 2012-5591)) (zie appendix: flowchart of the patients used in all studies of this thesis). Daarnaast hebben we voorbereidingen getroffen voor klinische implementatie op grotere schaal. Dit hoofdstuk vat de belangrijkste resultaten van dit proefschrift samen.

Om te laten zien dat er een betere methode voor de voorspelling van fractuurrisico nodig is, begonnen we in **Hoofdstuk 2** met de validatie van de huidige Nederlandse klinische behandelrichtlijnen. In deze richtlijnen staat dat femora met een axiale corticale aantasting van meer dan 30 mm gemeten op röntgenfoto's een verhoogd fractuurrisico hebben en daarom preventieve stabilisatie de voorkeur heeft. Voor de validatie werden 110 femora van patiënten uit de CT-femurstudie en de KWF-



femurstudie geïnccludeerd, waarvan 14 femora een fractuur hadden opgelopen. Fractuurrisico werd bepaald door drie waarnemers (een radiotherapeut-oncoloog, een radioloog en een orthopedisch chirurg) aan de hand van hun klinische ervaring en door de laesiedimensies te meten op conventionele röntgenfoto's die niet meer dan twee maanden voor de radiotherapie gemaakt waren. Als artsen het fractuurrisico schatten aan de hand van hun klinische ervaring was de sensitiviteit 43%, de specificiteit 78% en de positief (PPV) en negatief voorspellende waarde (NPV) respectievelijk 22% en 90%. Als de 30 mm-regel werd toegepast, was de sensitiviteit 86%, specificiteit 50%, PPV 20% en NPV 96. Er moet wel bij gezegd worden dat artsen het onderling niet altijd eens waren of een laesie wel of niet groter dan 30 mm was (Spearman correlaties tussen 0.44 en 0.57). We concludeerden dat de 30 mm-regel gebruikt zou moeten worden in klinische besluitvorming tot er een betrouwbaardere methode voor fractuurrisicovoorspelling beschikbaar is. Desalniettemin is er een betere methode nodig om het fractuurrisico voor patiënten met kanker en botmetastasen in het femur te voorspellen.

Om uit te zoeken of EE-modellen onderscheid kunnen maken tussen patiënten die wel en geen femurfractuur oplopen, hebben we patiëntdata uit de CT-femurstudie gebruikt. In deze prospectieve cohortstudie includeerden we 66 patiënten met pijnlijke botmetastasen in het femur die behandeld waren met een eenmalige bestraling (single fraction, SF,  $1 \times 8$  Gy) of bestraling verdeeld over meerdere fracties (multiple fractions, MF, 5 of  $6 \times 4$  Gy) uit drie Nederlandse radiotherapieafdelingen. De patiënten werden tot zes maanden na inclusie gevolgd. Patiënten ondergingen CT-scans vooraf en 1, 4 en 10 weken na radiotherapie. Tijdens deze CT-scans lagen de patiënten op een zogenaamd kalibratiefantoom met bekende dichtheden, waarmee de Hounsfield units (HU) van de CT scan om konden rekenen naar botmineraaldichtheid (BMD). We hebben deze scans uit de CT-femurstudie allereerst gebruikt om het effect van SF en MF radiotherapie op BMD te bepalen in het bestralingsveld en op de metastase, voor lytische, blastische en gemengde laesies apart (**Hoofdstuk 3**). De 42 patiënten die meer dan één CT-scan ondergingen werden hiervoor geïnccludeerd. We bepaalden de gemiddelde BMD op elke CT-scan voor elk proximale femur en meer gedetailleerd voor het specifieke kleinere gebied waarin de metastase zich bevond. We zagen een beperkte stijging in BMD van patiënten met botmetastasen in het femur tien weken na palliatieve radiotherapie waarbij er geen toegevoegd voordelig effect was van MF ten opzichte van SF radiotherapie. Er was geen verandering van BMD in lytische laesies, en een kleine stijging van BMD in gemengde en blastische laesies. Of dat laatste duidt op progressie van de ziekte of op remineralisatie is niet duidelijk, en het klinische effect op botsterkte van het femur moet nog onderzocht worden in de toekomst.

In **Hoofdstuk 4** onderzochten we aan de hand van patiënten uit de CT-femurstudie of het EE-model hun fractuurrisico kon voorspellen. Zeven patiënten hadden in totaal negen femora gebroken tijdens de zes maanden follow-up. We maakten EE-modellen van 39 patiënten (47 femora) met laesies die voornamelijk lytisch van aard waren, en berekenden faalkrachten. We onderzochten of de EE-modellen in staat waren om onderscheid te maken tussen patiënten die wel of geen fractuur hadden opgelopen tijdens follow-up. Op de drie radiotherapieafdelingen waren verschillende CT-scanners gebruikt, wat leidde tot inter-scannerverschillen, waardoor patiënten voor elk van de drie instellingen afzonderlijk werden geanalyseerd. We zagen binnen één van de instellingen dat de faalkracht significant lager

was voor patiënten met fractuur dan voor patiënten zonder fractuur. In de andere twee instellingen was het aantal patiënten dat een fractuur op had gelopen te laag om een duidelijk verschil te zien. Door het EE-model werden fracturen over het algemeen op dezelfde locatie voorspeld als te zien op röntgenfoto's die na de fractuur waren gemaakt. Het EE-model was beter in het identificeren van patiënten met een hoog fractuurrisico in vergelijking met artsen die het fractuurrisico hadden geschat aan de hand van digitaal gereconstrueerde röntgenfoto's (DRR's), met een sensitiviteit van 89% voor het EE-model versus 0% tot 33% voor de klinische schattingen. De specificiteit was 79% voor het EE-model versus 94% tot 95% voor klinische schattingen. De artsen gaven wel aan dat de kwaliteit van de DRR's niet zo goed was als die van de röntgenfoto's die ze normaal gebruiken voor hun fractuurrisicovoorspellingen. We concludeerden dat de EE-modellen waardevol kunnen zijn om de voorspellingen van het fractuurrisico te verbeteren voor patiënten met botmetastasen. In de toekomst zou de verbetering ten opzichte van de huidige klinische richtlijnen moeten worden bevestigd in een grotere patiëntenpopulatie.

Voor de KWF-femurstudie, waarin we 90 patiënten met botmetastasen in het femur includeerden die behandeld waren met radiotherapie in vier Nederlandse radiotherapie-instellingen, kwantificeerden we de verschillen tussen CT-scanners en -scanprotocollen, aangezien dit een van de voornaamste problemen was waar we in Hoofdstuk 4 tegenaan liepen. In **Hoofdstuk 5** onderzochten we daarom het effect van verschillende CT-scanners, reconstructieprotocollen, scanposities en lucht op HU en/of BMD bepaald met een kalibratiefantoom. Hiervoor gebruikten we fantomen die bekende dichtheden bevatten lijkend op de dichtheden van verschillende weefsels. Doordat we fantomen gebruikten, was het mogelijk om de opstelling in de verschillende CT-scanners precies te reproduceren. De fantomen, die ook botgelijkende materialen bevatten, werden gescand op de vier CT scanners. We vonden aanzienlijke verschillen in HU tussen CT-scanners, die kleiner werden als de HU werden gekalibreerd naar BMD. Daarnaast concludeerden we dat verschillende reconstructiekernels, veranderingen in positie binnen het *field of view* (FOV) en luchtruimtes tussen kalibratiefantoom en patiënt vermeden moeten worden. In een opvolgende studie (**Hoofdstuk 6**) wilden we deze effecten kwantificeren voor HU, BMD en EE-faalkracht van femurs van overleden mensen (zogenaamde kadaverfemora). Hiervoor includeerden we zes kadaverfemora die we plaatsten in een anatomisch model van het onderlichaam om een patiënt na te bootsen. Vervolgens scanden we die bovenop een kalibratiefantoom in de vier CT-scanners. De scans werden gemaakt met het standaard protocol van de patiëntenstudie, en daarnaast met variaties in plakdikte, FOV en reconstructiekernel. Een extra scan werd gemaakt met een luchtruimte tussen het model en het kalibratiefantoom. HU en gekalibreerde BMD werden bepaald in een corticaal en trabeculair gebied, en EE-modellen werden gemaakt om faalkrachten te berekenen. Opnieuw zagen we dat er aanzienlijke verschillen waren tussen CT scanners met een gemiddeld verschil van 17% in EE-faalkracht. Daarnaast vonden we een aanzienlijk effect van verschillende reconstructiekernels met een gemiddeld effect van 9% in EE-faalkracht. Lucht tussen het model en het kalibratiefantoom had een klein effect. De Hoofdstukken 5 en 6 lieten zien dat men bewust moet zijn van mogelijke verschillen in CT-waarden en daaruitvolgende EE-uitkomsten wanneer een multicenter studie wordt uitgevoerd waarbij verschillende CT-scanners of -protocols, met name

verschillende reconstructiekernels, worden gebruikt. Die verschillen moeten gecorrigeerd worden voordat de EE-resultaten samengevoegd kunnen worden.

Om EE-modellen te kunnen maken voor alle patiënten met botmetastasen waarvan een gewone klinische CT-scan zonder kalibratiefantom beschikbaar is, onderzochten we de mogelijkheid tot fantoomloos kalibreren in **Hoofdstuk 7**. Uit beide patiëntenstudies includeerden we 67 femora, waarvan we EE-modellen maakten op basis van de fantoomkalibratie en twee fantoomloze kalibratiemethodes: de “lucht-vet-spier”- en “niet-patiëntspecifieke” kalibraties. Voor de lucht-vet-spiervetkalibratie maakten we een histogram van de HU binnen een gestandaardiseerd gebied rond een been van de patiënt. Uit het histogram haalden we de pieken die overeenkwamen met lucht, vet- en spierweefsel, die we vervolgens lineair correleerden aan vastgestelde “BMD”-waarden van de betreffende weefsels om zo een kalibratiefunctie te krijgen. Voor de niet-patiëntspecifieke kalibratie bepaalden we de gemiddelde fantoomkalibratiefunctie over de 26 patiënten die op één van de vier CT-scanners waren gescand. Die gemiddelde fantoomkalibratiefunctie pasten we vervolgens op alle patiënten toe. EE-faalkrachten werden bepaald en vergeleken tussen de fantoom- en fantoomloze kalibratiemethodes. We zagen dat voor beide fantoomloze kalibratiemethodes de faalkrachten zeer goed correleerden met die van de fantoomkalibratie. Ondanks dat de verschillen in faalkracht tussen de fantoomkalibratie en zowel de lucht-vet-spieervet- als de niet-patiëntspecifieke kalibratie klein waren, adviseerden we om de lucht-vet-spiervetkalibratie te gebruiken. Bij het gebruik van de niet-patiëntspecifieke kalibratie weken de EE-faalkrachten namelijk af voor CT-scans die gemaakt waren met een ander reconstructiekern. Met de lucht-vet-spiervetkalibratie zal klinische implementatie van het EE-model makkelijker zijn, omdat EE-modellen gemaakt kunnen worden voor elke patiënt met botmetastasen in het femur, ook als er geen kalibratiefantom beschikbaar is.

In **Hoofdstuk 8** vergeleken we het fractuurrisicovoorspellende vermogen van de EE-modellen met dat van de huidige Nederlandse klinische behandelrichtlijnen waarin de mate van axiale corticale aantasting wordt geadviseerd te meten (bij > 30 mm preventieve operatie overwegen). Om dit te doen combineerden we de klinische inschattingen uit Hoofdstuk 2 met de EE-database, die we met een deel van de patiënten uit Hoofdstuk 4 (CT-femurstudie) en de KWF-femurstudie hadden opgebouwd. We vergeleken de diagnostische nauwkeurigheid (sensitiviteit, specificiteit, PPV en NPV) van beide methodes voor 50 femora (45 patiënten). We zagen dat de EE-modellen beter waren in het voorspellen van fractuurrisico in vergelijking met de klinische richtlijn: het fractuurrisico werd voor meer femora juist ingeschat (sensitiviteit van 100% vs. 86%, specificiteit van 74% vs. 42%, PPV van 39% vs. 19%, en NPV van 100% vs. 95%, voor het EE-model ten opzichte van de klinische richtlijn).

Tot slot reflecteerden we in **Hoofdstuk 9** op het werk beschreven in dit proefschrift en bediscussieerden we de toekomstige uitdagingen voor verdere verbetering en klinische implementatie van de EE-modellen. Omdat de EE-modellen behulpzaam zijn bij het schatten van het fractuurrisico, willen we beginnen met implementatie van de EE-modellen in de klinische praktijk. Het plan is dat artsen een EE-berekening (die we dan “Bone Strength score” of “BOS-score” noemen) aan kunnen vragen net zoals ze een laboratoriumtest aanvragen, vergelijkbaar met bijvoorbeeld genexpressieprofielings bij borst- of longkanker. Een patiëntspecifiek EE-model zal worden gemaakt en een advies over het

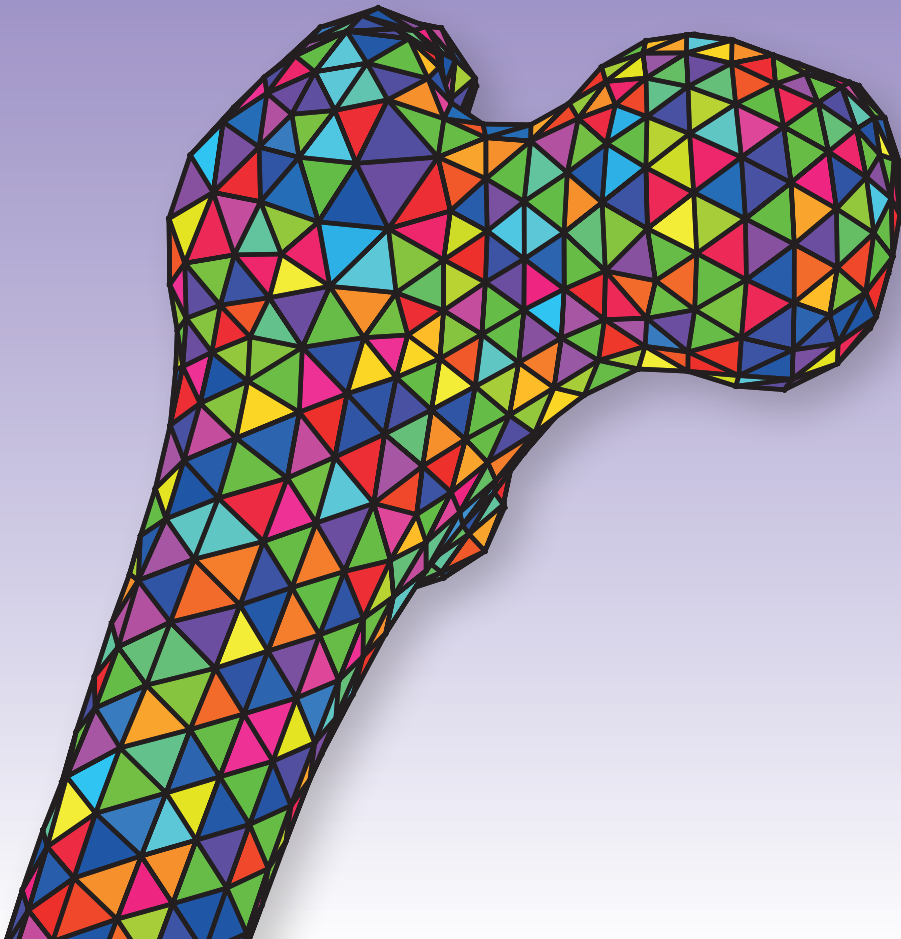
verwachte fractuurrisico zal terug worden gestuurd naar de betreffende arts. Vervolgens kan de arts deze informatie gebruiken in het overleg met de patiënt om de beste behandeling te kiezen. Op deze manier is het EE-model van waarde tijdens *shared decision making*: het proces waarbij arts en patiënt samen de behandelingsopties bespreken om zo rekening te houden met de wensen van de patiënt. Een kwantitatieve berekening van het fractuurrisico zal artsen en patiënten helpen bij het kiezen van de beste behandeling.

Met dit proefschrift en het werk dat we in de toekomst gaan uitvoeren willen we de behandeling van patiënten met uitgezaaide kanker en botmetastasen in het femur verbeteren door het aantal pathologische fracturen enerzijds en onnodige operaties anderzijds te verminderen. Daarmee hopen we te kunnen bijdragen aan de kwaliteit van leven van deze patiënten. In de tussentijd willen we de EE-database vergroten om zo de betrouwbaarheid van de EE-computermodellen als fractuurrisicovoorspellingsmethode te vergroten. Dit kan prospectief met de patiënten waarvan EE-modellen worden gemaakt tijdens de implementatiefase, maar ook retrospectief aan de hand van beschikbare klinische CT-databases.



# CHAPTER 12

DANKWOORD  
CURRICULUM VITAE  
LIST OF PUBLICATIONS  
DATA MANAGEMENT  
PHD PORTFOLIO





## Dankwoord









## Curriculum vitae

Florieke Eggermont was born on 5th February 1989 in Nijmegen, The Netherlands. She grew up in Kekerdom, a small town in the Ooijpolder. In 2007, she obtained her Gymnasiuim diploma from the Stedelijk Gymnasium in Nijmegen, after which she studied Biomedical Sciences at the Radboud University Nijmegen. She obtained her Bachelor's degree in 2010, followed by her Masters drgree with a focus on Clinical Human Movement Sciences in 2013. Her Master's thesis entitled "Biomechanical performance of a non-resorbable, anatomically shaped, total meniscal replacement. A human cadaveric study" was performed at the ORL under supervision of Anne Vrancken, Dennis Janssen and Nico Verdonschot. In 2014, she returned to the ORL to work as scientific researcher on the Bone Metastases project. During her studies and the first years of her professional career, Florieke played volleyball at a semi-professional level in several cities in the Netherlands. She competed in European Club Championships and was Dutch Champion and Cup Winner in 2012, and Cup Runner-up in 2014. After a serious knee injury, she was forced to quit playing.

Florieke is now working as scientific researcher at the ORL on implementation of the finite element model, and as scientific lecturer for the studies Biomedical Sciences and Medicine at the Radboud University Nijmegen.



## List of publications

Van der Wal CWPG, **Eggermont F**, Kroon HM, Ayu O, Fiocco M, Slot A, Snyers A, Rozema T, Verdonchot N, Dijkstra PDS, Tanck E, Van der Linden YM. Validation of a simple radiographic parameter to predict fracturing in metastatic femoral bone lesions. *Radiotherapy & Oncology*, in press.

**Eggermont F**, Van der Wal CWPG, Westhoff PG, De Jong MAA, Rozema T, Kroon HM, Ayu O, Derikx LC, Dijkstra PDS, Verdonchot N, van der Linden YM, Tanck E. Patient-specific finite element computer models improve fracture risk assessments in cancer patients with femoral bone metastases compared to clinical guidelines. *Bone*, in press.

**Eggermont F**, Verdonchot N, van der Linden Y, Tanck E. Calibration with or without phantom for fracture risk prediction in cancer patients with femoral bone metastases using CT-based finite element models. *PLoS ONE*. 2019; 14(7), e0220564.

**Eggermont F\***, Free J\*, Derikx L, van Leeuwen R, van der Linden Y, Jansen W, Raaijmakers E, Tanck E, Kaatee R. The effect of different CT scanners, scan parameters and scanning setup on Hounsfield units and calibrated bone density: a phantom study. *Biomedical Physics & Engineering Express*. 2018; 4(5), 055013.  
\*joint first authorship

**Eggermont F\***, Derikx LC\*, Verdonchot N, van der Geest ICM, de Jong MAA, Snyers A, van der Linden YM, Tanck E. Can patient-specific finite element models better predict fractures in metastatic bone disease than experienced clinicians? Towards computational modelling in daily clinical practice. *Bone Joint Res*. 2018; 7(6):430–439.  
\*joint first authorship

**Eggermont F**, Derikx LC, Free J, van Leeuwen R, van der Linden YM, Verdonchot N, Tanck E. Effect of different CT scanners and settings on femoral failure loads calculated by finite element models. *J Orthop Res*. 2018; 36(8), 2288-2295.

**Eggermont F**, Derikx LC, Verdonchot N, Hannink G, Kaatee RSJP, Tanck E, Van der Linden YM. Limited short-term effect of palliative radiotherapy on Quantitative Computed Tomography-derived bone mineral density in femora with metastases. *Advances in Radiation Oncology*. 2017;2(1): 53-61.

Vrancken AC, **Eggermont F**, van Tienen TG, Hannink G, Buma P, Janssen D, Verdonchot N. Functional biomechanical performance of a novel anatomically shaped polycarbonate urethane total meniscus replacement. *Knee Surg Sports Traumatol Arthrosc*. 2016;24(5):1485-94.

Jansen M, De Jong M, Coes HM, **Eggermont F**, Van Alfen N, De Groot IJ. The assisted 6-minute cycling test to assess endurance in children with a neuromuscular disorder. *Muscle Nerve*. 2012;46(4):520-30.

Smulders E, van Lankveld W, **Eggermont F**, Duysens J, Weerdesteyn V. Step performance in persons with rheumatoid arthritis: a case-control study. *Arch Phys Med Rehabil*. 2011;92(10):1669-74.

## Conference abstracts

**Eggermont F**, van der Wal G, Westhoff P, Laar A, de Jong M, Rozema T, Kroon HM, Ayu O, Derikx L, Dijkstra S, Verdonchot N, van der Linden Y, Tanck E. Patient-specific finite element computer models improve fracture risk predictions in cancer patients with femoral bone metastases compared to clinical guidelines. European Orthopaedic Research Society, 2019, Maastricht, The Netherlands.

**Eggermont F**, van der Linden Y, Derikx L, de Jong M, Snyers A, Rozema T, Verdonchot N, Tanck E. Patient-specific finite element models for fracture risk prediction in patients with cancer and bone metastases. 7th Dutch Bio-Medical Engineering Conference, 2019, Egmond aan Zee, The Netherlands.

**Eggermont F**, Verdonchot N, Janssen D, van der Linden Y, Tanck E. Phantom calibration versus phantomless calibration for finite element modelling of femurs with bone metastases in patients with cancer. 8th World Congress of Biomechanics, 2018, Dublin, Ireland.

**Eggermont F**, Derikx LC, Free J, Van der Linden Y, Verdonchot N, Tanck E. Effect of different CT scanners and settings on bone CT values for finite element predicted femoral failure load. 6th Dutch Bio-Medical Engineering Conference, 2017, Egmond aan Zee, The Netherlands.

**Eggermont F**, Derikx LC, Free J, Van der Linden Y, Verdonchot N, Tanck E. Effect of Different CT Scanners and Settings on Bone CT values for Finite Element Predicted Failure Load. European Society of Biomechanics 22nd Conference, 2016, Lyon, France.

**Eggermont F**, Derikx LC, Verdonchot N, Hannink G, Kaatee RSJP, Tanck E, van der Linden YM. Limited short-term effect of radiotherapy on bone density in metastatic femoral bone. ESTRO 35 Conference, 2016, Turin, Italy.

Vrancken ACT, **Eggermont F**, van Tienen TG, Hannink G, Janssen D, Verdonchot N, Buma P. Functional biomechanical performance of a novel anatomically shaped polycarbonate urethane total meniscus replacement. 10th Biennial ISAKOS Congress, 2015, Lyon, France.

Vrancken ACT, **Eggermont F**, van Tienen TG, Hannink G, Janssen D, Verdonschot N, Buma P. Biomechanical performance of an anatomically shaped polycarbonate urethane total meniscus replacement. NOV scientific meeting, 2015, Maastricht, the Netherlands.

Vrancken ACT, **Eggermont F**, van Tienen TG, Janssen D, Verdonschot N, Buma P. Biomechanical performance of an anatomically shaped polycarbonate urethane total meniscus replacement. World Congress of Biomechanics, 2014, Boston, USA.

Vrancken ACT, **Eggermont F**, van Tienen TG, Janssen D, Verdonschot N, Buma P. Biomechanical performance of a polycarbonate urethane total meniscus replacement. ORS Annual Meeting, 2014, New Orleans, USA.

## Awards

### **Best Poster Award at the 7th Dutch Conference on Bio-Medical Engineering 2019**

Eggermont F, van der Linden Y, Derikx L, de Jong M, Snyers A, Rozema T, Verdonschot N, Tanck E. Patient-specific finite element models for fracture risk prediction in patients with cancer and bone metastases.

**Science Award Radboud Institute for Health Sciences 2018** for best peer-reviewed scientific publication of 2018 within the Radboud Institute for Health Sciences

Eggermont F, Derikx LC, Verdonschot N, van der Geest ICM, de Jong MAA, Snyers A, van der Linden YM, Tanck E. Can patient-specific finite element models better predict fractures in metastatic bone disease than experienced clinicians? Towards computational modelling in daily clinical practice. Bone Joint Res. 2018 Jul 7;7(6):430–439.





## Data management

CT femur study protocol was approved by Regionale Toetsingscommissie Patiëntgebonden Onderzoek, Leeuwarden (NL12568.099.06). KWF femur study protocol was approved by Commissie Mensgebonden Onderzoek regio Arnhem – Nijmegen (2013/305). Both studies were conducted in accordance with the principles of the Declaration of Helsinki.

The privacy of the patients in these studies is warranted by use of encrypted and unique individual subject codes. All patients signed informed consent, which is stored within the patient records of the respective radiotherapy department, along with the encryption key to identify patients from subject codes. The researchers have no access to any patient identifying information.

Patients' baseline forms and questionnaires were filled in on paper and are stored in the fireproof safe at the department archive (Radboudumc, room M379.-1.219). All paper data were digitalized using Castor EDC. Data management and monitoring were also performed within Castor EDC. An audit trail was incorporated to provide evidence of the activities that has altered the original data.

All data obtained for all chapters of this thesis are archived on the secured department-specific drive on the Radboudumc network, which is only accessible for selected employees of the Orthopaedic Research Laboratory.



## PhD portfolio

Institute for Health Sciences  
**Radboudumc**

Name PhD candidate:	F.E. Eggermont	PhD period:	01-03-2014 – 18-12-2019
Department:	Orthopaedic Research Laboratory	Promotor:	Prof. N. Verdonchot
Graduate School:	Radboud Institute for Health Sciences	Co-promotors:	Dr E. Tanck Dr Y.M. van der Linden

### TRAINING ACTIVITIES

	Year(s)	ECTS
<b>a) Courses &amp; Workshops</b>		
Introduction day Radboudumc	2014	0.5
BROK course	2017	1.5
edX-course "Using Python for Research", HarvardX	2017	1.5
Workshop Science & Skills: Illustrator	2019	0.1
Introductie Nijmeegse curricula	2019	0.2
<b>b) Seminars &amp; lectures</b>		
Radboud Research Rounds	2016-2018	0.3
NVPHBV Spring meeting 2017	2017	0.3
<b>c) Symposia &amp; congresses (OP/ PP indicates oral or poster presentation)</b>		
ESTRO35, Turin, Italy (PP)	2016	1.8
Congress of the European Society of Biomechanics 22, Lyon, France (OP)	2016	1.5
RIHS PhD retreat, Wageningen, The Netherlands (OP)	2016	0.5
6th Dutch Bio-Medical Engineering Conference, Egmond aan Zee, The Netherlands (OP)	2017	1.0
RIHS PhD retreat, 's-Hertogenbosch, The Netherlands (OP)	2017	0.5
RIHS PhD retreat, 's-Hertogenbosch, The Netherlands (OP)	2018	0.5
7th Dutch Bio-Medical Engineering Conference, Egmond aan Zee, The Netherlands (PP)	2019	1.0
European Orthopaedic Research Society 2019, Maastricht, The Netherlands (OP)	2019	1.5
<b>d) Other</b>		
Journal Clubs Orthopaedic Research Lab	2014-2016	3.0
Lab presentations Orthopaedic Research Lab	2014-2019	5.3

## Continued

TEACHING ACTIVITIES		
	Year(s)	ECTS
<b>e) Lecturing</b>		
BSc course 5DT03 – Determinanten 3: Fysische Factoren	2016	0.5
BSc course MIN05 – Moving Questions	2016-2019	2.4
BSc course 6MBB – Belasting en Belastbaarheid	2017,2019	0.8
MSc course 5HM02 – Tissue: Biomechanics and Engineering	2016	1.5
MSc course BMS53 – Orthopaedic Biomechanics in Motion	2017-2019	3.6
VvBN Studentendag	2015	0.2
RIHS introduction for BMW master students	2016	0.2
<b>f) Supervision of internships / other</b>		
BSc internship:		
Arjonne Laar, RU Nijmegen	2017	1.3
Milan Baars, RU Nijmegen	2018	0.3
MSc internship:		
Lieke van den Wildenberg, TU/e Eindhoven	2017	1.0
Jelle Eikhout, RU Nijmegen	2019	1.5
Science Project Biomedical Sciences and Medicine, RU Nijmegen		
Inger Albeda, Miriam Boot, Nadiëh Jansen	2017	2.0
Literature thesis:		
Arjonne Laar, RU Nijmegen	2019	0.3
Research project Clinical Technology, TU Delft		
Renée van den Brand, Caspar Coco Martin, Severano Verhoef, Marlein Vogels	2019	0.6
<b>OTHER</b>		
Reviewing scientific articles	2017,2019	0.2
<b>TOTAL</b>		<b>37.4</b>



Appendix: Flowchart of patients in all studies

