Towards clinical application of biomechanical tools for the prediction of fracture risk in metastatic bone disease

Loes C. Derikx a,b,* , Nico Verdonschot a,c, Esther Tanck a

a Orthopaedic Research Laboratory, Radboud Institute for Health Sciences, Radboud university medical center, Nijmegen, The Netherlands
b Human Movement Biomechanics Research Group, KU Leuven, Leuven, Belgium
c Laboratory of Biomechanical Engineering, University of Twente, Enschede, The Netherlands

1. Introduction

Bone is a preferred organ for primary tumour cell seeding in common cancer types such as breast, prostate, lung, kidney and thyroid cancer (Coleman, 1997, 2006; Gralow et al., 2013; Laitinen et al., 2012). Particularly skeletal parts that contain bone marrow (e.g. the skull, ribs, spine, and long bones of the axial skeleton) provide a fertile environment for seeding and are therefore commonly invaded by tumour cells (Johnson and Knobf, 2008; Laitinen et al., 2012; Mavrogenis et al., 2012). In more progressive states bone metastases can, amongst others, cause pathological fractures (Coleman, 1997; Laitinen et al., 2012; Mantyh, 2013), in which case patients lose their full mobility and may develop severe complications (Mavrogenis et al., 2012; Ruggieri et al., 2010). Pathological fractures are treated with complex surgical procedures. Surgeons have to weigh the impact of the operation and rehabilitation against the physical status and expected survival of the patient (Attar et al., 2012). In addition, they must be convinced that the load capacity of the reinforced bone will sustain the daily loads for the life expectancy of the patient (Attar et al., 2012). In current clinical practice, metastatic lesions identified with an impending fracture are treated with preventive surgery. This treatment is less complex and has better survival rates than surgical treatment of actual pathological fractures (Laitinen et al., 2012; Mavrogenis et al., 2012; Ratasvuori et al., 2013). Lesions that do not jeopardise the mechanical integrity of the bone are treated conservatively with (a combination of) radiation therapy, analgesics, chemotherapy, hormonal therapy or bisphosphonates, with the aim to relieve pain (Van der Linden, 2005). However, it turns out to be extremely difficult, if not impossible, to assess clinical fracture risks based on conventional X-rays or CT images. Hence, even for experienced clinicians, it is impossible to make accurate predictions. This was well demonstrated in a study by Hipp et al. (1995). They used 10 paired cadaver femurs, with an artificial lesion drilled in one of the femurs of the pair. The failure load for each bone and the strength reduction within a pair was determined based on a mechanical axial loading experiment. Using CT-scans and roentgenograms of the bones, three orthopaedic surgeons were asked to report on the lesion size, the femoral failure load and the strength reduction within a femoral pair. There was a moderate agreement in defining the lesion size (mean difference 11%, range 2–47%), but there was no relationship between the failure load measured in the experiments and the failure load estimated by the surgeons. The same disappointing result was found for the strength reduction in the femoral pairs. In a comparable experiment we showed very similar results. Clinical experts were asked to rank femurs with and without artificial lesions on bone strength; the rank correlations between experimental bone strength and predictions by clinical experts ranged only between 0.45 and 0.53 (Fig. 2) (Derikx et al., 2012). This demonstrates that a more quantitative measure of bone strength in patients with metastatic bone disease is urgently needed.

In this paper we provide an overview of more objective and quantitative approaches that have been proposed for fracture risk prediction.
assessment in metastatic bone disease. We discuss their efficacy and the potential challenges that may come in with the clinical implementation of such tools.

2. Clinical methods for fracture risk assessment in metastatic bone disease

Finding an objective measure for fracture risk assessment of bones with metastases has been under study for several decades. By evaluating roentgenograms of patients who sustained a pathological fracture, the size of the lesion (Beals et al., 1971; Cheng et al., 1980; Harrington et al., 1976; Keene et al., 1986; Miller and Whitehill, 1984; Snell and Beals, 1964; Van der Linden et al., 2004; Zickel and Mouradian, 1976), the extent to which cortical bone was disrupted by the lesion (Van der Linden et al., 2004) and the radiographic appearance of the lesion (Beals et al., 1971; Bunting et al., 1985; Keene et al., 1986; Miller and Whitehill, 1984; Mirels, 1989; Snell and Beals, 1964; Van der Linden et al., 2004; Yazawa et al., 1990; Zickel and Mouradian, 1976) have been studied as potential predictors for the fracture risk. Pain has been included as well in these studies (Beals et al., 1971; Fidler, 1973; Harrington et al., 1976; Keene et al., 1986; Mirels, 1989; Parrish and Murray, 1970; Van der Linden et al., 2004), as it was hypothesised to be a measure for loss of mechanical strength (Mirels, 1989), or an indicator of excessive deformation (Fidler, 1973). Despite these documented efforts, none of the studies identified a powerful predictor for the fracture risk. The most recent clinical study in this field compared, amongst others, two guidelines: Mirels’ scoring system and a threshold for cortical disruption (Van der Linden et al., 2004). Mirels’ system scores the location of the lesion, pain and the appearance and size of the lesion. Patients with high scores need immediate surgery, while patients with low scores can be treated conservatively. Had Mirels’ scoring system been applied in the patients in the study of van der Linden et al., none of the impending fractures would have been missed but a large number of patients would have undergone unnecessary surgery (sensitivity=1.0, specificity=0.13). Alternatively, a threshold of 3 cm cortical disruption was proposed to identify impending pathological fractures. Had this method been used in Van der Linden’s work, some of the impending fractures would have been missed (sensitivity=0.86), but the power to identify non-fracture patients would have increased (specificity=0.58). Thus, the latter guideline improved upon Mirels’ scoring system but remained to have difficulties preventing unnecessary surgeries.

In summary, clinical studies so far have mainly focussed on lesion characteristics and pain, while the bone strength of the femur was largely ignored. In order to estimate the fracture risk, however, it is important to assess the reduction in bone strength caused by the lesion with respect to the initial bone strength.

3. Mechanical models to assess femoral bone strength

More recently, the focus has shifted towards mechanical models for fracture risk assessment. The most commonly proposed ones are computed tomography based rigidity analysis (CTRA) and patient-specific finite element (FE) modelling.

The use of composite beam theory in the context of fracture risk assessment has been extensively investigated over the last two decades (Leong et al., 2010; Snyder et al., 2009; Snyder et al., 2006; Windhagen et al., 1997). Starting in the spine, Windhagen et al. (1997) generated quantitative CT (QCT) scans from vertebral segments with artificial and actual metastatic lesions and mechanically loaded them until fracture. Using a calibration phantom, the grey values in the CT scans were converted to ash densities and Young’s moduli, respectively, and the axial rigidity for every CT slice was subsequently calculated based on composite beam theory. High correlations were found between the experimental failure load and axial rigidity ($R^2$ ranging from 0.79 to 0.85). No correlation was found between defect size and failure load, which confirms earlier findings showing that lesion characteristics alone cannot accurately predict fracture risk in metastatic bone. Additionally, CTRA was applied in the femur in a clinical setting. Snyder et al. (2006) included 36 patients with benign femoral lesions, 18 of which had sustained a fracture. Axial, torsional and bending rigidities were calculated for the affected bone and the intact contralateral bone, respectively (Fig. 1). Statistical analysis revealed no difference in lesion characteristics between the two groups, but the relative reduction in rigidity (i.e. the difference in rigidity between the intact and the affected bone) was significantly larger in the fracture group than in the non-fracture group. Thus, in the fracture group the lesions had weakened the bone to a larger extent than in the non-fracture group. This was the case for axial, bending and torsional rigidity. Based on these results, cut-off values were defined, on the

Fig. 1. The left panel shows the relative reduction in bone mineral content (BMC), axial rigidity (axial), minimum bending rigidity (min bend) and torsional rigidity (torsional) for every CT slice. This reduction is largest in the slice at location – 385 mm (＞50% for minimum bending rigidity) and indicated with the black line in the lateral plane (middle panel). In the right panel, the affected and unaffected femurs at this specific level are shown. Reprinted from (Snyder et al., 2006) with permission.
basis of which fracture patients and non-fracture patients could be accurately distinguished (sensitivity = 1, specificity = 0.94). This method is rather straightforward, and relatively quick, but, as acknowledged by the authors, it requires an intact contralateral femur for calculation of the reduction in rigidity. In metastatic bone disease, however, this contralateral femur is likely to be affected as well, which hampers the use of this method. Moreover, the outcome measures (axial, bending or torsional rigidities) are difficult to interpret in clinical practice. Finally, the method is not suited to assess effects of different external loading modes and is therefore difficult to use in a clinical setting where clinicians would like to advise the patient as to which daily activities can be safely performed and which may lead to bone fracture.

The other promising tool for the prediction of fracture risk in metastatic bone disease is patient-specific finite element (FE) analysis. Although this method is widely studied to calculate the fracture risk in osteoporosis (for example (Bessho et al., 2009; Keyak et al., 2013; Kopperdahl et al., 2013; Orwell et al., 2009)), few groups have used it for assessing failure load in metastatic bone disease in the femur (Cheal et al., 1993; Derikx et al., 2012; Keyak et al., 2005a, 2007; Spruijt et al., 2006; Tanck et al., 2009). Cheal et al. (1993) were one of the first to use FE modelling for this purpose. Unfortunately, they found large differences between the calculated failure loads and the failure loads measured in their experiments. These inferior results may be explained by the fact that they used a femoral FE model based on average anatomy and material behaviour data (Cheal et al., 1993), and therefore did not capture the relevant biomechanical differences that exist amongst bones of different subjects.

Some years later, Keyak and co-workers developed and extensively validated a full workflow for patient-specific FE modelling based on QCT images. They empirically established relationships between CT values and bone material properties (Keyak et al., 1996). Subsequently, in a mechanical test setup they loaded intact cadaver femurs until failure and found good agreement between the experiments and the FE simulations (Keyak, 2001; Keyak et al., 2005b). Additionally, they applied this workflow to femurs with simulated and actual metastatic lesions (Keyak et al., 2005a, 2007) and were able to accurately predict bone strength ($R^2 = 0.97$, $r = 0.98$ and $r = 0.94$, for intact femurs, and femurs with simulated and actual lesions, respectively) (Keyak et al., 2005a).

Building on the work by Keyak et al. (2005b), we have developed and validated a workflow for generating subject-specific finite element models. For the validation of our model, we experimentally tested 10 paired femurs under axial loading conditions (Derikx et al., 2012; Tanck et al., 2009). Before testing, we retrieved QCT scans of these femurs, on the basis of which we generated subject-specific FE models. The axial loading conditions from the experiments were mimicked, and the FE models were able to accurately predict the experimental failure load ($R^2 = 0.90$ for intact femurs and $R^2 = 0.93$ for metastatic femurs). We additionally showed that, under these simple loading conditions, FE models can outweigh the performance of clinical experts (Fig. 2) (Derikx et al., 2012). By implementing more realistic material behaviour, we were able to further improve the prediction of failure locations (Derikx et al., 2011). Although results were very promising in a cadaver study, it is still unclear if these methods can also improve fracture risk assessment in patients.

Currently, we are running a first prospective study to test if these FE approaches can improve upon standard clinical guidelines with regard to the prediction of bone fractures in patients. In this prospective cohort study, including 66 patients with femoral bone metastases, nine femurs fractured during follow-up. Preliminary results for 23 of these patients (with five fractures in three patients) show that the mean load capacity in the fracture group is significantly lower than in the non-fracture group, but confidence intervals between groups overlap. This indicates that the predictions are not yet accurate enough on patient level, which obviously is required for implementation in clinical practice. These preliminary results also show that FE models, with model definitions as applied in the cadaver study, are currently able to identify fracture cases and non-fracture cases with a sensitivity and specificity in the same order as previous clinical guidelines. Fortunately, a number of these model definitions can be further improved to better suit in vivo fracture risk assessments. As a result, the number of over- and undertreated patients could decrease, so that the FE models could outperform existing clinical guidelines. This would render the method safe to start clinical trials in which the FE models are used for making clinical treatment decisions.

4. Future opportunities for FE analysis in clinical fracture risk assessments

While current simulation results are promising, there are a number of opportunities to further improve the predictive capacity of these models.

First of all, it is important to develop realistic loading conditions (using musculoskeletal modelling) and to couple them to FE-models. This would enable modelling of actual daily life activities, which could lead to more tailored fracture risk predictions. For example, the femoral load capacity of a femur under axial loading would be greatly affected by a lesion in the medial shaft, but far less by a lesion in the greater trochanter. The latter case would have a great influence on the bone strength if the applied loading condition included the abductor muscles. It should be noted, however, that the use of such musculoskeletal models is not straightforward and may in turn lead to (additional) errors entering the model (Lund et al., 2012). Hence, the subsequent effect on the accuracy of the fracture risk predictions should be elucidated.

Secondly, metastatic bone tissue may require adapted material models. The material models used in FE modelling are generally based on empirical studies investigating the relationship between CT intensity and material behaviour of healthy bone tissue, as is the case in our FE models. Yet, the composition and consequent material behaviour of diseased bone tissue may be rather different from healthy tissue. Although Keyak et al. (2005a) showed that the use of the healthy material model is valid, adapted material models may be needed in case of extensive sclerotic or mixed type lesions.

In addition, the transition from modelling cadaver experiments to in vivo fracture risk assessments using FE models should be done with caution. When scanning patients in vivo, bony structures and soft tissues affect the CT attenuation in the femur. These effects are obviously dependent on anatomy and are therefore patient-specific. In order to minimise these beam hardening effects, a calibration phantom should be used, which enables to establish patient-specific or even image-specific calibration lines for converting CT intensities to calcium equivalent values.

It may be challenging to perform multicentre studies and deal with subsequent differences in CT images. These centres probably use CT scanners from multiple manufacturers, who apply different algorithms to reconstruct the CT images. Robust calibration procedures are needed to correct for these differences in order to reliably establish material properties. Recent work by Carpenter et al. (2014) showed large differences in calculated femoral bone strength based on QCT images retrieved from two different scanners, especially under single leg stance loading (mean difference $-1100$ N, 95% CI between 390 and $-2526$ N, approximately; independent of femoral strength). The use of hydroxyapatite calibration phantoms could not sufficiently correct for these differences between scanners. Obviously, such measurement errors are unacceptable when using FE predictions for clinical decision making on a patient-specific basis, and rigorous alternative calibration protocols should be developed for multicentre studies.
Another opportunity to improve the predictive capacity is the implementation of anisotropic material behaviour of bone. Since bone anisotropy cannot be quantified using a clinical CT resolution, considerable work should be done on multi-level modelling in order to find a proper way of implementing these mechanical properties in a patient-specific manner. The first steps in that direction have been taken by Hazrati Marangalou et al. (2013). Although their findings were not yet confirmed using clinical CT images, the results were promising and suggest that extrapolating anisotropic material behaviour from micro-level CT data to macro-level FE models is possible. In addition, imaging techniques are vastly improving, and it is only a matter of time before this micro-level information can be gleaned from clinical CT images with a radiation dose acceptable for in vivo scanning. Alternatively, pre-clinical research on bone specimens showed that ultrasound can be used to determine the structural properties of bone (Lin et al., 2012). However, whether this new technique is applicable in vivo as well, remains to be seen.

From a more practical point of view, the workflow for fracture risk assessment using FE models should be accelerated beyond a clinically acceptable limit. Clinical implementation is currently hampered by the fact that the procedure to calculate the fracture risk needs specific modelling software and engineering knowledge. In order to make these mechanical tools available for experts in clinical practice, the workflow should be further automated. A promising method to do so is probabilistic modelling (Taylor et al., 2013), which would use principal component analysis to select characteristics from the FE models that are statistically predictive for the fracture risk. If these significant components are determined for every patient, the statistical model can be used to calculate the individual fracture risk. In this way, the extensive patient-specific modelling becomes redundant, and the fracture risk assessment will be accelerated to a clinically acceptable time span.

By addressing the opportunities mentioned above, further improvements on in vivo fracture risk predictions can be established, which will lead to decreased numbers over- and undertreated patients. Obviously, the FE models can only be introduced into clinical practice if they outperform existing guidelines. Since these guidelines poorly predict the patient-specific fracture risk, there is room for improvement of clinical decision making using mechanical models. Hence, although the validity of these models is not yet perfect, in the near future they may become of great added value. There is no doubt that with the modelling and imaging opportunities ahead, these tools will find their way to clinical practice in the very near future.

5. Conclusion

In this paper we discussed the clinical practice of metastatic bone disease, which lacks an accurate predictor for the fracture risk.
Current scoring systems omit mechanical parameters (such as initial bone strength) that are crucial for sound fracture risk assessments, and therefore result in relatively high numbers of over- and under-treated patients. Alternatively, biomechanical tools, such as CTRA and FE modelling, have been proposed for clinical fracture risk assessment in metastatic bone disease. In an experimental setting, both methods show promising results; FE models have shown to outperform clinical experts. Obviously these models need to prove their benefit in clinical trials before this technology can be made available for clinical experts treating patients with metastatic bone disease. This transition to in vivo fracture risk assessment is challenging, but there are a number of modelling and imaging opportunities to further improve the predictive capacity of these models. With such improvements ahead, these tools may become of great added value for the fracture risk assessment in metastatic bone disease and as such will find their way to clinical practice.

Conflict of interest statement

All authors declare that they do not have any conflict of interest that could inappropriately influence this work.

Note from the authors

Rik loved science. Rik was science. He enjoyed discussing and developing theories about the mechanisms of bone adaptation and taught us the basics of Wolff’s law. Nowadays, we ourselves teach students, in Rik’s spirit, how bone adapts and often use the phrase ‘use it or lose it’ which Rik used on many occasions. In this paper we focus on metastatic bone disease, a devastating illness in which bone is resorbed by cancer cells and normal bone adaptation rules do not apply anymore. Although Rik favoured fundamental research, he definitely stimulated applied research. Not because he wished to be involved but mainly because he saw the urgency to improve the connexion between biomechanics and clinical practice. The research within the lab in Nijmegen has progressed in his spirit and we still feel the presence of his critical and stimulating mind. We hope to make him proud!

Acknowledgements

This work was funded by the Dutch Cancer Foundation and Fonds NutsOhra.

References


