Prediction of femoral fracture load using automated finite element modeling

Joyce H. Keyak, Stephen A. Rossi, Kimberly A. Jones, Harry B. Skinner

Department of Orthopaedic Surgery, University of California Irvine Medical Center, Orange, CA, U.S.A.
Rehabilitation Research and Development Service, Department of Veterans Affairs Medical Center, San Francisco, CA, U.S.A.
Bioengineering Graduate Group, University of California, San Francisco/Berkeley, CA, U.S.A.

Received in final form 17 November 1997

Abstract

Hip fracture is an important cause of morbidity and mortality among the elderly. Current methods of assessing a patient's risk of hip fracture involve local estimates of bone density (densitometry), and are limited by their inability to account for the complex structural features of the femur. In an effort to improve clinical and research tools for assessing hip fracture risk, this study investigated whether automatically generated, computed tomographic (CT) scan-based finite element (FE) models can be used to estimate femoral fracture load in vitro. Eighteen pairs of femora were examined under two loading conditions — one similar to loading during the stance phase of gait, and one simulating impact from a fall. The femora were then mechanically tested to failure and regression analyses between measured fracture load and FE-predicted fracture load were performed. For comparison, densitometry measures were also examined. Significant relationships were found between measured fracture load and FE-predicted fracture load ($r = 0.87$, stance; $r = 0.95$, fall; $r = 0.97$, stance and fall data pooled) and between measured fracture load and densitometry data ($r = 0.78$, stance; $r = 0.91$, fall). These results indicate that this sophisticated technique, which is still early in its development, can achieve precision comparable to that of densitometry and can predict femoral fracture load to within $\pm 40\%$ to $\pm 60\%$ with $95\%$ confidence. Therefore, clinical use of this approach, which would require additional X-ray exposure and expenditure for a CT scan, is not justified at this point. Even so, the potential advantages of this CT/FE technique support further research in this area. © 1998 Elsevier Science Ltd. All rights reserved.

Keywords: Hip fracture; Osteoporosis; Finite element; Femur; Bone strength

1. Introduction

Hip fracture is an important cause of morbidity and mortality among the elderly. In the U.S.A. over 260,000 hip fractures occur each year in patients 65 yr of age or older (Graves, 1992). A great proportion of these fractures lead to permanent disability and/or death (Cummings et al., 1985; Kenzora et al., 1984; Mullen and Mullen, 1992; White et al., 1987), and in about 50% of all cases, recovery is not achieved within one year (Holbrook et al., 1984; Miller, 1978). The high cost of treatment, estimated at $7.1 billion annually in the U.S.A. (Praemer et al., 1992), and the disabling nature of this injury point to the importance of preventing hip fracture. An inexpensive and accurate means of assessing hip fracture risk would enable patients at high risk to be identified so that preventive measures could be taken.

Current methods of assessing hip fracture risk are based on the assumption that reduced bone strength is related to fracture incidence and employ two-dimensional (2-D) or three-dimensional (3-D) imaging techniques (i.e. densitometry), such as dual-energy X-ray absorptiometry (DXA) or quantitative computed tomography (QCT), to obtain estimates of bone mineral density in the femur or other bones (Cann, 1988; Cummings et al., 1993; Glueer et al., 1994; Grampp et al., 1993; Lang et al., 1991; Mazess et al., 1988; Melton et al., 1986; Perloff et al., 1991; Riggs et al., 1982; Vega et al., 1991). These methods have been only partly successful in identifying patients who fracture, and are inherently limited by
their inability to account for complex geometry, bone heterogeneity, loading conditions, or variable fracture location. Attempts to improve densitometry-based techniques have been directed at developing better predictors of femoral fracture load by including simple measures of femoral geometry in the evaluation (Alho et al., 1988; Courtney et al., 1994, 1995; Faulkner et al., 1993; Lotz and Hayes, 1990), or by performing 2-D structural engineering analyses (Beck et al., 1990; Gies et al., 1985). However, these efforts have resulted in only marginal improvements.

In an effort to improve clinical and research tools for assessing hip fracture risk, this study investigated whether automatically generated, computed tomographic (CT) scan-based finite element (FE) models can be used to estimate femoral fracture load. To address this question, two loading conditions were studied in vitro — one similar to loading investigated previously (Backman, 1957; Lotz and Hayes, 1990); opposing forces were applied during mechanical testing. For the femora tested in the fall configuration, a cup was also made for the greater trochanter.

For FE model generation, the femur and attached PMMA base were immersed in water and placed atop a calibration phantom (0–200 mg cm$^{-3}$ K$_2$HPO$_4$ in water) (Lang et al., 1991), and a CT scan of the femur was taken using a GE 9800 Research Scanner (General Electric, Milwaukee, WI, U.S.A.) (80 kVp, 280 mAs, contiguous 3.0 mm-thick slices, 1.08 mm pixels, 320 × 320 matrix, standard reconstruction). The CT scan data were transferred to an Indigo Iris computer (Silicon Graphics, Mountain View, CA), and a 3-D FE model using heterogeneous linear isotropic mechanical properties was automatically generated for each femur (Keyak et al., 1990). Linear, eight-noded cube-shaped elements measuring 3 mm on a side were used so that the elements matched the thickness of the CT scan images. The FE models, which each took about 30 min to generate, consisted of 6876–19,151 nodes and 5152–15,552 elements, depending on femur size (Fig. 2).

The elastic modulus and strength of each element were computed as follows. The CT scan data were calibrated in terms of K$_2$HPO$_4$ equivalent density, and apparent ash density was estimated from these data using the relation reported by Les et al. (1994) for equine bone. This relation is close to one found previously for human tibial trabecular bone (Keyak et al., 1994) and is valid for both cortical and trabecular bone. Modulus and strength were computed from ash density using reported correlations for trabecular bone (Keyak et al., 1994) and cortical bone (Keller, 1994) (Table 1). To speed FE analysis, all elements with a modulus below 5 MPa were assigned a new modulus of 0.01 MPa (essentially zero), and the remaining element moduli were grouped, requiring an adjustment of no more than $\pm$ 2.5% for each element, to obtain approximately 170 modulus values. The final moduli which met this $\pm$ 2.5% criterion were 5.125, 5.381, ..., 21.531, 22.607 MPa, or equivalently, $5.125 \times (1.05)^{n-2}$ MPa, where $n = 2, 3, \ldots, 174$. A constant Poisson’s ratio of 0.4 was assumed (Reilly and Burstein, 1975; Van Buskirk and Ashman, 1981).

Boundary conditions were applied to the FE models to represent the conditions of mechanical testing with a quasi-static force of 1000 N. Force magnitude was arbitrary because model linearity enabled the stress/strain results to be scaled to obtain results for any magnitude. For each model, the direction and location of the applied force were extrapolated from CT scan coordinates of the PMMA base attached to the femur, thereby enabling accurate representation of the experimental loads. Force was equally divided over nodes on the loaded portion of the femoral head, and the distal portion of the model was fully restrained at the level corresponding to the most proximal portion of the PMMA base. For the fall load configuration, nodes representing the portion of the greater trochanter that was in contact.
Fig. 1. Stance (a) and fall (b) loading conditions used in this study. The FE models attempted to represent these conditions.
with the PMMA cup were restrained in the direction of the applied load. The models were analyzed using ABAQUS software (version 5.4, Hibbitt, Karlsson, and Sorensen, Providence, RI).

To predict the fracture load of a femur, a factor of safety (FOS) for each element was computed at the element centroid using the distortion energy theory of failure:

$$\text{FOS} = \frac{\text{element strength (derived from the CT scan data)}}{\text{element von Mises stress (computed the FE model)}}$$

FOS less than one indicated element failure. Elements at the model surface were disregarded because partial volume effects made results at the surface unreliable. Neglecting these elements may have been a source of error if fracture began at the bone surface; however, this error was partially offset by the fact that, on average, these elements consisted of only 50% bone by volume and therefore played a minor structural role. In addition, multiple elements were used to identify the fracture load (see below), thereby reducing the effect of disregarding individual surface elements.

The predicted femoral fracture load ($F_{FE}$) was defined as the load at the onset of local material failure, as indicated by the element FOS values. This methodology assumed that fracture began in the region where FOS values were less than one. Because failure of just a few finite elements was thought to be structurally insignificant and could have been prone to artifacts, $F_{FE}$ was defined a priori as the load at which the factors of safety for 15 contiguous nonsurface elements were less than 1 (Fig. 3).1 Linearity of the models enabled $F_{FE}$ to be computed after a single FE analysis by simply finding the 15 contiguous elements with lowest FOS and then scaling the load, FE data, and FOS data until those 15 FOS values were less than or equal to 1. This post-processing of the FE data was performed in just a few minutes using in-house software.

After $F_{FE}$ was computed, the measured femoral fracture load ($F_{Meas}$) was determined by mechanical testing to failure under displacement control at 0.5 mm s$^{-1}$ (Bionix 858 Test System, MTS, Eden Prairie, MN). Grease was placed between the PMMA cups and loading platens to minimize transverse forces. $F_{Meas}$ was defined as the load at the first peak of the force–displacement curve. Given the definition of $F_{FE}$ used in this study, a more appropriate definition of $F_{Meas}$ would have been the load at the measured onset of failure. However, this latter definition could not be implemented because the onset of failure was, by nature, highly localized and consequently had a negligible effect on the force–displacement curve. As a result, $F_{Meas}$ tended to overstate the load at the onset of failure.

To provide comparative data for one currently available technology, two QCT-based densitometry measures were examined. For the femora tested in the stance configuration, average density in the subcapital region ($\rho_{SC}$) was computed (Esses et al., 1989). For the femora tested in the fall configuration, intertrochanteric density ($\rho_{IT}$) multiplied by intertrochanteric area ($A_{IT}$) was obtained (Lotz and Hayes, 1990). These densitometry measures were selected a priori, and were computed using the same CT scan, calibration, and geometry data that were used to generate the FE models.

The abilities of the FE models and densitometry measures to predict $F_{Meas}$ were assessed separately for each loading condition by performing simple linear regression on $F_{Meas}$ vs $F_{FE}$, and on $F_{Meas}$ vs densitometry values. When the data appeared not to be linearly related and did not satisfy the assumption of homoscedasticity (tested by a Spearman rank correlation between the absolute values of the residuals and $F_{Meas}$), power relations were computed by performing simple linear regression on logarithmic transformations of the data.

For the FE approach, the data for the two loading conditions were compared by pooling the data and

---

1$F_{FE}$ was relatively insensitive to the number of contiguous elements required by the failure criterion as long as more than 8–10 elements were required. As the number of elements was increased from 11 to 15, $F_{FE}$ increased by 0.4–5% (mean, 2.7%) for the stance condition and 0–11% (mean, 5.1%) for the fall condition.
Table 1
Relationships used to compute modulus and strength from ash density

<table>
<thead>
<tr>
<th>Relationship</th>
<th>Density range</th>
<th>Correlation coefficient</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E = 33,900 \rho_{ash}^{2.20}$</td>
<td>$\rho_{ash} \leq 0.27$ g cm$^{-3}$</td>
<td>0.92</td>
<td>Keyak et al. (1994)</td>
</tr>
<tr>
<td>$E = 10,200 \rho_{ash}^{3.6}$</td>
<td>$\rho_{ash} \geq 0.6$ g cm$^{-3}$</td>
<td>0.82</td>
<td>Keller (1994)</td>
</tr>
<tr>
<td>$E = 5307 \rho_{ash} + 469^a$</td>
<td>$0.27 &lt; \rho_{ash} &lt; 0.6$ g cm$^{-3}$</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$S = 137 \rho_{ash}^{1.88}$</td>
<td>$\rho_{ash} &lt; 0.317$ g cm$^{-3}$</td>
<td>0.96</td>
<td>Keyak et al. (1994)</td>
</tr>
<tr>
<td>$S = 114 \rho_{ash}^{1.72}$</td>
<td>$\rho_{ash} \geq 0.317$ g cm$^{-3}$</td>
<td>0.94</td>
<td>Keller (1994)</td>
</tr>
</tbody>
</table>

$^a$ This equation represents linear interpolation between the extremes of the two previous relations for modulus.

$^b$ The strength correlations were similar and intersected at $\rho_{ash} = 0.317$ g cm$^{-3}$.

$E$ = elastic modulus (MPa); $S$ = strength (MPa); $\rho_{ash}$ = apparent ash density (g cm$^{-3}$).

performing backward stepwise regression ($F$-to-remove = 3.9, $p = 0.0565$; $F$-to-enter = 4.0, $p = 0.0535$). The initial regression equation took the form

$$F_{Meas} = b_0 + b_1 G + b_2 G F_{FE} + b_3 F_{FE},$$

where $G$ was a dummy variable equal to +1 for the stance configuration and -1 for the fall configuration, and $b_0$, $b_1$, $b_2$, and $b_3$ were determined by the regression analysis. Significance of the terms containing $G$ would indicate that the relationship between $F_{Meas}$ and $F_{FE}$...
depended on loading condition. The change in coefficient of determination \( R^2 \) caused by removal of terms during the stepwise regression indicated the amount of additional variance in \( F_{\text{meas}} \) that had been accounted for by those terms. Finally, to determine whether the use of two, theoretically non-independent data points from each donor influenced the final regression equation, an additional multiple regression analysis that accounted for between-subjects variability was performed (Glanz and Slinker, 1990), and the coefficients of the resulting equation and the original equation were compared using \( t \)-tests with \( z = 0.05 \).

Comparison of the CT/FE technique and densitometry methods utilized a limited data set of 18 pairs, so statistically significant differences between the methods were not expected. Rigorous statistical comparison of correlation coefficients for the two methods would have required over 100 pairs of specimens, assuming underlying correlation coefficients of 0.8 and 0.9, 80% statistical power, and \( z = 0.05 \) (Zar, 1984).

3. Results

\( F_{\text{meas}} \) and \( F_{\text{FE}} \) were linearly related for the stance configuration \( [r = 0.867, p < 0.0001, \text{standard error of the estimate (S.E.E.)} = 1.56 \text{kN}; \text{Fig. 4a}] \) and non-linearly related for the fall configuration \( [r = 0.949, p < 0.0001, \text{S.E.E.} = 0.0909 \log_{10}(\text{kN}); \text{Fig. 4b}] \). Logarithmic transformations of the data were required for the fall configuration due to both nonlinearity and heteroscedasticity of the data \( (p = 0.008) \). Data for one femur to be tested in the fall condition were lost due to mechanical error.

When the data for both loading conditions were pooled, logarithmic transformations were again necessary (Spearman test, \( p < 0.0001 \)), and backward stepwise regression revealed that the regression equation did not depend on loading condition \( (p = 0.08) \). More importantly, inclusion of the terms related to loading condition increased \( R^2 \) from 0.936 to 0.947, indicating that these terms accounted for only an additional 1.1% of the variance in \( F_{\text{meas}} \). In addition, the coefficients of the regression equation were not significantly different after accounting for between-subjects variability. Therefore, the data for both loading conditions were described by a single power relation between \( F_{\text{meas}} \) and \( F_{\text{FE}} \) \( [r = 0.967, p < 0.0001, \text{S.E.E.} = 0.0970 \log_{10}(\text{kN}); \text{Fig. 5}] \). This equation and the equation for the fall loading condition alone were nearly identical, further supporting the finding that the relationships for the stance and fall configurations were the same. Finally, the 95% confidence interval for an additional observation for this relationship spanned from \(-40\%\) to \(+60\%\) of \( F_{\text{meas}} \) (Fig. 5).

Significant correlations between \( F_{\text{meas}} \) and the densitometry data were found for the stance \( (r = 0.781, p = 0.0001, \text{S.E.E.} = 1.95 \text{kN}; \text{Fig. 6a}) \) and fall \( (r = 0.906, p < 0.0001, \text{S.E.E.} = 0.568 \text{kN}; \text{Fig. 6b}) \) loading conditions. These correlation coefficients were not significantly different from those for the CT/FE technique. However, statistical power was low, as explained previously.

4. Discussion

In an effort to improve clinical and research tools for assessing hip fracture risk, we have shown that automatically generated CT scan-based FE models can predict
femoral fracture load in vitro for two very different loading conditions with a precision of −40% to +60%. This level of precision is comparable to that of the densitometry-based measures examined here. With respect to accuracy, the CT/FE technique tends to underestimate $F_{\text{FE}}$ as indicated by the regression results. However, this underestimation can be attributed to the fact that $F_{\text{M}}$ was defined as the load at the first peak of the force-displacement curve. Therefore, the underestimation of $F_{\text{M}}$ should not be viewed as a limitation of the CT/FE technique, but rather as a result of the experimental method.

The precision of the CT/FE technique is comparable to that of densitometry methods reported previously. For the stance configuration, the correlation for the CT/FE technique ($r = 0.87$) is as strong as correlations found in previous studies using dual photon absorptiometry ($r = 0.74$) (Beck et al., 1990), QCT ($r = 0.79$) (Alho et al., 1988), and engineering-related methods ($r = 0.89$ in two studies) (Beck et al., 1990; Gies et al., 1985). For the fall configuration, the correlation coefficient for the CT/FE technique ($r = 0.95$) is statistically comparable to coefficients for the strongest correlations reported by Courtney et al. (1994, 1995) ($r = 0.80$–$0.96$), who used DXA and loading representing impact from a fall to the side.

It should be noted that the CT/FE technique is still in its infancy and significant improvements in precision may be achieved through refinements, most importantly in the area of material representation. For example, the current approach relied upon isotropic material properties that had been determined by testing specimens in compression in their most stiff and strong direction; element moduli and strengths for the other directions were therefore overstated. Use of anisotropic mechanical properties (Ciarelli et al., 1991; Keyak et al., 1994; Reilly and Burstein, 1975; Van Buskirk and Ashman, 1981) and accounting for the fact that bone material strength is about 30% lower in tension than in compression (Currey, 1970; Keaveny et al., 1994a,b; Reilly and Burstein, 1975) may improve FE model predictions. However, such

---

**Fig. 5.** Measured fracture load versus FE-predicted fracture load when data for both load configurations were pooled. A single power relation described the data for both configurations, indicating the robustness of the CT/FE technique. Circles and triangles indicate femora tested in the stance and fall configurations, respectively. Dashed lines indicate the bounds of the 95% confidence intervals for the regression equation (short dashes) and additional observations (long dashes).

**Fig. 6.** Measured fracture load versus subcapital density for the stance configuration (a), and measured fracture load versus intertrochanteric density multiplied by area for the fall configuration (b). The correlation coefficients for these regression equations are statistically comparable to those for the CT/FE technique (Fig. 4). Dashed lines indicate the bounds of the 95% confidence intervals for the regression equations (short dashes) and additional observations (long dashes).
improvements would require special failure theories, and none have been validated for bone at this time. A more practical first step toward improving model precision may be achieved by simply using smaller elements to better represent material heterogeneity and bone geometry. Finally, restriction of the analysis to linear material behavior, which maintained computational time at practical levels, was an important limitation because it precluded modeling past the onset of failure.

The ability of the FE models to account for loading condition represents a major advance over densitometry techniques, making it possible to calculate fracture loads for real-life loading conditions on a patient-specific basis. These fracture loads can then be compared with loads occurring in vivo in order to assess fracture risk, either for research or for clinical applications. Specifically, use of this technology in research may help us identify the most risky loading conditions and improve our understanding of the mechanisms of hip fracture so that preventive measures, either medical (Ettinger, 1985, 1993; Liberman et al., 1995; Reid et al., 1994; Stone, 1993) or biomechanical (Lauritzen et al., 1993), can be directed optimally.

The results of the present study are applicable to most of the elderly population because there were no restrictions based on medical history. Although all donors were Caucasian, lack of diversity should not affect generality of the results because the CT/FE technique can account for differences in bone size and density, the most important race-related parameters. Three of the donors died from cancer, and the cancer had metastasized extensively to the bones in one of these cases. In addition, it is likely that other femora came from donors with other diseases or treatments that could have affected bone strength. Despite these potentially confounding variables, there were no obvious outliers in the data. This result is especially surprising in the case of the femora with metastatic tumors, as it has not been shown that the material property relationships used in generating the FE models are valid for bone with tumors. Thus, these results suggest that CT scan-based FE models may be useful for assessing fracture load in highly pathological conditions.

This study has shown that automatically generated CT scan-based FE models can be used to estimate femoral fracture load under two very different loading conditions. However, the precision of this technique is comparable to that of densitometry methods such as DXA and QCT. Therefore, clinical use of this approach, which would require additional X-ray exposure and expenditure for a full CT scan of the femur, is not justified at this stage of development. Even so, the potential advantages of this CT/FE technique, i.e. its ability to account for the effects of 3-D geometry, material heterogeneity, and loading conditions, all at relatively low cost, support further research in this area.

Acknowledgements

This work was funded by the Department of Veterans Affairs, Rehabilitation R & D Service. Mechanical testing was performed at the University of California, San Francisco Orthopaedic Biomechanics Laboratory. Femora were provided by the University of California, San Francisco Tissue Bank; University of California, San Francisco Department of Anatomy; University of California, Irvine Organ and Tissue Bank; and National Disease Research Interchange. The authors thank Marina Pejic for her assistance.

References


