Published in final edited form as:

J Biomech. 2015 November 26; 48(15): 4043–4052. doi:10.1016/j.jbiomech.2015.09.044.

Fusion of Clinical and Stochastic Finite Element Data for Hip Fracture Risk Prediction

Peng Jiang^a, Samy Missoum^{a,*}, and Zhao Chen^b

^aAerospace and Mechanical Engineering Department, University of Arizona, Tucson, AZ, USA

^bMel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, AZ, USA

Abstract

Hip fracture affects more than 250,000 people in the US and 1.6 million worldwide per year. With an aging population, the development of reliable fracture risk models is therefore of prime importance. Due to the complexity of the hip fracture phenomenon, the use of clinical data only, as it is done traditionally, might not be sufficient to ensure an accurate and robust hip fracture prediction model. In order to increase the predictive ability of the risk model, the authors propose to supplement the clinical data with computational data from finite element models. The fusion of the two types of data is performed using deterministic and stochastic computational data. In the latter case, uncertainties in loading and material properties of the femur are accounted for and propagated through the finite element model. The predictive capability of a support vector machine (SVM) risk model constructed by combining clinical and finite element data was assessed using a Women's Health Initiative (WHI) dataset. The dataset includes common factors such as age and BMD as well as geometric factors obtained from DXA imaging. The fusion of computational and clinical data systematically leads to an increase in predictive ability of the SVM risk model as measured by the AUC metric. It is concluded that the largest gains in AUC are obtained by the stochastic approach. This gain decreases as the dimensionality of the problem increases: a 5.3% AUC improvement was achieved for a 9 dimensional problem involving geometric factors and weight while a 1.3% increase was obtained for a 20 dimensional case including geometric and conventional factors.

Keywords

Data fusion; Hip fracture prediction; Support vector machines; Finite element modeling

*Corresponding author at: Aerospace and Mechanical Engineering Department, University of Arizona, 1130 N. Mountain Ave., Tucson, AZ 85721, USA. Tel.: +1 (520) 626-5226, smissoum@email.arizona.edu (Samy Missoum).

7. Conflict of Interest

None of the authors has any conflict of interest regarding the work presented in this manuscript.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

Hip fracture is increasingly common in older populations and is an important cause of disability and mortality. It is estimated that 50% of women aged 50 or older may suffer from a fragility fracture in their remaining life (Jacobs, 2008). The economic and societal burdens associated with hip fracture will keep increasing. It is therefore of prime importance to develop prediction tools to identify the population at high risk of hip fracture and reduce its incidence.

The occurrence of hip fracture depends on a large number of factors, including age, Bone Mineral Density (BMD), weight, ethnicity (Benetos et al., 2007). In fact, some of the important factors have long been identified and risk models, such as FRAX (Kanis et al., 2008, 2005), have been developed and deployed on the Internet. Important advances in hip fracture prediction also include the use of the femur geometry derived through Dual-energy X-ray Absorptiometry (DXA) and a dedicated Hip Structure Analysis (HSA) algorithm (Beck et al., 1990; Beck, 2007).

Hip fractures have been the focus on many studies at the national (Burger, 2006; Diab and Watts, 2013) and international levels (Kannus et al., 1996; World Health Organization, 2003). Currently, risk models are solely based on clinical dataset such as the one provided by the Women's Health Initiative (WHI) (The WHI Study Group, 1998) which is used in this work. However, due to the large number of factors involved in the hip fracture process, clinical data alone, even of large size, might not be sufficient to ensure a robust and reliable hip fracture prediction model.

In order to improve the predictive ability of hip fracture prediction model, the authors propose to combine the clinical data with finite element analysis (FEA) data. The computational data complement the clinical data by providing mechanical quantities such as strains or stresses. By adding FEA data, the risk models are therefore no longer simply based on statistical correlation but also include a mechanical dimension with a well understood, although not exhaustive chain of causes and effects.

The computational data are generated using a fully parameterized finite element (FE) models of a femur which can represent a wide range of hip geometries and bone properties without having to build traditional patient-specific image-based models. In other words, the proposed FE model, although it is a simplified, can efficiently generate the femur models for all the participants within a cohort even in large number. This task would be very tedious if traditional imaging to FE model techniques were used.

In this work, the hip fracture risk model is constructed using a support vector machine (SVM) (Cristianini and Shawe-Taylor, 2000; Burges, 1998), which is a classification-based machine learning technique. Several other machine learning techniques could have been used (e.g., Random Forest or Neural Networks) but SVM has become an established and trusted tool for which the authors have gained substantial experience. The fusion between the clinical database and FEA is performed by complementing the clinical data with a selected FEA output which is considered as a new risk factor. This approach is referred to as the "augmented space" technique for which two versions are proposed. First, the new factor,

also referred to as dimension, is chosen as the FEA predicted maximum principal strain within the femur. In the second version of the approach, the maximum principal strain is also used but the effects of uncertainties in material properties and loading on the femur are accounted for. A stochastic quantity, in the form of a probability to exceed a given strain threshold, is added to the database. Both approaches are compared in terms of the improvement of predictive ability of the SVM-based risk model.

This article is organized as follows: A background section 2 provides information on the clinical data and the construction of risk models using SVM. Section 3 describes the fusion of clinical and computational data for hip fracture prediction. This section also explains the computation of the stochastic quantity from the femur FE model. The results section 4 lists the various experiments and compares the risk model's predictive abilities using the various approaches. The results are followed by a discussion section 5.

2. Background

2.1. Clinical dataset

The dataset used in this paper is a subgroup of the Women's Health Initiative (WHI) which is the largest US nationwide study of postmenopausal women. The subgroup has a total of 11,497 postmenopausal women with 186 hip fractures reported in the first 10 years and is characterized by the fact that its participants received BMD testing at 3 of the 40 WHI participating clinical centers (Pittsburgh, PA; Birmingham, AL; and sister facilities in Tucson and Phoenix, AZ). The average time to fracture is 7.4 years from the baseline measurements.

The WHI is a longitudinal study which is composed of a clinical trial (CT) component and an observational study (OS) component. The clinical trial component included a calcium plus vitamin D trial, a female hormone trial, and a low fat intervention trial. Participants, who were all volunteers, were recruited and screened according to the inclusion and exclusion criteria for the WHI OS and CT. The selection criterion for the CT participants was stricter than for the OS group. The inclusion criteria can be found in (Langer et al., 2003; The WHI Study Group, 1998). Further details about the WHI study design and data collection methods can be found in (Robbins et al., 2007; Ritenbaugh et al., 2003; Jackson et al., 2003; Stefanick et al., 2003).

In our study, the OS group is used for model development (training set) while the CT group, with different inclusion criteria than the OS group, is used for risk model validation. The two groups include common factors such as age, weight, BMD, etc. The database also includes hip geometry information extracted from each patient's DXA image using HSA (Figure 1). Table 1 lists all WHI clinical parameters used in this study. Details and baseline characteristics of the WHI dataset are provided in Table 2.

2.2. Hip fracture risk model based on support vector machine (SVM)

For data that are binary in nature (e.g., a hip is fractured or not), there are several choices of risk models such as logistic regression (Hosmer Jr and Lemeshow, 2004) which is one of the most widely used approaches. In this study, a more general classification-based model,

referred to as support vector machine (SVM), is used for fracture risk prediction (Cristianini and Shawe-Taylor, 2000; Burges, 1998). SVM is now a widely accepted machine learning technique that has been successfully applied to many fields (Basudhar et al., 2008; Zhang and Rajapakse, 2009; Tay and Cao, 2001; Basudhar and Missoum, 2010), including hip fracture prediction by the authors (Jiang et al., 2014). SVM has been used as a risk model in several other biomedical studies (Klement et al., 2014; Osowski et al., 2004; Soong et al., 2008).

In simple terms, an SVM defines an “optimal” boundary between samples belonging to two classes such as “fractured” and “non-fractured” in a d -dimensional space. Given a set of N training samples $\mathbf{x}_i$ with classes $y_i = \pm 1$, the general form of an SVM is:

$$s(\mathbf{x}) = b + \sum_{i=1}^N \lambda_i y_i K(\mathbf{x}_i, \mathbf{x}) = 0 \quad (1)$$

The classification of any arbitrary point $\mathbf{x}$ is given by the sign of $s(\mathbf{x})$. In 1, b is the “bias” and K is the kernel function which can have several forms such as the Gaussian kernel, which is used in this work:

$$K(\mathbf{x}_i, \mathbf{x}) = \exp \left(-\gamma \|\mathbf{x}_i - \mathbf{x}\|^2 \right), \quad \gamma > 0 \quad (2)$$

where γ is the width parameter of the Gaussian kernel. Beside the width parameter, SVM also depends on another parameter C , referred to as the cost coefficient which is important when the data is non-separable. More details are provided in Section A.1. Note that the SVM only provides a predicted class. In order to obtain a “probability of hip fracture” as the one obtained with a logistic regression model, several approaches can be used after the SVM model is constructed: probabilistic SVM (Platt, John, 1999) and bagging (Breiman, 1996). The probability information is not needed to gauge the predictive capability of the SVM model and will therefore not be used in this article.

2.3. Optimal SVM parameters

The SVM parameters γ and C are found by maximizing the mean of the area under the curve (AUC) obtained using ten-fold cross validation (McLachlan et al., 2004; Kohavi, 1995). The SVM is trained and the AUC calculated on the training (OS) group. In order to maximize the AUC, a grid search in a logarithmic space is used with the following ranges for $\log_2(C)$ and $\log_2(\gamma)$: $[-10, 19]$ and $[-25, 3]$ respectively.

3. Methodology

3.1. Combining clinical and computational data for hip fracture prediction

Due to the complexity of the hip fracture phenomenon, the use of clinical data for the construction of risk models might require very large datasets and cohorts. In addition, risk models based on “statistical correlation” provide little or no understanding on the influence of mechanical factors on the occurrence of hip fracture. To overcome or at least alleviate these obstacles, the authors propose to use computational data from computer simulations as

a supplement to the clinical dataset. The hypothesis is that such a hybrid model combining clinical and computational data will have a better predictive ability.

This paper proposes to combine both types of data by constructing an SVM-based prediction model in an “augmented space” where the clinical data are complemented by outputs from computational simulations. In other words, computational data are used as additional factors for hip fracture prediction (Fig. 2). The prediction model is thus constructed in a higher dimensional space.

In this article two variations of the “augmented space” approach will be provided. In one instance, the maximum principal strain as predicted by a finite element model (see Section 3.2) will be used as the added factor (see Fig. 2). The use of a strain-based criterion for fracture of the cortical bone has been reported in several studies (Schileo et al., 2008). In the second approach, uncertainties in mechanical quantities such as the loading on the femoral head and bone material properties are considered. These uncertainties are propagated through the FE model to provide a stochastic quantity such as the probability for the maximum principal strain to be larger than a threshold value. Although the propagation of these uncertainties is computationally involved, this technique might provide a more realistic description compared to a deterministic FEA output. Details about the stochastic approach are provided in Section 3.3.

3.2. A fully parameterized FE model of a femur

3.2.1. Geometry and meshing—In order to be able to assign computationally evaluated strain values to all the participants of the database, a fully parameterized three dimensional finite element model has been created using ANSYS’ APDL language (ANSYS Inc., 2013). For given values of geometric attributes depicted in Fig. 1, the APDL code builds and meshes a corresponding simplified femur geometry. The geometric assumptions used to build the three dimensional model are as follows: the shaft and neck are circular, the larger trochanter is elliptical. These assumptions are consistent with the DXA geometric assumptions (Beck, 2007). Details about the geometry of the femur FE model and the section characteristics are provided in Fig. 3. The length of the femoral shaft is assumed constant at 10 cm. The mesh is composed of 12200 hexahedral solid elements for most of the femur and 4147 tetrahedral elements for the femoral head. The number of elements was selected by studying the convergence of the maximum principal strain which is used in this study as an indicator for fracture.

Fig. 4 provides an example of the flexibility of the approach through three automatically generated femur models with different geometries. Note that this model is not limited to the participants geometry and could represent any individual.

3.2.2. Material properties—Transversely isotropic material is used to simulate the cortical bone. That is, the behavior is isotropic within any plane transverse to the longitudinal direction of the bone, which was implemented through the construction of local coordinate systems (see Fig. 4(c)). From (Reilly et al., 1974), the mean values of the Young’s modulus of the cortical bone in the longitudinal direction is 17 GPa. In the transverse direction, the Young’s modulus of the cortical bone is 11.5 GPa. For the

trabecular bone, the Young's modulus is 1 GPa. The longitudinal direction is dictated by the coordinate systems depicted in Fig. 4 (c). In Section 3.3, the values of the material properties will no longer be assumed constant but varied following specific distributions.

3.2.3. Loading and boundary conditions—In this study and for simplicity, the loading scenario chosen in this study is a single-limb stance. The load transmitted from the pelvis to the femoral head is assumed to have a magnitude of 2.5 times the body weight (Bergmann et al., 2001). The load direction is set at a default value of 16° with respect to the shaft axis but will also be varied in this study (Hölzer et al., 2013) (see Section 3.3). The load is applied to the top half of the femoral head, modeled as a hemisphere, using nodal forces. The base of the femoral shaft is assumed clamped.

3.2.4. Output from FE model—In this work, the maximum principal strain is used as a predictor for fracture (Schileo et al., 2008). Other measures and failure criteria could have been chosen (e.g., stress-based measures), but the focus of this article is to demonstrate initial results for the fusion of computational and clinical data. An example of maximum principal strain contour obtained with the parameterized finite element model is provided in Fig. 4(d).

3.3. Stochastic FE output

The deterministic strain (ε_D) from FEA assumes that quantities such as loads and bone material properties are exactly known. However, for each individual there might be large uncertainties which can only be described in a statistical manner. For this reason, the second approach presented in this article is to increase the dimensionality of the space using a stochastic quantity. The quantity chosen is the probability that the maximum principal strain is larger than a threshold value. This probability is calculated by assuming that the load direction and the material properties of the trabecular and cortical bones are in fact random variables with known distributions.

3.3.1. Strain distribution for a given participant—In order to obtain a distribution of the maximum principal strain, Monte-Carlo (MC) simulation is used. Fig. 5 provides an example of distribution for a given participant. However, this would not be practical if the actual finite element model was repeatedly called. For this purpose the maximum principal strain is approximated by a Kriging surrogate (Cressie, 1990) and trained with a design of experiments (Sacks et al., 1989). Because the Kriging approximation is efficiently evaluated, MC simulations can be performed efficiently, thus enabling the propagation of uncertainties in load and material properties. In this work, the Kriging surrogate is constructed from a Latin Hypercube sampling (LHS) design of experiments (DOE) in the space of all the chosen factors and the random factors. Each DOE sample is evaluated using the femur FE model. The total number of DOE samples needed is selected so that the accuracy of the Kriging approximation, as calculated on a validation set from the DOE, converges. For the interested reader, further technical details and results are provided in A.2.

3.3.2. Probability calculation—Based on a given strain threshold θ , the probability that the maximum principal strain ($\varepsilon_{\max}$) is larger than θ is noted $P_f(\varepsilon_{\max}, \theta)$ and can be

calculated from the strain distribution. This probability is then used in the augmented space approach as an additional factor replacing the deterministic FEA strain prediction ε_D :

$$P_f(\varepsilon_{\max}, \theta) \approx \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} I_g(\tilde{\varepsilon}_{\max}^{(i)}, \theta) \quad (3)$$

where N_{MC} is the number of MC samples. $\tilde{\varepsilon}_{\max}^{(i)}$ is the approximation of the maximum principal strain for the i^{th} MC sample using the Kriging approximation (see Section A.2). I_g is an indicator function that is defined as:

$$I_g(x, \theta) = \begin{cases} 1 & \text{if } x \geq \theta \\ 0 & \text{if } x < \theta \end{cases} \quad (4)$$

Fig. 5 also depicts an example of P_f calculated for an arbitrary participant in the WHI dataset.

3.3.3. Determination of the threshold θ —The strain threshold θ must be carefully chosen since it dictates the values of P_f and eventually the predictive ability of the SVM risk model. The choice of the threshold could be made on mechanics considerations. However the complexity and large variability from participant to participant in the database makes this task unrealistic. In this work, the threshold is chosen so as to maximize the AUC calculated based the probability P_f taken as a single factor. It is noteworthy that this is the probability calculated based on Monte-Carlo simulations as described in the previous section and not a probability obtained from a risk model. In other words, the AUC calculated to find the threshold is independent of the risk model chosen. Formally, the optimal value θ^* is obtained by solving the following one-dimensional optimization problem:

$$\theta^* = \arg \max_{\theta} AUC_{P_f}(\theta) \quad (5)$$

Once the optimal strain threshold θ^* is obtained, the corresponding patient-specific probability can be calculated. In the remainder of this paper, the probability $P_f(\varepsilon_{\max}, \theta^*)$, will be referred as the stochastic quantity ε_S .

4. Results

In order to assess the gains in predictive capability through the fusion of clinical and computational data, several results are computed:

1. AUC for SVM risk model constructed using the WHI database only.
2. AUC when deterministic maximum principal strain (ε_D) is added as a factor.
3. AUC when the stochastic quantity (ε_S) as constructed in Section 3.3.3 is used instead of the deterministic strain (ε_D).

Based on the parameters listed in Table 1, five different experiments using various choices of parameters will be used:

1. Geometric parameters and weight
2. Conventional parameters without BMD
3. Conventional parameters
4. Geometric parameters, and conventional parameters without BMD
5. Geometric parameters, and conventional parameters

Although there are a total of 11497 participants in the WHI dataset, 6495 of them have missing values for at least one of the parameters listed in Table 1. For this reason, the dataset was reduced to participants with all parameter values available. This led to a set with 5002 participants of which 122 have a fractured hip within 10 years. This database is split into a training set with 2889 participants (84 fractures) and a validation set with 2113 participants (38 fractures).

4.1. Predictive ability of the FE model

Before performing the experiments, it is important to validate the FE model and quantify the quality of its prediction. The FE model does not require training data but its outcome can be tested against the outcomes (fracture or not) of the clinical data (5002 participants). This is done by using the fully parameterized model of the femur which provides the maximum principal strain for all the participants of the WHI dataset automatically. Once all the strains are obtained, they are ranked thus allowing to evaluate the AUC of the FE model against the outcomes of the WHI database (see Section 2.1). The distribution of strain is depicted in Fig. 6 as well as the corresponding ROC curve. The AUC is 74% (95% CI: [0.70, 0.78]).

4.2. Fusion of clinical and computational data. Deterministic approach

As described in Section 3.1, the maximum principal strain ε_D computed from FEA is used as an additional factor for the construction of the risk model in an “augmented space” as shown in Fig. 2. This augmented space has therefore a dimensionality of $N + 1$, where N is the number of clinical parameters involved in each experiment. In this first set of results, the maximum principal strain is assumed deterministic. It is computed from the FE model by considering that all parameters are fixed.

In order to characterize the importance of the newly added strain factor, its means for the fractured and non-fractured population are compared. For the fractured population, the mean is $\mu_{\varepsilon_D} = 0.65\%$ and $\mu_{\varepsilon_D} = 0.53\%$ for the non-fractured population. Therefore there is a 22% increase in strain between the non-fractured and fractured population (the corresponding p value is 2.4×10^{-5}). This justifies the use of the maximum principal strain as a computational factor.

Table 3 lists the changes in predictive ability as measured by the AUC between the SVM risk model constructed using the WHI dataset only and the SVM trained using both clinical and computational data. Using the deterministic approach, the SVM predictive capability increases by 3.6% and 1.5% in experiments I and II respectively. However, no noticeable increase is observed for experiments III through V. These results confirm, as would be

expected, that the efficiency of the deterministic approach reduces as the dimensionality of the problem increases.

4.3. Fusion of clinical and computational data. Stochastic approach

This section uses a stochastic quantity ε_S instead of the deterministic strain values. As a demonstrative example, uncertainties from bone material properties in the form of their Young's moduli and the loading direction are considered. Normal distributions are assigned to load angle ($LA \sim \mathcal{N}(16, 3^2)$ degrees), Young's moduli of the cortical ($E_c \sim \mathcal{N}(17, 3^2)$ GPa) and the trabecular bone ($E_t \sim \mathcal{N}(1, 0.2^2)$ GPa). The notation $\mathcal{N}(\mu, \sigma^2)$ indicates that the variable follows a normal distribution of mean μ and standard deviation σ .

As describes in Section 3.3, the stochastic quantity chosen is the probability P_f that the strain is larger than a given threshold value. The threshold value θ^* is chosen so as to maximize the AUC when using P_f calculated with 10^4 MC samples for each WHI participant, as the only predictor (see Fig. 7(a)).

Once the strain threshold θ^* is found, the corresponding P_f is denoted as ε_S . The mean of ε_S for the fractured and non-fractured populations are $\mu_{\varepsilon_S} = 1.87\%$ and $\mu_{\varepsilon_S} = 0.53\%$ respectively. Therefore the mean of the stochastic measure is more than 2.5 times larger in the fracture population. The corresponding p value is 1.5×10^{-18} .

The results show (Table 3) that the stochastic approach increases AUC scores by 5.3%, 4.5%, 2.0%, 1.5%, and 1.3% for experiments I through V. In addition, these increases are larger than the ones obtained with the deterministic approach. As in the deterministic case, the gains in predictive ability decrease with the dimensionality. As an example, the ROC curve corresponding to experiment V with stochastic strain is provided in Fig. 7(b).

5. Discussion

In this attempt to combine clinical and computational data to increase the predictive ability of the hip fracture risk model, several algorithmic and computational choices have been made and are discussed here:

- **What is the level of computational fidelity needed?** In this article a three dimensional linear elastic finite element model of the femur using volume elements was used to generate computational data. However, for the sake of comparison, a simplified two-dimensional beam finite element was also used to obtain the strain in a side study. It was observed that the predictive abilities of the volume and beam finite element models calculated by the AUC on the WHI dataset were almost identical. However, the strain values predicted are different. The reason why the AUC values are similar is that the AUC is only a function of the ordering of the strain values and not their actual values. In other words, the beam and volume models tend to rank the strain values of the database participants, in the same way. For this reason, the next steps of this research will involve the beam model for ease of computation but also use information from a more complex volume finite element model (e.g., multi-scale modeling involving volume fractions). This multi-

fidelity approach might provide the best approach to improve the predictive ability with reasonable computational efficiency.

- **Failure modes.** In this article, the maximum principal strain has been used as a failure criterion. Several other failure criteria, stress-based, could be used as reviewed in (Doblaré and García, 2003). Because a classification approach such as SVM is used for the construction of the risk model, several criteria can be used simultaneously without creating separate risk models for each criterion.
- **Sources of uncertainties.** Sources of uncertainties have been added at the computational level in addition to the inherent uncertainties originating from the clinical data. The Young's moduli of the trabecular and cortical bones as well as the direction of the load has been used. The use of uncertainties is essential since quantities such as loading and material properties cannot be known exactly but only statistically. Several sources of uncertainties could be added associated with a more sophisticated finite element model as well as other load cases (e.g., dynamic loadings to simulate a fall). The current results indicate that the use of a stochastic quantity involving these uncertainties enables a marked increase in AUC. Therefore, the propagation of other uncertainties needs to be investigated as it might further contribute to the increase in predictive ability.
- **Missing data.** In the database, not all patients have all the data. In order to compare all the results in the consistent manner, we have used the minimum set of data corresponding to the largest number of factors. In order to mitigate this problem and obtain more data, we could have used techniques for missing data such as imputation. However, this was not done in this study whose purpose is primarily to demonstrate the potential increase in predictive ability by combining clinical and computational data.
- **Limitations of the data fusion approach.** Although improvements in predictive ability have been achieved using the augmented space approach, the dimensionality of the space has been increased by one factor while maintaining the same number of database participants. As a consequence, the improvements in predictive ability due to the deterministic and stochastic approaches tend to reduce as the dimensionality of the problem increases from 9 to 20. This reduction in efficacy, which was expected, could be mitigated by removing the less influential factors (dimensionality reduction) and/or by adding different computational information (other than strain) to the database. This last aspect is being investigated as well as another approach that would not change the dimensionality of the space.

6. Conclusion

This work introduces a novel approach for the construction of a more robust and reliable hip fracture risk prediction model by supplementing clinical data with finite element data. To combine the two types of data, two approaches are proposed. A deterministic approach is first applied, which uses direct FE strain outputs as an additional parameter for risk model construction. The second approach fuses stochastic FE outputs and clinical data by propagating uncertainties in loads and material properties. It was observed that although the

deterministic approach tends to improve the predictive ability of the risk model, the stochastic approach further improves the predictive capability of the hip fracture prediction model.

The next step of this research will investigate further the data fusion scheme. Specifically, the authors will investigate what type of computational information, beyond strain, might be added to further improve the predictive ability of the risk model. In addition, other sources of uncertainty to be propagated through the finite element model will be included. Based on the risk models developed through this research, the authors plan to release an online hip fracture risk calculator.

Acknowledgments

Research reported in this publication was supported by NIAMS of the National Institutes of Health under award number 1R21AR060811. Their support is gratefully acknowledged. The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C.

References

- ANSYS Inc. ANSYS Mechanical APDL and Mechanical Applications Theory Reference. 2013. Release 15.0
- Basudhar A, Missoum S. An improved adaptive sampling scheme for the construction of explicit boundaries. *Structural and Multidisciplinary Optimization*. 2010; 42(4):517–529.
- Basudhar A, Missoum S, Harrison Sanchez A. Limit state function identification using Support Vector Machines for discontinuous responses and disjoint failure domains. *Probabilistic Engineering Mechanics*. 2008; 23(1):1–11.
- Beck TJ. Extending DXA beyond bone mineral density: understanding hip structure analysis. *Current osteoporosis reports*. 2007; 5(2):49–55. [PubMed: 17521505]
- Beck TJ, Ruff CB, Warden KE, Scott WW, Rao GU. Predicting femoral neck strength from bone mineral data. A structural approach. *Investigative radiology*. 1990; 25(1):6–18. [PubMed: 2298552]
- Benetos IS, Babis GC, Zoubos AB, Benetou V, Soucacos PN. Factors affecting the risk of hip fractures. *Injury*. 2007; 38(7):735–744. [PubMed: 17307178]
- Bergmann G, Deuretzbacher G, Heller M, Graichen F, Rohlmann A, Strauss J, Duda G. Hip contact forces and gait patterns from routine activities. *Journal of Biomechanics*. 2001; 34(7):859–871. [PubMed: 11410170]
- Breiman L. Bagging predictors. *Machine Learning*. 1996; 24(2):123–140.
- Burger HG. Hormone therapy in the WHI era. *Australian and New Zealand Journal of Obstetrics and Gynaecology*. 2006; 46(2):84–91. [PubMed: 16638027]
- Burges CJ. A tutorial on Support Vector Machines for pattern recognition. *Data mining and knowledge discovery*. 1998; 2(2):121–167.
- Chang CC, Lin CJ. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*. 2011; 2:1–27.
- Cressie N. The origins of kriging. *Mathematical Geology*. 1990; 22(3):239–252.
- Cristianini, N.; Shawe-Taylor, J. An introduction to Support Vector Machines and other kernel-based learning methods. Cambridge university press; 2000.
- Diab DL, Watts NB. Postmenopausal osteoporosis. *Current Opinion in Endocrinology, Diabetes and Obesity*. 2013; 20:501–509.
- Doblaré M, García J. On the modelling bone tissue fracture and healing of the bone tissue. *Acta Científica Venezolana*. 2003; 54(1):58–75. [PubMed: 14515768]

- Glantz, S.; Slinker, B. *Primer of Applied Regression & Analysis of Variance*. McGraw-Hill Education; 2000.
- Hölzer A, Schröder C, Woiczinski M, Sadoghi P, Scharpf A, Heimkes B, Jansson V. Subject-specific finite element simulation of the human femur considering inhomogeneous material properties: a straightforward method and convergence study. *Computer methods and programs in biomedicine*. 2013; 110(1):82–88. [PubMed: 23084242]
- Hosmer, DW., Jr; Lemeshow, S. *Applied logistic regression*. John Wiley & Sons; 2004.
- Jackson RD, Lacroix AZ, Cauley JA, McGowan J. The women's health initiative calcium/vitamin d trial: overview and baseline characteristics of participants. *Annals of Epidemiology*. 2003; 13(9, Supplement):S98–S106. [PubMed: 14575942]
- Jacobs, JJ. *The Burden of Musculoskeletal Diseases in the United States*. American Academy of Orthopaedic Surgeons; 2008. p. 247
- Jiang P, Missoum S, Chen Z. Optimal SVM parameter selection for non-separable and unbalanced datasets. *Structural and Multidisciplinary Optimization*. Oct; 2014 50(4):523–535. [PubMed: 25258621]
- Kanis JA, Borgstrom F, De Laet C, Johansson H, Johnell O, Jonsson B, Oden A, Zethraeus N, Pflieger B, Khaltav N. Assessment of fracture risk. *Osteoporosis International*. 2005; 16(6):581–589. [PubMed: 15616758]
- Kanis, JA., et al. *Assessment of osteoporosis at the primary health care level*. WHO Collaborating Centre, University of Sheffield; 2008.
- Kannus P, Parkkari J, Sievänen H, Heinonen A, Vuori I, Järvinen M. Epidemiology of hip fractures. *Bone*. 1996; 18:57S–63S. [PubMed: 8717549]
- Klement RJ, Allgäuer M, Appold S, Dieckmann K, Ernst I, Ganswindt U, Holy R, Nestle U, Nevinny-Stickel M, Semrau S, et al. Support vector machine-based prediction of local tumor control after stereotactic body radiation therapy for early-stage non-small cell lung cancer. *International Journal of Radiation Oncology Biology Physics*. 2014; 88(3):732–738.
- Kohavi, R. *International joint conference on artificial intelligence*. Vol. 14. Lawrence Erlbaum Associates Ltd; 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection; p. 1137–1145.
- Langer RD, White E, Lewis CE, Kotchen JM, Hendrix SL, Trevisan M. The women's health initiative observational study: baseline characteristics of participants and reliability of baseline measures. *Annals of Epidemiology*. 2003; 13(9, Supplement):S107–S121. [PubMed: 14575943]
- McLachlan, GJ.; Do, K-A.; Ambrose, C. *Analyzing microarray gene expression data*. Vol. 422. Wiley-Interscience; 2004.
- Osowski S, Hoai LT, Markiewicz T. Support Vector Machine-Based Expert System for Reliable Heartbeat Recognition. *IEEE Transactions on Biomedical Engineering*. 2004; 51(4):582–589. [PubMed: 15072212]
- Osuna, E.; Freund, R.; Girosi, F. *Support Vector Machines: training and applications*. 1997.
- Platt, John. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*. 1999; 10(3):61–74.
- Reilly DT, Burstein aH, Frankel VH. The elastic modulus for bone. *Journal of biomechanics*. 1974; 7(3):271–275. [PubMed: 4846264]
- Ritenbaugh C, Patterson RE, Chlebowski RT, Caan B, Fels-Tinker L, Howard B, Ockene J. The women's health initiative dietary modification trial: overview and baseline characteristics of participants. *Annals of Epidemiology*. 2003; 13(9, Supplement):S87–S97. [PubMed: 14575941]
- Robbins J, Aragaki AK, Kooperberg C, Watts N, Wactawski-Wende J, Jackson RD, LeBoff MS, Lewis CE, Chen Z, Stefanick ML, Cauley J. Factors associated with 5-year risk of hip fracture in postmenopausal women. *JAMA: the journal of the American Medical Association*. 2007; 298(20):2389–2398. [PubMed: 18042916]
- Sacks J, Welch WJ, Mitchell TJ, Wynn HP. Design and analysis of computer experiments. *Statistical science*. 1989:409–423.
- Schileo E, Taddei F, Cristofolini L, Viceconti M. 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. *Journal of Biomechanics*. 2008; 41(2):356–367. [PubMed: 18022179]

Soong TT, Wrzeszczynski KO, Rost B. Physical protein-protein inter-actions predicted from microarrays. *Bioinformatics*. 2008; 24(22):2608–2614. [PubMed: 18829707]

Stefanick ML, Cochrane BB, Hsia J, Barad DH, Liu JH, Johnson SR. The women's health initiative postmenopausal hormone trials: overview and baseline characteristics of participants. *Annals of Epidemiology*. 2003; 13(9, Supplement):S78–S86. [PubMed: 14575940]

Tay FE, Cao L. Application of support vector machines in financial time series forecasting. *Omega*. 2001; 29(4):309–317.

The WHI Study Group. Design of the Women's Health Initiative clinical trial and observational study. Controlled clinical trials. 1998; 19(1):61–109. [PubMed: 9492970]

Vapnik, V. The nature of statistical learning theory. Springer; 1999.

World Health Organization. Diet, nutrition and the prevention of chronic diseases. Tech rep. 2003

Zhang, Y.; Rajapakse, JC. Machine learning in bioinformatics. Vol. 4. John Wiley & Sons; 2009.

A. Appendix

A.1. Support Vector Machine

Clinical data is often non-separable and unbalanced. In other words, there are far more non-fractured participants than fractured ones. In order to balance the data, Osuna and Vapnik (Osuna et al., 1997; Vapnik, 1999) proposed using different cost coefficients (i.e., weights) for the different classes in the SVM formulation. In the linear case, the optimization problem becomes:

$$\begin{aligned} \min_{\mathbf{w}, \xi, b} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C^+ \sum_{i=1}^{N^+} \xi_i + C^- \sum_{i=1}^{N^-} \xi_i \\ \text{s.t.} \quad & y_i(\mathbf{w} \cdot \mathbf{x}_i - b) \geq 1 - \xi_i \\ & \xi_i \geq 0, i=1, \dots, N \end{aligned} \quad (6)$$

where C^+ and C^- are cost coefficients for +1 and -1 class separately. N^+ and N^- are numbers of samples from +1 and -1 classes. The coefficients are typically chosen as (Chang and Lin, 2011):

$$\begin{aligned} C^+ &= C \times w^+ \\ C^- &= C \times w^- \end{aligned} \quad (7)$$

where C is the common cost coefficient for both classes, w^+ and w^- are the weights for +1 and -1 class respectively. In this formulation, used in the present study, the weights are chosen as $w^+ = 1$ and $w^- = N^+/N^-$.

A.2. Kriging

In the stochastic augmented space approach, the probabilities are calculated through MC simulations using a Kriging approximation (Sacks et al., 1989) of the maximum principal strain. A Kriging approximation is the sum of an “average” (trend) approximation (typically a low order polynomial) and a Gaussian process. In this work, the Kriging surrogate is trained using a Latin Hypercube Sampling design of experiments in a 12 dimensional space (nine FE implemented parameters listed in Table 1 as well as the three random variables

discussed in Section 4.3). Additional samples are also used for validation of the Kriging approximation.

To determine a reasonable number of samples for Kriging construction, a convergence study of error metrics as a function of the number of training samples is carried out. 1500 validation samples are used. The coefficient of determination (R^2 , Eqn. 8) and the relative maximum absolute error (Glantz and Slinker, 2000) ($RMAE$, Eqn. 9) are used to quantify the Kriging generalization error. R^2 is a global measure of the accuracy between the actual and predicted responses while $RMAE$ indicates the maximum local error. With 190 samples, the constructed Kriging can achieve more than 0.99 for R^2 while with 1075 samples, $RMAE$ can be reduced to less than 0.25 based on the validation set. Thus, with 1075 or more samples, the Kriging model is accurate both globally and locally.

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2} \quad (8)$$

$$RMAE = \max_{i=1, \dots, N} |y_i - \hat{y}_i| / \sigma(y) \quad (9)$$

Once the Kriging model is built, it is used to find the probability that the strain is larger than the strain threshold.

A.3. Baseline characteristics

The baseline characteristics of the study population are listed in Table 2.

Figure 1.
Femoral geometry and corresponding parameters.

Figure 2. Combining clinical and computational data for hip fracture prediction in the “augmented space”. Computational data are used as additional factors (e.g., strain from FEA).

Figure 3.

Geometry of the three dimensional femur FE model. The shaft and neck are assumed circular while the intertrochanter regions are composed of elliptical sections. The major axes of the trochanter sections are controlled by the distance from the ellipse centers C_i to the origin of the local coordinate systems O_i ($C_1O_1 : C_2O_2 : C_3O_3 : C_4O_4 : C_5O_5 = 10\% : 25\% : 65\% : 100\% : 110\%$). $O_1A_1 = O_2A_2 = O_3A_3 = O_4A_4 = O_5A_5 = R_5$, where R_5 is the radius of femoral shaft. The minor axis for the greater trochanter was arbitrarily set to 20% larger than the radius of the femoral shaft.

Figure 4.

Fully-parameterized FE model of the femur (a). Exaggerated hip geometry with large neck-shaft angle to demonstrate the flexibility of the FE model. (c) Local coordinate systems for implementation of the transverse isotropy of the cortical bone. (d) Contour of computed maximum principal strain at the femoral neck.

Figure 5.

Distribution of maximum principal strain considering uncertainties in load angle and material properties of the trabecular and cortical bone.

Figure 6.

(a) Histogram of maximum principal strain from the FE model for all the participants in the WHI dataset. (b) Predictive ability of the FE model based on the WHI dataset.

Figure 7.

(a) Evolution of predictive ability of the femur FE model with uncertainty as the strain threshold θ increases. This threshold is used to calculate the femur fracture risk for each individual using Monte Carlo simulation. The max AUC is 77.8% and the corresponding threshold of maximum principal strain is 0.84%. (b) ROC curve for experiment V with ε_S (20 clinical factors).

Table 1

Clinical parameters for risk model construction. Baseline values.

Category	Parameter	Used in FEA	Individual factor AUC
Conventional parameters*	Ethnicity		0.55
	Self-reported health		0.62
	Any fracture on/after age 55		0.52
	Physical activity		0.55
	Smoking status		0.51
	Parent broke hip		0.52
	Corticosteroid use		0.51
	Diabetes treatment		0.53
	Age		0.77
	Height		0.52
	Weight	Yes	0.58
Hip geometry	Hip total BMD		0.77
	Neck-shaft angle	Yes	0.53
	Neck length	Yes	0.53
	Outer diameter of NN	Yes	0.63
	Cortical bone thickness of NN	Yes	0.77
	Outer diameter of IT	Yes	0.65
	Cortical bone thickness of IT	Yes	0.75
	Outer diameter of S	Yes	0.61
	Cortical bone thickness of S	Yes	0.71

* The first 11 conventional parameters are selected based on (Robbins et al., 2007).

† NN: femoral narrow neck, IT: inter-trochanter, and S: shaft.

Table 2

Baseline characteristics of the WHI clinical parameters used in this study.

Parameter	Training		Validation	
	No fracture (N=2805)	Fracture (N=84)	No fracture (N=2075)	Fracture (N=38)
Ethnicity	American Indian/ Alaskan native	26 (0.9%)*	0 (0%)	1 (2.6%)
	Asian/Pacific islander	6 (0.2%)	0 (0%)	0 (0%)
	African-American	256 (9.1%)	5 (6.0%)	349 (16.8%)
	Hispanic/Latino	153 (5.5%)	3 (3.6%)	103 (5.0%)
	Non-Hispanic white	2350 (83.8%)	75 (89.3%)	1590 (76.6%)
	Other	14 (0.5%)	1 (1.2%)	13 (0.6%)
Self-reported health	Excellent	509 (18.1%)	8 (9.5%)	308 (14.8%)
	Very good	1097 (39.1%)	18 (21.4%)	795 (38.3%)
	Good	925 (33.0%)	46 (54.8%)	772 (37.2%)
	Fair	256 (9.1%)	11 (13.1%)	190 (9.2%)
	Poor	18 (0.6%)	1 (1.2%)	10 (0.5%)
				0 (0%)
Fracture on/after age 55	No	1951 (69.6%)	54 (64.3%)	1445 (69.6%)
	Yes	367 (13.1%)	26 (31.0%)	251 (12.1%)
	N/A	487 (17.4%)	4 (4.8%)	379 (18.3%)
Physical activity (METs/week)	0, inactive	418 (14.9%)	17 (20.2%)	435 (21.0%)
	<5	574 (20.5%)	23 (27.4%)	516 (24.9%)
	5 – 12	646 (23.0%)	15 (17.9%)	501 (24.1%)
	>= 12	1167 (41.6%)	29 (34.5%)	623 (30.0%)
				8 (21.1%)
Smoking status	Nonsmoker	1514 (54.0%)	49 (58.3%)	1192 (57.4%)
	Past smoker	1105 (39.4%)	27 (32.1%)	732 (35.3%)
	Current smoker	186 (6.6%)	8 (9.5%)	151 (7.3%)
Parent broke hip	No	1706 (60.8%)	52 (61.9%)	1244 (60.0%)
	Yes	1099 (39.2%)	32 (38.1%)	831 (40.0%)
Corticosteroid use	No	2783 (99.2%)	81 (96.4%)	2069 (99.7%)
	Yes	22 (0.8%)	3 (3.6%)	6 (0.3%)
				0 (0%)

Parameter	Training			Validation		
	No fracture (N=2805)	Fracture (N=84)	No fracture (N=2075)	Fracture (N=38)	No fracture (N=2075)	Fracture (N=38)
Diabetes treatment	No	2716 (96.8%)	78 (92.9%)	1952 (94.1%)	31 (81.6%)	7 (18.4%)
	Yes	89 (3.2%)	6 (7.1%)	123 (5.9%)	7 (18.4%)	7 (18.4%)
Age (yr)		62.0 (7.0)	68.8 (6.9)	62.0 (7.2)	69.5 (5.7)	69.5 (5.7)
Height (cm)		161.8 (6.6)	161.9 (7.7)	161.8 (6.1)	161.3 (6.6)	161.3 (6.6)
Weight (kg)		71.4 (15.8)	68.0 (13.0)	76.1 (15.4)	71.4 (14.0)	71.4 (14.0)
Hip total BMD (g/cm ²)		0.85 (0.1)	0.7 (0.1)	0.9 (0.1)	0.7 (0.1)	0.7 (0.1)
Neck-shaft angle (°)		130.7 (5.0)	131.9 (5.7)	130.7 (4.9)	132.4 (4.9)	132.4 (4.9)
Neck length (cm)		3.0 (0.2)	3.1 (0.3)	3.0 (0.2)	3.1 (0.2)	3.1 (0.2)
Outer diameter of NN [†] (cm)		0.14 (0.03)	0.1 (0.02)	0.1 (0.03)	0.1 (0.02)	0.1 (0.02)
Cortical bone thickness of NN (cm)		4.7 (0.5)	4.7 (0.6)	4.7 (0.5)	4.7 (0.6)	4.7 (0.6)
Outer diameter of IT (cm)		5.1 (0.3)	5.3 (0.4)	5.1 (0.3)	5.2 (0.3)	5.2 (0.3)
Cortical bone thickness of IT (cm)		0.3 (0.1)	0.2 (0.05)	0.3 (0.06)	0.2 (0.04)	0.2 (0.04)
Outer diameter of S(cm)		2.8 (0.2)	2.9 (0.2)	2.8 (0.2)	2.9 (0.2)	2.9 (0.2)
Cortical bone thickness of S (cm)		0.4 (0.1)	0.4 (0.08)	0.4 (0.08)	0.3 (0.06)	0.3 (0.06)

* Mean (standard deviation) for continuous variables.
Count (percentage) for categorical variables.

[†] NN: femoral narrow neck, IT: inter-trochanter, and S: shaft.

Table 3
Predictive ability of SVM-based risk model for 4 configurations of parameters (see Table 1).

Exp	# Param.	Parameters	Validation dataset		
			AUC	95% CI	Increase
I	9	Geometric parameters + weight	0.773	[0.71, 0.83]	-
		Geometric parameters + weight + ε_D	0.809	[0.74, 0.87]	0.036
		Geometric parameters + weight + ε_S	0.826	[0.76, 0.89]	0.053
II	11	Conventional parameters w/o BMD	0.819	[0.74, 0.88]	-
		Conventional parameters w/o BMD + ε_D	0.834	[0.76, 0.89]	0.015
		Conventional parameters w/o BMD + ε_S	0.864	[0.80, 0.92]	0.045
III	12	Conventional parameters	0.853	[0.80, 0.90]	-
		Conventional parameters + ε_D	0.861	[0.82, 0.92]	0.008
		Conventional parameters + ε_S	0.873	[0.83, 0.93]	0.020
IV	19	Geometric + conventional w/o BMD	0.856	[0.79, 0.91]	-
		Geometric + conventional w/o BMD + ε_D	0.857	[0.79, 0.91]	0.001
		Geometric + conventional w/o BMD + ε_S	0.871	[0.81, 0.93]	0.015
V	20	Geometric + conventional parameters	0.868	[0.81, 0.92]	-
		Geometric + conventional parameters + ε_D	0.870	[0.81, 0.92]	0.002
		Geometric + conventional parameters + ε_S	0.881	[0.83, 0.92]	0.013