Clinical Orthopaedics and Related Research®
A Publication of The Association of Bone and Joint Surgeons®

CrossMark

Preventing Heterotopic Ossification in Combat Casualties— Which Models Are Best Suited for Clinical Use?

Keith A. Alfieri MD, Benjamin K. Potter MD, Thomas A. Davis PhD,
Matthew B. Wagner PhD, Eric A. Elster MD, Jonathan A. Forsberg MD

Abstract

Background   To prevent symptomatic heterotopic ossi-fication (HO) and guide primary prophylaxis in patients with combat wounds, physicians require risk stratification meth-ods that can be used early in the postinjury period. There are no validated models to help guide clinicians in the treatment for this common and potentially disabling condition.

Questions/purposes   We developed three prognostic models designed to estimate the likelihood of wound-specific HO formation and compared them using receiver operating characteristic (ROC) curve analysis and decision curve analysis (DCA) to determine (1) which model is most accurate; and (2) which technique is best suited for clinical use.

Methods   We obtained muscle biopsies from 87 combat wounds during the first débridement in the United States, all of which were evaluated radiographically for development of HO at a minimum of 2 months postinjury. The criterion for determining the presence of HO was the ability to see radiographic evidence of ectopic bone formation within the zone of injury. We then quantified relative gene expression from 190 wound healing, osteogenic, and vascular genes. Using these data, we developed an Artificial Neural Net-work, Random Forest, and a Least Absolute Shrinkage and

K. A. Alfieri, B. K. Potter, J. A. Forsberg
Department of Orthopaedics, Walter Reed National Military Medical Center, Bethesda, MD, USA

K. A. Alfieri, B. K. Potter, T. A. Davis, M. B. Wagner, J. A. Forsberg (✉)
Regenerative Medicine Department, Naval Medical Research Center, 503 Robert Grant Avenue, Silver Spring, MD 20910, USA
e-mail: jaforsberg@me.com

K. A. Alfieri, B. K. Potter, T. A. Davis, M. B. Wagner, E. A. Elster, J. A. Forsberg
Department of Surgery, Uniformed Services University of the Health Sciences, Bethesda, MD, USA

Selection Operator (LASSO) Logistic Regression model designed to estimate the likelihood of eventual wound-specific HO formation. HO was defined as any HO visible on the plain film within the zone of injury. We compared the models accuracy using area under the ROC curve (area under the curve [AUC]) as well as DCA to determine which model, if any, was better suited for clinical use. In general, the AUC compares models based solely on accuracy, whereas DCA compares their clinical utility after weighing the consequences of under- or overtreatment of a particular disorder.
Results Both the Artificial Neural Network and the LASSO logistic regression models were relatively accurate with AUCs of 0.78 (95% confidence interval [CI], 0.72–0.83) and 0.75 (95% CI, 0.71–0.78), respectively. The Random Forest model returned an AUC of only 0.53 (95% CI, 0.48–0.59), marginally better than chance alone. Using DCA, the Artificial Neural Network model demonstrated the highest net benefit over the broadest range of threshold probabilities, indicating that it is perhaps better suited for clinical use than the LASSO logistic regression model. Specifically, if only patients with greater than 25% risk of developing HO received prophylaxis, for every 100 patients, use of the Artificial Network Model would result in six fewer patients who unnecessarily receive prophylaxis compared with using the LASSO regression model while not missing any patients who might benefit from it.
Conclusions Our findings suggest that it is possible to risk-stratify combat wounds with regard to eventual HO formation early in the débridement process. Using these data, the Artificial Neural Network model may lead to better patient selection when compared with the LASSO logistic regression approach. Future prospective studies are necessary to validate these findings while focusing on symptomatic HO as the endpoint of interest.
Level of Evidence Level III, prognostic study.

Introduction

Heterotopic ossification (HO) is the formation of mature, lamellar bone in nonosseous tissue. Up to 64% [3, 15] of combat casualties develop radiographically evident HO, which is far greater than that reported in civilian trauma

E. A. Elster
Department of Surgery, Walter Reed National Military Medical Center, Bethesda, MD, USA

J. A. Forsberg
Section of Orthopaedics and Sports Medicine, Department of Molecular Medicine and Surgery, Karolinska Institute, Karolinska University Hospital, Stockholm, Sweden

B. K. Potter, M. B. Wagner, E. A. Elster, J. A. Forsberg
Surgical Critical Care Initiative, Bethesda, MD, USA

literature [6–9, 17]. Approximately one-third of these will eventually undergo surgical excision for symptomatic lesions.

The causes of combat-related HO formation are just beginning to be elucidated. Previous studies identified variables associated with the eventual formation of HO. Recently published data demonstrate that it is most likely a result of blast mechanism of injury, impaired wound healing, bioburden, and local and systemic inflammatory dysregulation [2, 3]. Further work by Jackson et al [11] demonstrated that muscle-derived progenitor cells in blast-injured tissues are multipotent and may differentiate into adipocytes, chondrocytes, and osteoblasts. Likewise, Davis et al [2] found that high-energy combat wounds that eventually form HO have more connective tissue progenitor cells committed to osteogenic differentiation than wounds that do not form HO. Most importantly, the authors also suggested that osteogenic gene signatures may be detectable very early in the wound healing process, which forms the basis for the present study.

When symptomatic, HO affects many important aspects of recovering patients' convalescence. Pain, neurovascular compromise, primary ankylosis or secondary arthrofibrosis of joints, and skin ulceration are common. In amputees, HO may delay or complicate prosthetic fitting, which can ultimately degrade patient mobility and independence. Socket modification, rest, injection of neuromata, and medication adjustments can successfully treat the majority of symptomatic lesions [15]. However, for the patients who fail these conservative measures, operative excision, which is potentially debilitating and fraught with complications, remains the only treatment option [15].

Because the problem is so severe, identifying those patients at greatest risk to develop HO can help treating physicians target those individuals for prophylaxis such as nonsteroidal antiinflammatory drugs and radiotherapy. Unfortunately, other than vague demographic parameters such as age, sex, and presence or absence of head injury [6–9], which have little ability to discriminate among the mostly male, mostly young population, often seriously injured patients who present with these injuries, no reliable tools are available. We therefore were interested to see whether wound-specific gene expression analysis, which can be easily performed at the time of initial surgical débridement, might be useful in this regard. However, the large number of potential gene transcripts calls for a sophisticated approach. In selecting our models for analysis, we first looked at our data, which comprised almost 200 gene transcripts, and realized that we would need very sophisticated tools that would be able to discern relationships between genes with seemingly unrelated function. The Artificial Neural Network, Least Absolute Shrinkage and Selection Operator (LASSO) logistic regression, and Decision Tree are all excellent candidates because they have all of the qualities required for our data

set: ultimately, to generate an easily reproducible test that will produce a result that could guide HO prophylaxis.

Better means of risk stratification are needed to guide therapy as well as to support clinical trials evaluating novel means of primary prophylaxis currently in development. We therefore sought to determine the feasibility of risk-stratifying combat wounds for HO formation early in the postinjury period using mRNA transcripts isolated from wound muscle tissue. Toward this end, we developed Artificial Neural Network (ANN), Random Forest (RF), and LASSO logistic regression models for HO formation based on expression of inflammatory, angiogenic, osteogenic, and wound healing genes. We chose these models for their unique discriminatory ability as well as their ability to evaluate large quantities of analytic data.

We then asked (1) which model was most accurate using receiver operating characteristic (ROC) analysis; and (2) which model performed best on decision curve analysis (DCA) [19] and is therefore best suited for clinical use.

Patients and Methods

After institutional review board approval, we screened 670 patients for enrollment at our institution between January 2007 and December 2011. Of these, 72 patients with 87 wounds gave informed consent after first meeting the inclusion criteria—consisting of the presence of one or more high-energy, combat-related extremity wound > 75 cm^2. Twelve patients who otherwise met inclusion criteria or their legally authorized representative declined to participate. All 72 patients had at least 2 months of radiographic followup. All patients were male with a median age of 22 (interquartile range [IQR] 21, 26) and the majority (86%) sustained a blast injury. Forty-three patients (60%) presented with soft tissue injuries and 15 patients (21%) with major limb amputations. Twenty-nine wounds (33%) developed radiographic evidence of HO at a minimum of 2 months postinjury.

Muscle biopsies were obtained from healthy-appearing tissue after the initial irrigation and débridement procedure. This was performed after patients arrived at our facility from overseas at a median of 5 days (IQR 4, 7 days) after the initial injury. From these samples, we assessed the gene transcript expression of 190 wound healing, inflammatory, osteogenic, and vascular genes using a custom-designed TaqMan® Low Density Array (Applied Biosystems, Rockville, MD, USA). Each gene was quantified by normalizing to the 18S expression, and mRNA transcript levels were assessed in duplicate.

The presence or absence of HO for a given wound was confirmed by a two-author review (JAF, BKP) of good-quality orthogonal radiographs at a minimum of 2 months

postinjury [3, 15]. There was no disagreement regarding the presence or absence of HO between reviewers. Using these data, we developed three models, a modified (LASSO) logistic regression, a RF model, and an ANN, to estimate the likelihood of eventual wound-specific HO formation. The first uses a traditional regression approach. The latter two methods are computer-intensive and use machine learning to identify patterns within the data as well as their association(s) with the outcome of interest.

Each model was created using the same data and trained to estimate the likelihood of HO formation based on the gene transcript products. We developed the ANN model using the Oncogenomics Online Artificial Neural Network Analysis system [14], which was designed to be used with small sample sizes and a relatively large number of candidate features. First, data were transformed with a log transformation to normalize distributions. Next, principal component analysis was performed on all 190 candidate features to identify the top 10 linearly uncorrelated variables with the largest variance. This was done in an effort to simplify the model as well as mitigate overfitting to the training data and potentially maximize applicability to other populations. The network was composed of three layers: an input layer consisting of the 10 principal components identified, a hidden layer (which may change the relative importance placed on data from each of the inputs) with five nodes, and an output layer producing a committee vote discriminating two possible outcomes for a given wound (development of HO: yes or no). We then performed internal validation using the 10-fold crossvalidation method. Briefly, we first randomized the data into 10 matching train-and-test sets. Each set consisted of a training set composed of 90% of patient records and a test set composed of the remaining 10% of records. Stratification of the data by patients with multiple wounds was not considered necessary, because the outcome measures were assessed on a wound-specific basis. For instance, wound-specific gene expression is not likely to yield prognostic information for a remote wound. These 10 iterations of crossvalidation yield 10 models with different parameter weightings that are then evaluated using the ROC and area under the curve (AUC) characteristics.

The RF model was developed using R® Version 3.1.1 statistical software [16]. The RF is composed of multiple decision trees using classification and regression tree methodologies. As stated previously, all data were transformed with a log transformation to normalize distributions. We accommodated the small sample size and more numerous candidate features to reduce overfitting using the random subspaces method [10]. The RF generates multiple models using the training data that are aggregated into the final prediction (development of HO: yes or no) while controlling for number of trees, complexity, and

resampling. Tenfold crossvalidation was performed as described previously and evaluated with the other models.

For comparison purposes, we also developed a modified logistic regression model using the LASSO method in R® Version 3.1.1 statistical software [16]. Only potentially significant variables identified on univariate analysis, p < 0.3, were entered into the multivariate model to reduce overfitting the patient population. All three modeling methods produced data appropriate for ROC analysis and DCA. Tenfold crossvalidation was performed. The LASSO model parameters were determined through the 10-fold crossvalidation and selected to yield the minimum mean crossvalidated error.

We then directly compared the models using two methods. First, we assessed the accuracy (AUC) by evaluating the ROC curves. ROC curves plot the true-positives of a diagnostic test or model versus the false-positives. A more accurate model is represented by a curve that falls above a diagonal line with the slope of 1, which represents a "flip of a coin" or 50% accuracy. The area under the ROC curve is used to quantify accuracy and can be used to compare different models. Finally, we compared each model using DCA, a technique that weighs the clinical consequence of "wrong answers" (false-positives and false-negatives) generated by the models. Net benefit, defined as patients who duly receive primary prophylaxis after appropriate risk stratification, was calculated and plotted versus the threshold probability (p_t) of HO formation. The p_t is the probability in which a surgeon is indecisive about whether to give prophylaxis for a particular wound. Each p_t is related to how surgeons weigh the relative consequences of over- or undertreating the patient and is dependent on the safety profile of the method of primary prophylaxis being considered as well as patient factors including associated injuries, concomitant fractures, etc. By plotting p_t along a continuum, we are able to evaluate each model over all possible thresholds (0–1), thus making the DCA of these particular models applicable to a variety of settings and important in considering the diversity of the combat-wounded patient population as well as the safety profiles of all current (and future) means of primary prophylaxis. The code for all analyses except for the ANN development is included as supplementary material (Appendices 1–5; the supplementary material can be opened with R, which is free; you can get R at: www.r-project.org [Supplemental materials are available with the online version of CORR®.]).

Results

The most reliable models based on ROC analysis were the ANN and the LASSO logistic regression, both of which were superior to the RF model. On internal validation, the AUC for the ANN was 0.78 (95% confidence interval [CI], 0.72–0.83) compared with 0.75 (95% CI, 0.71–0.78) for the LASSO model (p = 0.19) and 0.53 (95% CI, 0.48–0.59) for the RF model (p < 0.0001) (Fig. 1). The ANN model identified an eight-gene signature including EGR1, CX3CL1, SMAD6, FADD, TGFB2, CCL11, CXCL11, and HMGB1 that successfully estimated the likelihood of eventual wound-specific HO formation. The RF model identified 15 genes including MMP1, MPO, BMP5, IGFBP6, SMAD6, TIMP2, BMP4, CCL28, CX3CL1, NCAM2, BMP1, CCL19, ECGF1, GDF5, and MMP11. The LASSO modeling method produced a 19-gene
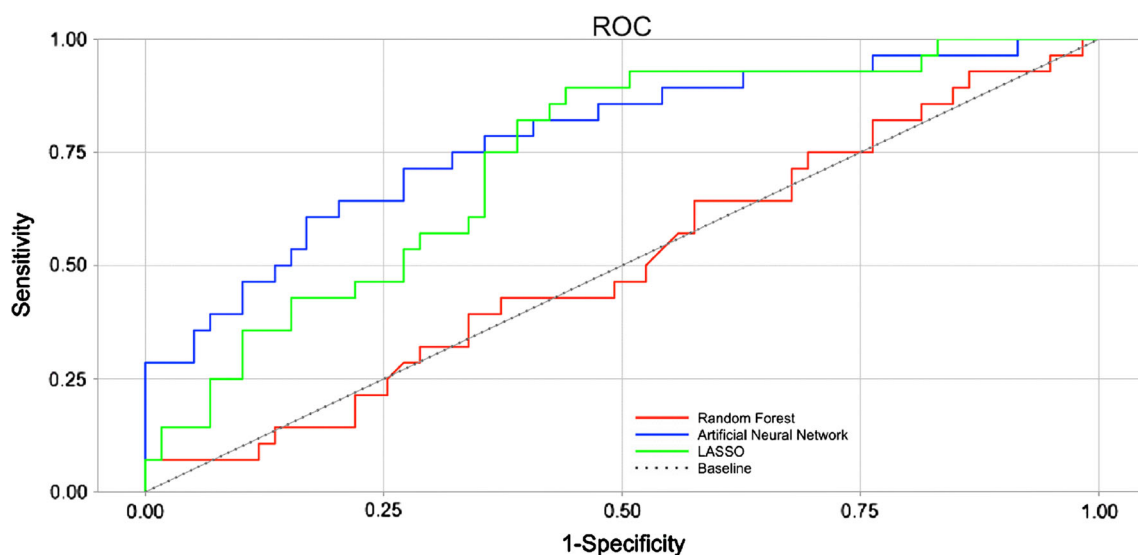

Fig. 1 ROC curve analysis demonstrates an AUC for the ANN of 0.78 (95% CI, 0.72–0.83) compared with 0.75 (95% CI, 0.71–0.78) for the LASSO model and 0.53 (95% CI, 0.48–0.59) for the RF model.
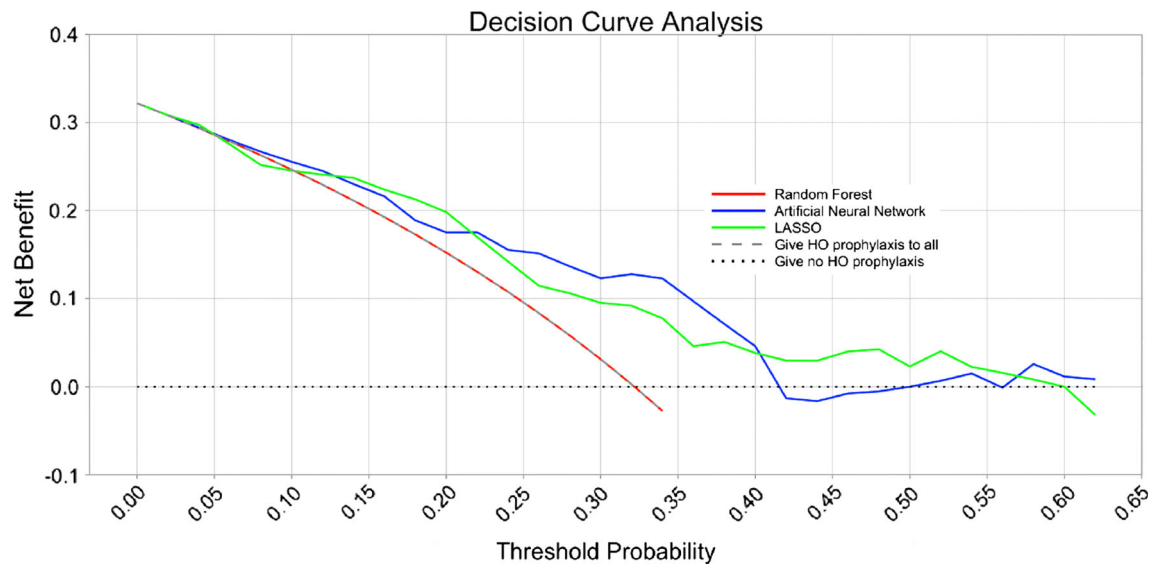
Fig. 2 DCA demonstrates use of the ANN and the LASSO models result in positive net benefit, indicating either could be used rather than assume all patients or none receive primary prophylaxis.

signature, ACTA2, ANGPT1, BMP3, BMP5, CCL28, CXCL1, ECGF1, FGF5, GAPDH, GDF3, GDF5, IGFBP6, IL12A, IL17A, MMP3, PF4, SERPINE1, SLPI, and TGFB.

Although the DCA revealed the ANN and LASSO models had a positive net benefit (Fig. 2), indicating that each could potentially be used clinically, the ANN model resulted in a higher net benefit (y-axis) when compared with the LASSO model across a broader range of threshold probabilities (x-axis). These results suggest if only patients with greater than 25% risk of developing HO received prophylaxis, for every 100 patients, use of the ANN model would reduce the number who unnecessarily receive prophylaxis by 18 (six more than the LASSO regression model) while not missing any patients who duly require it. The RF model was only marginally more accurate than chance alone and provided no better net benefit than assuming all patients should receive prophylaxis.

Discussion

For the combat-injured patient, HO formation can be an important barrier to functional mobility, independence, and return to active duty. Similarly, HO formation as a result of high-energy civilian trauma, especially in the acetabulum and elbow, can cause significant disability and may also benefit from this research. Unfortunately, there are currently no methods to risk-stratify individual patients or wounds to guide the use of local and/or systemic means of primary prophylaxis. We therefore used results from mRNA assays of tissue samples taken from the first débridement performed at Walter Reed National Military Medical Center to develop

three models, an ANN, RF, and a LASSO model, capable of risk-stratifying combat-related wound-specific HO formation early in the débridement process. We found that two models, ANN and LASSO logistic regression, demonstrated superior and near equivalent accuracy and that the ANN provided the best clinical utility.

This study has limitations. First, this study focuses on estimating the likelihood of any HO formation, not necessarily the lesions that would go on to be symptomatic. Although we believe risk-stratifying wounds early in the débridement process is an important finding of this study, further research geared toward estimating the likelihood of symptomatic HO is arguably as important and is the logical next step in future analyses. This retrospective analysis included only combat-related patients enrolled in a clinical study, who had at least 2 months of radiographic followup. The results may not be applicable to other patients with less severe extremity wounds or even those who sustain civilian trauma. By the same token, we report internal validation statistics, which are known to overestimate model accuracy. However, these yield upper-bound confidence limits of how the models may perform when confronted with external validation data while reducing the likelihood of overfitting (the process of modeling "noise" within the data). As such, until external validation is complete, the ANN or LASSO models reported are not ready for widespread clinical use. Second, we used only wound-specific gene expression data for these analyses. This was done in an effort to derive a wound-specific risk stratification tool; however, incorporation of systemic measures of inflammation such as procalcitonin or interleukin-6 [4] may improve accuracy, but this is unproven and deserving of

further study. In addition, the ANN and LASSO models require complete input data, which, in contrast to other techniques, may limit their use when information is missing. The RF model, however, can accommodate missing data while maintaining accuracy; however, this technique did not result in a useful model. Still, any future external validation study would simply obtain the necessary transcription data for the requisite genes.

Whenever new tests are developed, cost is a factor to consider. The costs for conducting gene expression studies using custom or commercially available low-density microarrays are relatively inexpensive. For example, in this study we assessed mRNA gene transcript for 192 different target genes (including control housekeeping) in duplicate using a 384-array platform. The total cost for the arrays and reagents for mRNA isolation and QC validation is approximately USD 225 or USD 0.58 per reaction for the materials alone (when factoring in equipment and staff time, although an estimate of USD 1 to 2 per reaction may be more realistic). However, finally, these models may only be applicable to the patient population in which they were developed—that is, for use in blast- and otherwise combat-related extremity wounds. A final limitation that applies here is that this study examined only 87 wounds from 72 patients, which is a relatively small sample size considering the large number of candidate features (190 genes). We acknowledge this and attempted to mitigate it by using an ANN specifically designed for this setting and by incorporating the Random Subspace Method [10] into the RF model. Still, it is possible that overfitting occurred, further emphasizing the importance of external validation.

Our findings suggest that ANN and LASSO models, but not the RF model, are capable of estimating the likelihood of wound-specific HO early in the débridement process. This is evidenced by AUCs of 0.78, 0.75, and 0.53 for the ANN, LASSO, and RF models, respectively. We were surprised to find the RF model was least accurate despite incorporation of a relatively large amount of transcriptomic information. Although accuracy is important in medical decision-making, it must be appropriately tempered with a measure of clinical use.

When developing Clinical Decision Support tools, approaches that focus predominantly on accuracy should be avoided [1]. DCA has been used previously in orthopaedic surgery to weigh the relative consequences of a falsely positive or negative prediction by the model [5]. Performing DCA enables one to evaluate the risk of over- or undertreatment and assess which model, if any, is best suited for clinical use. Depending on the threshold probability (the probability of HO formation at which the surgeon becomes indecisive about offering prophylaxis), the ANN and LASSO models appeared best at some point along the continuum. However, the ANN model resulted in the highest net

benefit over the broadest range of threshold probabilities, which translates to better patient selection when compared with the LASSO model. Therefore, the ANN may be consistently the most useful model when applied in a clinical setting. At the extreme ($p_t > 0.6$), more patients would be appropriately treated (duly offered prophylaxis or not) if surgeons assumed no patient would develop HO rather than use the model output. This is important when one considers p_t is patient-, surgeon-, treatment-, and scenario-dependent. It is incumbent on the treating surgeon to determine his or her p_t based on a variety of factors. For instance, his or her p_t is likely to be higher when treating patients with many contraindications to prophylaxis (eg, recent spine fusion, multiple long bone fractures, or a history of gastric ulcerations if considering nonsteroidal antiinflammatory drugs as a means of primary prophylaxis) compared with patients with few, if any, overt contraindications. Importantly, DCA evaluates and compares models over a range of threshold probabilities so an exact p_t need not be specified a priori. In fact, the DCA curves produced by these data may be applicable not only to current therapies, but also future means of primary prophylaxis currently in development.

Of the eight genes identified by the ANN, two gene transcripts, EGR1 and CX3CL1, were upregulated; one transcript, SMAD6, was downregulated; whereas FADD, TGFβ2, CCL11, CXCL11, and HMGB1 were found to be unchanged when compared with wounds that did not form HO. EGR1, TGFβ2, and SMAD6 are all involved with regulation of bone formation; CX3CL1, CCL11, CXCL11, and HMGB1 are all involved with regulation of inflammatory response; FADD regulates apoptosis [12, 18]. The RF model identified 15 genes; however, only two of them (CX3CL1 and SMAD6) were also identified by the ANN model. This illustrates differences in the feature selection process between the two techniques and may help explain the relative inaccuracy achieved by the RF model. In contrast, the LASSO model, using a frequentist approach, identified a 19-gene signature, ACTA2, ANGPT1, BMP3, BMP5, CCL28, CXCL1, ECGF1, FGF5, GAPDH, GDF3, GDF5, IGFBP6, IL12A, IL17A, MMP3, PF4, SERPINE1, SLPI, and TGFβ. BMP3 and 5, GDF-5 and GDF3 and 5, and TGFβ, members of the TGFβ superfamily, play an important role in mesenchymal stem cell differentiation and endochondral bone formation.

Although a mechanistic discussion is beyond the scope of this study, it is important to note that the genes identified by each of the modeling processes may not have biological significance from a mechanistic standpoint. For instance, the ANN model achieved the best discriminatory ability (accuracy) by including only eight transcripts regardless of whether they were upregulated, downregulated, or remained unchanged when stratified by wound outcome. The less accurate RF model identified 15 transcripts, likely

because it introduces randomness into the model by including potentially less important features. In theory, this serves to decrease the propensity to overfit to the training data and also serves to accommodate smaller sample sizes such as ours. Taken independently, the individual genetic transcripts identified by either the ANN or the RF models do not indicate potential HO development. However, they do so only in the context of the complete, coherent model. In contrast, the LASSO model describes direct associations between individual transcripts and wound outcome. As such, the association between BMP4 and GDF3 with eventual HO formation, reported previously [13] in animal models, is deserving of further study in humans.

In conclusion, we successfully estimated the likelihood of HO using wound-specific gene expression data available early in the débridement process. The ANN and LASSO logistic regression models were both found to be accurate; however, the ANN may be better suited for clinical use because it results in better patient selection than both the LASSO logistic regression model and the current standard of care, which is heavily driven by clinical judgment and seldom guided by formal diagnostic testing. Although these results are encouraging, external validation, currently underway, is absolutely required before recommending that this model be used clinically. Importantly, these results suggest that wounds are committed to form HO very soon after injury. Additional studies are necessary to characterize the mechanisms behind this phenomenon as well as to evaluate means of primary prophylaxis that can be used immediately after a blast or other combat-related injury.

References

1. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*. 2007;115:928–935.
2. Davis TA, O'Brien FP, Anam K, Grijalva S, Potter BK, Elster EA. Heterotopic ossification in complex orthopaedic combat wounds: quantification and characterization of osteogenic precursor cell activity in traumatized muscle. *J Bone Joint Surg Am*. 2011;93:1122–1131.
3. Forsberg JA, Pepek JM, Wagner S, Wilson K, Flint J, Andersen RC, Tadaki D, Gage FA, Stojadinovic A, Elster EA. Heterotopic ossification in high-energy wartime extremity injuries: prevalence and risk factors. *J Bone Joint Surg Am*. 2009;91:1084–1091.
4. Forsberg JA, Potter BK, Polfer EM, Safford SD, Elster EA. Do inflammatory markers portend heterotopic ossification and wound failure in combat wounds? *Clin Orthop Relat Res*. 2014;472:2845–2854.
5. Forsberg JA, Sjoberg D, Chen Q-R, Vickers A, Healey JH. Treating metastatic disease: Which survival model is best suited for the clinic? *Clin Orthop Relat Res*. 2013;471:843–850.
6. Garland DE. Clinical observations on fractures and heterotopic ossification in the spinal cord and traumatic brain injured populations. *Clin Orthop Relat Res*. 1988;233:86–101.
7. Garland DE, Dowling V. Forearm fractures in the head-injured adult. *Clin Orthop Relat Res*. 1983;176:190–196.
8. Garland DE, Rothi B, Waters RL. Femoral fractures in head-injured adults. *Clin Orthop Relat Res*. 1982;166:219–225.
9. Garland DE, Toder L. Fractures of the tibial diaphysis in adults with head injuries. *Clin Orthop Relat Res*. 1980;150:198–202.
10. Ho TK. The random subspace method for constructing decision forests. *IEEE Trans Pattern Anal Mach Intell*. 1998;20:832–844.
11. Jackson WM, Aragon AB, Bulken-Hoover JD, Nesti LJ, Tuan RS. Putative heterotopic ossification progenitor cells derived from traumatized muscle. *J Orthop Res*. 2009;27:1645–1651.
12. National Center for Biotechnology Information. National Institutes of Health. Available at: www.ncbi.nlm.nih.gov/gene. Accessed February 4, 2013.
13. Olmsted-Davis E, Gannon FH, Ozen M, Ittmann MM, Gugala Z, Hipp JA, Moran KM, Fouletier-Dilling CM, Schumara-Martin S, Lindsey RW, Heggeness MH, Brenner MK, Davis AR. Hypoxic adipocytes pattern early heterotopic bone formation. *Am J Pathol*. 2007;170:620–632.
14. Online artificial neural network analysis tool. Available at: http://home.ccr.cancer.gov/oncology/oncogenomics/tool.htm. Accessed April 13, 2012.
15. Potter BK, Burns TC, Lacap AP, Granville RR, Gajewski DA. Heterotopic ossification following traumatic and combat-related amputations. Prevalence, risk factors, and preliminary results of excision. *J Bone Joint Surg Am*. 2007;89:476–486.
16. R Core Team. R: A language and environment for statistical computing. Available at: http://www.R-project.org/. Accessed April 2, 2014.
17. Steinberg GG, Hubbard C. Heterotopic ossification after femoral intramedullary rodding. *J Orthop Trauma*. 1993;7:536–542.
18. *The Gene Ontology*. Available at: http://www.geneontology.org. Accessed February 4, 2013.
19. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006;26:565–574.