



Published in final edited form as:

*Wound Repair Regen.* 2016 January ; 24(1): 181–188. doi:10.1111/wrr.12384.

## Rapid identification of slow healing wounds

**Kenneth Jung, MS<sup>1</sup>, Scott Covington, MD<sup>2</sup>, Chandan K. Sen, PhD<sup>3</sup>, Michael Januszyk, MD<sup>4</sup>, Robert S. Kirsner, MD, PhD<sup>5</sup>, Geoffrey C. Gurtner, MD<sup>4</sup>, and Nigam H. Shah, MBBS, PhD<sup>6</sup>**

<sup>1</sup>Program in Biomedical Informatics, Stanford University, Stanford, California

<sup>2</sup>Healogics Inc., Jacksonville, Florida

<sup>3</sup>Departments of Surgery, Comprehensive Wound Center, Davis Heart and Lung Research Institute, Centers for Regenerative Medicine and Cell-based Therapies, The Ohio State University Medical Center, Columbus, Ohio

<sup>4</sup>Department of Surgery, Division of Plastic and Reconstructive Surgery, Stanford University School of Medicine, Stanford, California

<sup>5</sup>Department of Dermatology and Cutaneous Surgery, University of Miami Miller School of Medicine, Miami, Florida

<sup>6</sup>Center for Biomedical Informatics Research, Stanford University, Stanford, California

### Abstract

Chronic nonhealing wounds have a prevalence of 2% in the United States, and cost an estimated \$50 billion annually. Accurate stratification of wounds for risk of slow healing may help guide treatment and referral decisions. We have applied modern machine learning methods and feature engineering to develop a predictive model for delayed wound healing that uses information collected during routine care in outpatient wound care centers. Patient and wound data was collected at 68 outpatient wound care centers operated by Healogics Inc. in 26 states between 2009 and 2013. The dataset included basic demographic information on 59,953 patients, as well as both quantitative and categorical information on 180,696 wounds. Wounds were split into training and test sets by randomly assigning patients to training and test sets. Wounds were considered delayed with respect to healing time if they took more than 15 weeks to heal after presentation at a wound care center. Eleven percent of wounds in this dataset met this criterion. Prognostic models were developed on training data available in the first week of care to predict delayed healing wounds. A held out subset of the training set was used for model selection, and the final model was evaluated on the test set to evaluate discriminative power and calibration. The model achieved an area under the curve of 0.842 (95% confidence interval 0.834–0.847) for the delayed healing outcome and a Brier reliability score of 0.00018. Early, accurate prediction of delayed healing wounds can improve patient care by allowing clinicians to increase the aggressiveness of intervention in patients most at risk.

---

Reprint requests: Kenneth Jung, Stanford Center for Biomedical Informatics Research, Medical School Office Building, 1265 Welch Rd, Room X-215, Stanford, CA 94305-7994, USA. Tel: (415) 640-5289; Fax: (650) 725-7944; kjung@stanford.edu.

Additional Supporting Information may be found in the online version of this article.

## INTRODUCTION

Chronic wounds are those that fail to heal in a timely manner,<sup>1</sup> and affect an estimated 6.5 million people in the United States (2% of the population).<sup>2</sup> Chronic wounds are at increased risk of complications, such as amputation and infection, which can have a severe negative impact on patient well-being. The cost of treating these wounds is high—up to \$50 billion annually<sup>3</sup>—and the incidence of chronic wounds is expected to increase due to an aging population and rising risk factors such as diabetes and obesity.<sup>4</sup> Knowing in advance that a given wound is likely to be problematic despite standard care could enable care providers to make better decisions about treatment options such as early intervention with hyperbaric oxygen therapy (HBOT) or negative pressure wound therapy that may improve outcomes.<sup>5–10</sup>

Previous work has attempted to identify factors of prognostic value in predicting delayed wound healing, but most of these studies were restricted to specific wound types such as venous leg ulcers.<sup>11–16</sup> They used patient cohorts of modest size and drawn from single sites or enrolled in clinical trials, limiting the generalizability of these results to the diversity of patients and wound types seen in clinical practice.<sup>17</sup> Furthermore, these models have not yet been validated in independent datasets with respect to either discriminatory power or calibration.

Among the best work to date is that of Margolis et al.<sup>11,12</sup> who developed prognostic models for venous leg ulcers and diabetic neuropathic foot ulcers using data from tens of thousands of patients across geographically diverse outpatient wound care centers, and carefully validated the models, achieving excellent calibration but only modest discriminative power (area under the curves [AUCs] ranging from 0.63 to 0.71). These results led to the study by Kurd et al.,<sup>18</sup> a clustered multicenter trial demonstrating that providing prognostic information from these models to clinicians improved healing rates even without specific guidance about treatment options.

In this work, we report the development and validation of a novel prognostic model that uses data from an Electronic Health Record (EHR) collected at the onset of care in outpatient wound care centers. De-identified data was obtained under an institutionally reviewed agreement between Stanford University and Healogics Inc. for the purpose of building a wound healing prediction model. These data comprise basic patient demographic information such as age and gender, along with clinically observable wound characteristics such as wound type, anatomic location, dimensions, and qualities such as rubor and erythema. Our model achieves an AUC of 0.842 (95% confidence interval 0.836–0.849) on held out data and provides well-calibrated probabilities for delayed wound healing. Our model is constructed and validated on a dataset comprising tens of thousands of patients with over a 100,000 wounds from geographically diverse wound care centers. All wound types are considered in this work, and thus the model is applicable to the full diversity of wounds seen in clinical practice. We note that similar to risk models such as APACHE III,<sup>19</sup> our model would need to be prospectively evaluated for settings other than Healogics wound care centers. Our contribution is to demonstrate that modern statistical learning techniques,

applied to a large clinical dataset, can provide accurate prognostic information about delayed wound healing.

## MATERIALS AND METHODS

### Dataset

Figure 1 provides an overview of the dataset, and Supporting Information S1 contains a full list of the variables collected, along with basic patient and wound statistics. In brief, the dataset was compiled from data collected at 68 Healogics wound care centers in 26 states from 2009 to 2013. Patients underwent serial assessment of their wounds, with appropriate interventions as recommended in evidence based clinical guidelines based on factors such as vascular status, debridement requirements, and presence of infection or inflammation. Treatment decisions include dressing selection, pain management, pressure relief, systemic disease management, and the use of advanced modalities such as HBOT. This includes data on 180,696 wounds from 59,953 patients, with wounds receiving standard care, and assessed weekly until healed or otherwise resolved (e.g., because treatment was discontinued due to noncompliance, admission to a hospital, etc.). Patient information includes demographic information along with ICD9 codes at each encounter. Wound information included quantitative wound characteristics such as length, width, depth (measured during each wound assessment with a handheld ruler), and area, along with categorical wound characteristics such as “rubor,” “erythema,” “undermining,” and “tunneling.” There were 40 wound types (e.g., pressure ulcer, venous ulcer) and 37 anatomic locations. Venous leg ulcers, pressure ulcers, and diabetic wounds on the lower extremities (not otherwise specified) accounted for 48.8% of wounds, with trauma and surgery contributing an additional 13.4 and 9.5%, respectively. Consistent with the preponderance of certain wound types, wounds on the lower leg, foot, toes, or ankles accounted for 68.5% of all wounds. After removal of wounds whose final disposition was unknown at the end of the study period or containing outlier values for quantitative features as detailed below, the dataset consisted of 150,277 wounds from 53,354 patients. Thus, each patient had on average 2.82 wounds over the study period. Patients ranged from 1 to over 100 years old (median of 67 years) and were 52.7% male. The mean duration of care was 52.2 days. Overall, 11.6% of wounds exhibited delayed wound healing, defined as 15 or more weeks to closure. Figure 2 shows the distribution of the number of wounds per patient, patient age, wound types, and wound locations, along with the duration of care.

### Data preprocessing

In order to guard against spurious values in the data, we removed all records for patients who had implausible outlier values in any of the primary quantitative wound characteristics—i.e., wound area, length, width, and depth. Outliers were determined by manual inspection of the distribution of values for each feature. This process identified 658, 137, 227, and 58 wounds, respectively, for each wound measurement. Additionally, we removed all wound records for patients with negative or missing values for these fields or age in any of their wound records. This process yielded a final set of 53,354 patients and 150,277 wounds for model development and validation.

## Feature construction

Previous work has shown that variables that capture the rate of change of wound dimensions have prognostic value. Therefore, we calculated features that encode both absolute and relative changes in wound dimensions (area, length, width, and depth) between the first wound assessment and the second wound assessment occurring 1 week later. We also note that patients on average have 2.8 wounds over the time period of the study. It is possible that there are patient specific factors that may impact the risk of delayed wound healing. These factors may be modeled as latent variables, but estimation of such models is generally problematic, especially at large scales.<sup>20</sup> Thus, in this study, we have calculated features that reflect patient specific factors that may have an impact on delayed wound healing based directly on observable characteristics, such as total number of wounds, history of delayed wound healing, and total wound surface area. Note that these characteristics are calculated at the time of first assessment for each wound, and thus do not peek ahead into the future by using patient and wound information that is not available at prediction time.

## Model development

Model development is summarized in Figure 3. We used data from the first and second wound assessments (weeks one and two) to construct predictors of delayed wound healing. We chose delayed wound healing as the outcome because long term, nonhealed wounds are a significant source of morbidity for patients, increase cost to the healthcare system and, in a capitated payment model, a significant risk for the care provider.

Patients were randomly assigned into training (42,683 patients) or test (10,671 patients) sets; wounds were then assigned to training or test sets according to the assignment of the corresponding patients. Doing so prevents the test and training data from containing wounds from the same individual. This resulted in 120,473 training and 29,804 test wounds. Quantitative predictors were standardized using mean and variance estimates from the training set; these parameters were also used to standardize the test data. Categorical variables were encoded as binary values. In all, 865 predictors were used, 834 of which were binary (0 or 1) encodings of categorical variables.

We fit L1 regularized logistic regression (lasso), random forest, and gradient boosted tree models on the training data using the `glmnet`,<sup>21</sup> `randomForest`,<sup>22</sup> and `gbm`<sup>23</sup> packages, respectively. Model hyperparameters were set by evaluating model performance in held out subsets of the training data. In particular, GBMs have hyperparameters that are tuned to prevent overfitting to training data—the bag fraction, the learning rate, and the number of iterations. The bag fraction was left at the default value of 0.5. We found optimal values for the other hyperparameters by fitting the model to 90% of the training data and monitoring the error on the remaining 10% of the training data. The models were then trained on the full training set using these hyperparameter settings. We found that the best single model was a gradient boosted tree model. The remainder of this study thus focused solely on that model. The ability of the model to predict delayed wound healing was measured by area under the receiver-operator curve (AUC) calculated with the `pROC` R package, with 95% confidence intervals estimated by bootstrapping with  $B=10,000$ .<sup>24</sup> Model calibration (i.e., the

agreement of the predicted probabilities of delayed wound healing to observed frequencies) was measured using Brier reliability.<sup>25</sup>

## RESULTS

A prognostic model for delayed wound healing should discriminate between normal and delayed healing wounds, and provide accurate probabilities for those outcomes.<sup>26,27</sup> We evaluated discrimination by AUC, and calibration by Brier reliability (Figure 4). The AUC summarizes the ability of the model to discriminate between normal and nonhealing wounds over the range of tradeoffs between sensitivity (recall) and specificity. An AUC of 0.5 means the model performs no better than random guessing, while an AUC of 1 means the model discriminates perfectly.

Our model achieved an AUC of 0.842 (95% confidence interval 0.836–0.849) over all wound types. For the two wound types previously examined by Margolis et al.<sup>11,12</sup> (venous leg ulcers and diabetic neuropathic ulcers), we achieve AUCs of 0.827 and 0.823, respectively, a significant improvement over the previously reported best results of 0.71 and 0.70. Performance in other wound types well represented in the test set ( $N > 500$ ) ranged from 0.799 for lower extremity diabetic wounds to 0.838 for pressure ulcers.

We also examined the importance of different variables to the model using relative influence, a measure of the contribution to the discriminative performance of the model.<sup>28</sup> Relative influence is normalized such that the relative influences of all variables sum to 1; higher numbers indicate greater importance. Interestingly, whether or not the patient is undergoing palliative care is the single most important predictor. This is because patients falling into this category are disproportionately likely to exhibit delayed wound healing. Other important variables are patient age and features for wound dimensions and their changes between the first and second assessments. Somewhat surprisingly, wound type and location had relatively little importance, with each cumulatively accounting for only 4.4 and 2.8 of total relative influence, respectively. In all, 381 out of 865 variables had zero relative influence, indicating that they may be dropped from the model entirely. The bulk of the zero-influence variables consisted of ICD9 codes, whose 587 codes collectively accounted for only 3.1 of 100 relative influence points. The top 100 features account for over 95% of the total relative influence. It is not possible to directly translate cumulative relative influence to the performance using a subset of features, but these results suggest that it may be possible to fit a model with much of the predictive power of our model with a significantly smaller set of inputs—an experiment fitting a model with only the top 100 most important features achieved a test set AUC of 0.841. The top 10 most important variables are shown in Table 1; Supporting Information S2 lists all features and their relative influence in the final model.

Brier reliability measures agreement between predicted probabilities and observed frequencies for stratified samples on a scale between 0 and 1, with small values indicating good agreement and 0 indicating perfect agreement. Test cases were stratified based on their predicted probability of delayed wound healing. The model scored a Brier reliability of 0.00018 indicating that the predicted probabilities in each stratum closely matched observed frequencies of delayed wound healing (see Figure 4).

## DISCUSSION

From Ambrose Paré to antiquity, the care and treatment of wounds has been central to the identity and practice of surgery. General, vascular and plastic surgeons all regularly see chronic wounds as part of their practice. Having an accurate predictive computational algorithm would be transformational in the triage and care of these patients. Early knowledge of wound severity would alter the care of patients by allowing clinicians to increase the aggressiveness of intervention in patients most at risk, while diverting costly treatments from patients predicted to heal easily. Kurd et al.<sup>18</sup> has already demonstrated the value of even relatively inaccurate prognostic information in the wound care center setting in a cluster randomized multicenter trial. We have used modern machine learning techniques to develop a prognostic model for delayed wound healing that significantly advances the state of the art in accuracy and the types of wounds for which a prediction can be made. The model uses data routinely captured in EHRs during the first week of care (two wound assessments spaced 1 week apart) at wound centers and achieves best-to-date discriminative power and calibration on validation data. Unlike previous work, our model is also applicable to the full range of wound types. Furthermore, by changing the cutoff threshold, it is possible to make trade offs between positive predictive value and sensitivity (recall) for specific uses of the model (Figure 3B). For instance, it may be desirable to trade off low positive predictive value for a high sensitivity when making decisions regarding referral to specialized wound care centers. Once referred to a wound center, however, it may be desirable to have a high positive predictive value and sacrifice sensitivity when making decisions about high cost interventions that may improve outcomes in potentially problematic cases. Because the model outputs are well-calibrated probabilities, they can be used directly to help care providers and patients make informed referral or intervention decisions.

It is worth emphasizing how our methods differ from previous efforts.<sup>11</sup> Traditional prognostic models have used a small number of manually selected features and were developed using relatively small datasets. Often, these models took the form of checklists, where clinicians added up numbers corresponding to the presence of risk factors for the outcome of interest. Such models were common for two main reasons. First, more complex models with more features would have been a great risk of over fitting the small amount of training data (i.e., performing well on the training data but poorly in actual use, i.e., on new data). Second, such models can be put into practice without requiring electronic medical records systems. However, advances in statistical learning and the widespread adoption of EHRs allow us to move beyond such limitations. These developments have motivated the creation of increasingly sophisticated models that can leverage very large datasets without overfitting, while also automatically learning which features are important. Several scientific domains have benefited from such advanced methods.<sup>29</sup> In this work, we demonstrate the use of such methods to predict delayed wound healing. The resulting model significantly outperforms previous models.

Our approach does have limitations. Although our model was developed on data from 68 geographically distributed wound care centers in 26 states, it may not generalize to centers other than those that participated in this study. Indeed, we observe considerable inter-center



variability in practices and patient populations. Second, it is possible that the model may not generalize to cases outside wound care centers, limiting its utility for making referral decisions. More generally, we do not envision applying the model as-is in non-Healogics wound care centers without prospective validation. Our contribution is to demonstrate that modern statistical learning techniques applied to a large clinical dataset provide accurate prognostic information about the risk of delayed wound healing—therefore, given sufficient data, it is possible to train institution specific models as described in this study. Third, the model may not perform as well on new cases despite our evaluation on a hold out validation set; a prospective study is required to best estimate performance in practice. We also note that some aspects of our construction of the training and test sets, such as the removal of wounds with implausible quantitative wound measurements and without a definitive outcome (because patients were lost to follow-up before the 15 week threshold) may induce bias in our estimates of performance. Another potential source of bias is confounding by indication.<sup>30</sup>

Finally, as with all predictive models, it would be necessary to monitor performance on recent cases to guard against changes in practices, patient populations, and other factors degrading performance.

We note that simple models, such as those of Margolis et al., can be easily put into practice by clinicians as simple checklists. However, this ease of use comes at the cost of significantly lower accuracy. In contrast, our model is more complex and not intended to be applied directly by clinicians using pen and paper. Instead it would be deployed within the EHR system and clinicians would use its output to help guide their management of wounds. In exchange for embracing a modern, computational approach, we have achieved significantly higher accuracy than was possible with the simpler models. With the widespread and growing adoption of EHRs, the practice of medicine is growing increasingly data-driven, with intense interest in the use of predictive models to advance clinical care.<sup>31,32</sup> We would argue that it is time to shift some of the burden of such prognostication to automated systems that leverage modern data-mining techniques applied to very large datasets to achieve significantly higher accuracy than is possible with simpler models.

These limitations notwithstanding, our model, to the best of our knowledge, is the first that has been developed and validated on the full diversity of patients and wound types seen in clinical practice. It achieves significantly better discrimination than those of previous efforts, and is a case study for the meaningful secondary use of routinely collected EHR data to improve treatment strategies for surgically relevant disease.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

This study was funded by a research grant from Healogics Inc. NHS and KJ acknowledge additional funding from NIH grant U54 HG004028 for the National Center for Biomedical Ontology, NLM grant R01 LM011369, NIGMS grant R01 GM101430, and the Smith Stanford Graduate Fellowship. NS is scientific advisor and co-founder of Kyron Inc., and is an advisor to Apixio Inc. MJ is funded by NIH grant NIDDK DK-074095. CS, GG and MJ report

no conflicts of interest. RSK is a paid advisor to Healogics Inc. SC and DS are employed by Healogics Inc. We thank Jaime Campbell of Healogics Inc. for her assistance extracting data from the Healogics EHR.

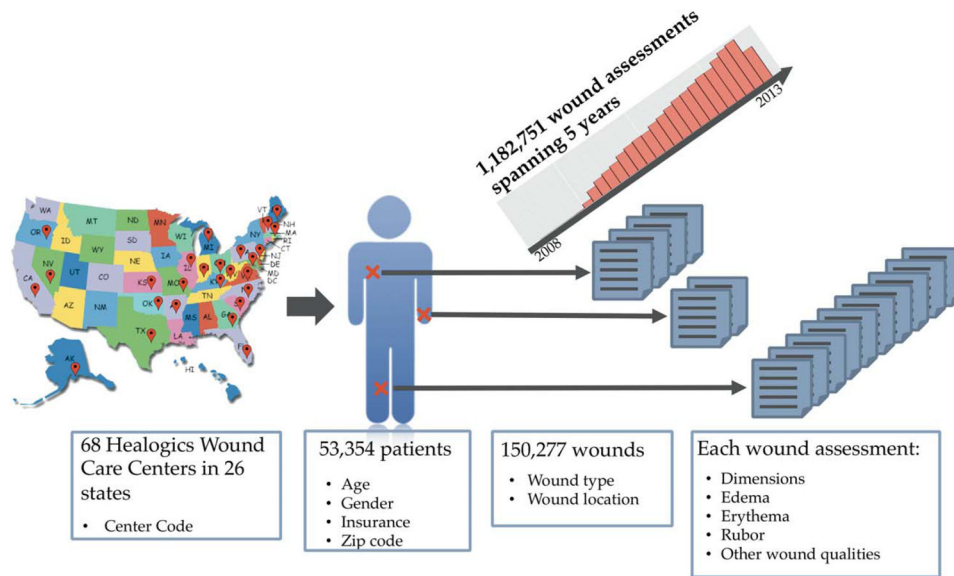
*Conflicts of Interest and Source of Funding:* This study was funded by a research grant from Healogics Inc. RSK is a paid advisor to Healogics Inc. SC and DS are employed by Healogics Inc.

## References

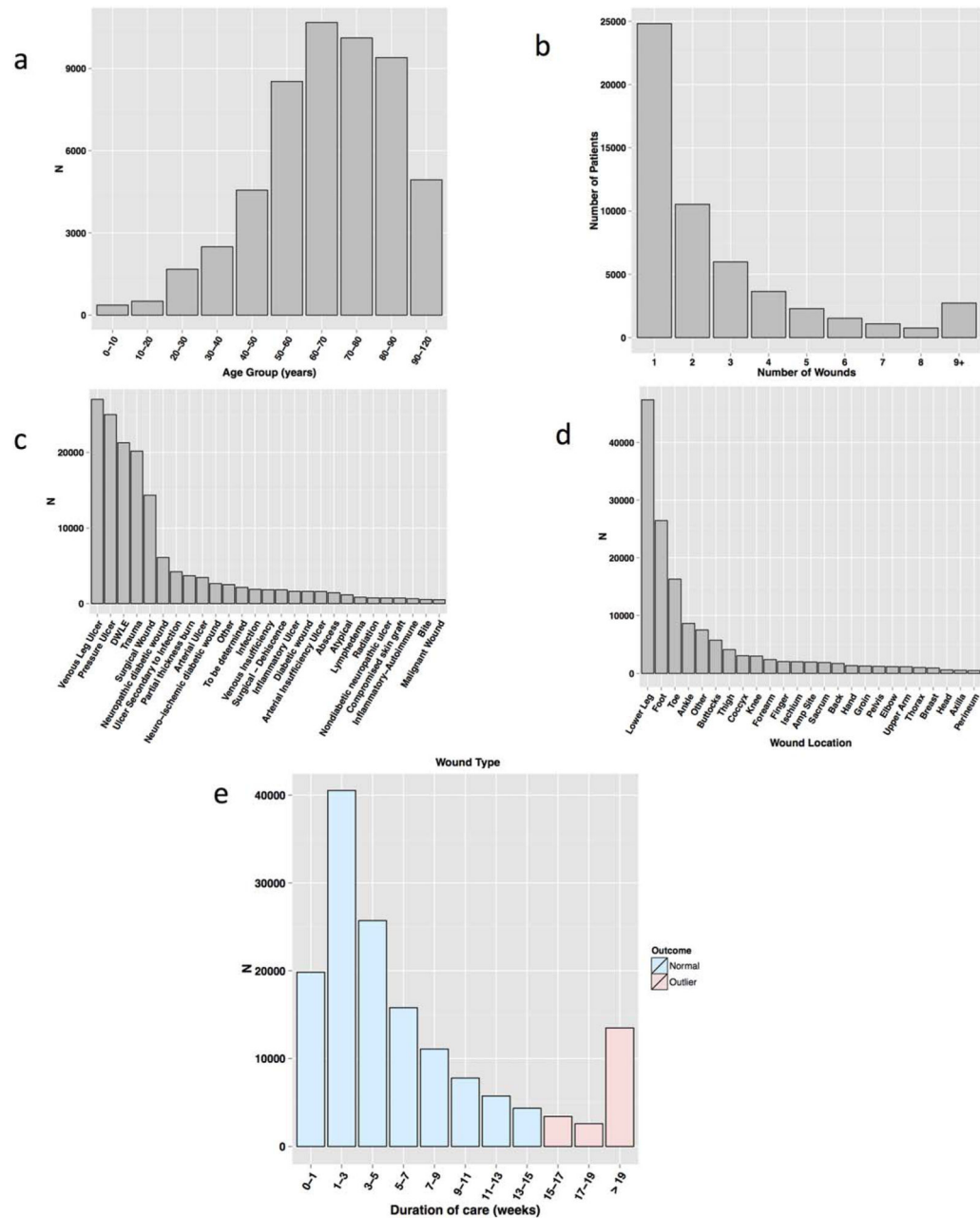
1. Lazarus GS, Cooper DM, Knighton DR, Percoraro RE, Rodeheaver G, Robson MC. Definitions and guidelines for assessment of wounds and evaluation of healing. *Wound Repair Regen.* 1994; 2:165–70. [PubMed: 17156107]
2. Fife CE, Carter MJ, Walker D, Thomson B. Wound care outcomes and associated cost among patients treated in US out-patient wound centers: data from the US wound registry. *Wounds.* 2012; 24:10–7. [PubMed: 25875947]
3. Driver VR, Fabbi M, Lavery LA, Gibbons G. The costs of diabetic foot: the economic case for the limb salvage team. *J Vasc Surg.* 2010; 52:17S–22S. [PubMed: 20804928]
4. Sen CK, Gordillo GM, Roy S, Kirsner R, Lambert L, Hunt TK, et al. Human skin wounds: a major and snowballing threat to public health and the economy. *Wound Repair Regen.* 2009; 17:763–71. [PubMed: 19903300]
5. Andros G, Armstrong DG, Attinger CE, Boulton AJ, Frykberg RG, Joseph WS, et al. Consensus statement on negative pressure wound therapy (V.A.C. Therapy) for the management of diabetic foot wounds. *Ostomy Wound Manage.* 2006; (Suppl):1–32. [PubMed: 17007488]
6. Armstrong DG, Ayello EA, Capitulo KL, Fowler E, Krasner DL, Levine JM, et al. New opportunities to improve pressure ulcer prevention and treatment: implications of the CMS inpatient hospital care present on admission indicators/hospital-acquired conditions policy: a consensus paper from the International Expert *Wound Care Advisory Panel.* *Adv Skin Wound Care.* 2008; 21:469–78. [PubMed: 18836326]
7. Clemens MW, Parikh P, Hall MM, Attinger CE. External fixators as an adjunct to wound healing. *Foot Ankle Clin.* 2008; 13:145–56. vi–vii. [PubMed: 18328418]
8. Melling AC, Leaper DJ. The impact of warming on pain and wound healing after hernia surgery: a preliminary study. *J Wound Care.* 2006; 15:104–8. [PubMed: 16550663]
9. Stojadinovic A, Carlson JW, Schultz GS, Davis TA, Elster EA. Topical advances in wound care. *Gynecol Oncol.* 2008; 111:S70–S80. [PubMed: 18793796]
10. Wu SC, Armstrong DG. Clinical outcome of diabetic foot ulcers treated with negative pressure wound therapy and the transition from acute care to home care. *Int Wound J.* 2008; 5:10–6. [PubMed: 18577133]
11. Margolis DJ, Allen-Taylor L, Hoffstad O, Berlin JA. Diabetic neuropathic foot ulcers: predicting which ones will not heal. *Am J Med.* 2003; 115:627–31. [PubMed: 14656615]
12. Margolis DJ, Allen-Taylor L, Hoffstad O, Berlin JA. The accuracy of venous leg ulcer prognostic models in a wound care system. *Wound Repair Regen.* 2004; 12:163–8. [PubMed: 15086767]
13. Cardinal M, Eisenbud DE, Phillips T, Harding K. Early healing rates and wound area measurements are reliable predictors of later complete wound closure. *Wound Repair Regen.* 2008; 16:19–22. [PubMed: 18211575]
14. Ubbink DT, Lindeboom R, Eskes AM, Brull H, Legemate DA, Vermeulen H. Predicting complex acute wound healing in patients from a wound expertise centre registry: a prognostic study. *Int Wound J.* 2015; 12:531–6. [PubMed: 24007311]
15. Kantor J, Margolis DJ. Efficacy and prognostic value of simple wound measurements. *Arch Dermatol.* 1998; 134:1571–4. [PubMed: 9875195]
16. Wicke C, Bachinger A, Coerper S, Beckert S, Witte MB, Konigsrainer A. Aging influences wound healing in patients with chronic lower extremity wounds treated in a specialized Wound Care Center. *Wound Repair Regen.* 2009; 17:25–33. [PubMed: 19152648]
17. Carter MJ, Fife CE, Walker D, Thomson B. Estimating the applicability of wound care randomized controlled trials to general wound-care populations by estimating the percentage of individuals excluded from a typical wound-care population in such trials. *Adv Skin Wound Care.* 2009; 22:316–24. [PubMed: 20375969]



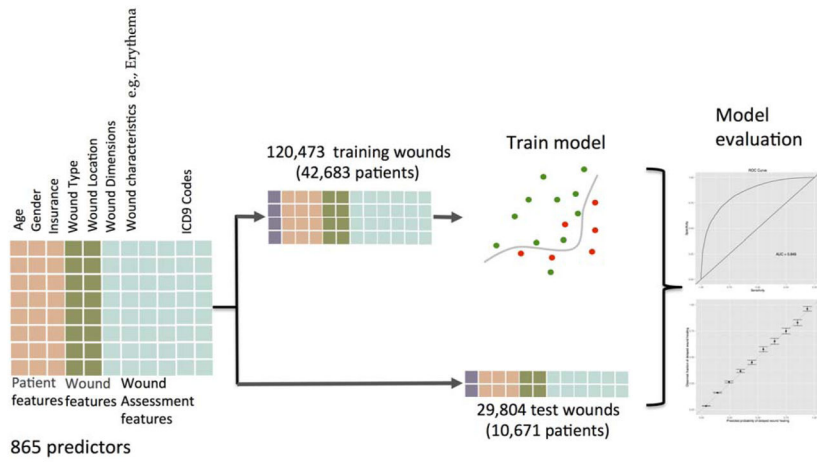
18. Kurd SK, Hoffstad OJ, Bilker WB, Margolis DJ. Evaluation of the use of prognostic information for the care of individuals with venous leg ulcers or diabetic neuropathic foot ulcers. *Wound Repair Regen.* 2009; 17:318–25. [PubMed: 19660039]
19. Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest.* 1991; 100:1619–36. [PubMed: 1959406]
20. Jordan MI, Wainwright MJ. Graphical models, exponential families, and variational inference. *Found Trend Mach Learn.* 2007; 1:1–305.
21. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw.* 2010; 33:1–22. [PubMed: 20808728]
22. Liaw A, Wiener M. Classification and regression by random-Forest. *R News.* 2002; 2:18–22.
23. Ridgeway, G. *gbm: generalized boosted regression models.* R Package Ver. 2.1. 2007. Available at <http://cran.r-project.org/web/packages/gbm/>
24. Margolis DJ, Bilker W, Boston R, Localio R, Berlin JA. Statistical characteristics of area under the receiver operating characteristic curve for a simple prognostic model using traditional and bootstrapped approaches. *J Clin Epidemiol.* 2002; 55:518–24. [PubMed: 12007556]
25. Brier GW. Verification of forecasts expressed in terms of probability. *Monthly Weather Rev.* 1950; 78:1–3.
26. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation.* 2007; 115:928–35. [PubMed: 17309939]
27. Cook NR. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. *Clin Chem.* 2008; 54:17–23. [PubMed: 18024533]
28. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat.* 2001; 29:1189–232.
29. Mjolsness E, DeCoste D. Machine learning for science: state of the art and future prospects. *Science.* 2001; 293:2051–5. [PubMed: 11557883]
30. Paxton C, Niculescu-Mizil A, Saria S. Developing predictive models using electronic medical records: challenges and pitfalls. *AMIA Annu Symp Proc.* 2013; 2013:1109–15. [PubMed: 24551396]
31. Kohane IS, Drazen JM, Campion EW. A glimpse of the next 100 years in medicine. *N Engl J Med.* 2012; 367:2538–9. [PubMed: 23268669]
32. Schneeweiss S. Learning from big health care data. *N Engl J Med.* 2014; 370:2161–3. [PubMed: 24897079]



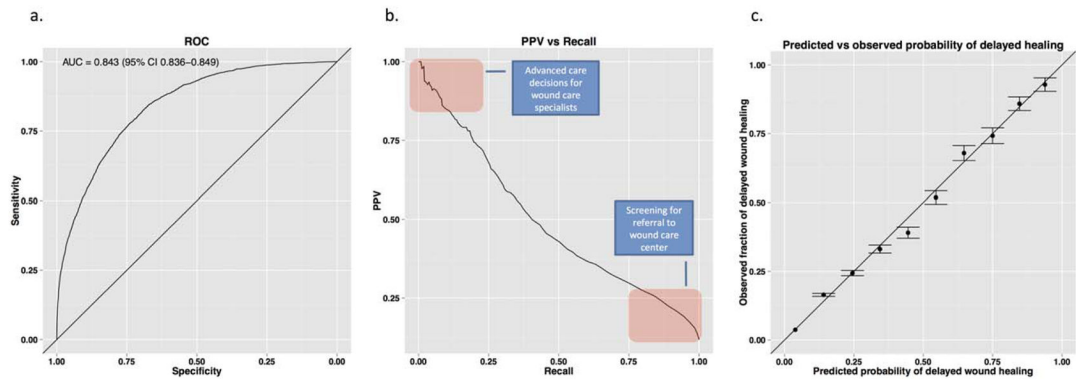
**Figure 1.** Dataset characteristics. The dataset is drawn from 68 Healogics wound care centers in 26 states over a period spanning 2009–2013 and comprises 181,716 wounds from 59,958 patients. Patient and wound information is recorded in an EHR at each weekly wound assessment. There were 40 distinct wound types spanning 37 anatomical locations. We use the information recorded in the first two assessments to predict delayed wound healing.



**Figure 2.** Patient and wound characteristics. (A) Distribution of patient age, binned by decade. (B) Number of patients with a given number of wounds. (C) Number of wounds by wound type. (D) Number of wounds by wound location. (E) Number of weeks until wound status resolution, in weeks. The threshold for delayed wound healing is 15 or more weeks; the pink bars correspond to delayed wound healing.



**Figure 3.** Model development and evaluation. A total of 865 predictors were used for each of the 150,277 wounds. The wounds were split into training and test sets. The training set was used to develop a prognostic model that was evaluated on the test set.



**Figure 4.**

Model evaluation. (A) Our model achieves an AUC curve of 0.842 (95% CI 0.836–0.849). (B) Because delayed wound healing is a low prevalence outcome, high specificity does not translate into high precision. The Precision vs. Recall plot for the final model shows that the tradeoff between precision and recall at different cutoffs for calling a case positive for delayed wound healing. (C) The probabilities of delayed wound healing output by the model on the test data agree well with observed frequencies of delayed wound healing (Brier reliability =0.00018 with patients stratified by predicted probabilities of delayed wound healing in bins of 0.1; standard error of the observed frequencies shown). Smaller values of Brier reliability indicate higher agreement actual observations with value of 0 indicating perfect agreement.

**Table 1**

Most influential features used by the model

Feature	Relative influence
PalliativeCare—yes	23.92
Fraction—iHealArea	11.09
Depth	7.04
iHealArea	4.96
AbsDelta—iHealArea	3.90
Granulation quality pale—no	2.94
Pending amputation on presentation—no	2.76
Fraction sum calculated area	2.66
Patient age	1.46
Fraction product area	1.44

iHealArea is an estimate of wound area recorded at the point of care. Fraction and AbsDelta features encode the fraction of quantitative measurements of wound dimensions at the second assessment relative to the same dimension at the first assessment (so small values mean the wound is smaller over time) and the absolute change in wound dimensions between the first and second assessments. The Sum and Product of areas refer to the sum and product of wound areas for all wounds for a given patient at the time of the first wound assessment for a given wound; these variables are intended to capture the total wound burden of a given patient at specific point in time.