

Predicting Emergency Department Visits

Sarah Poole^{1,2}, Shaun Grannis³ and Nigam H. Shah, MBBS, PhD¹

¹Stanford Center for Biomedical Informatics Research, Stanford, CA

²Stanford Biomedical Informatics Training Program, Stanford, CA

³Center for Biomedical Informatics, Regenstrief Institute, Indianapolis, IN

Abstract

High utilizers of emergency departments account for a disproportionate number of visits, often for non-emergency conditions. This study aims to identify these high users prospectively. Routinely recorded registration data from the Indiana Public Health Emergency Surveillance System was used to predict whether patients would revisit the Emergency Department within one month, three months, and six months of an index visit. Separate models were trained for each outcome period, and several predictive models were tested. Random Forest models had good performance and calibration for all outcome periods, with area under the receiver operating characteristic curve of at least 0.96. This high performance was found to be due to non-linear interactions among variables in the data. The ability to predict repeat emergency visits may provide an opportunity to establish, prioritize, and target interventions to ensure that patients have access to the care they require outside an emergency department setting.

Introduction

One of the central goals of the Affordable Care Act is to improve access to health care [1, 2], which was expected to decrease the number of emergency department (ED) visits as patients relied on their primary physicians rather than visiting the ED for non-emergency conditions [3]. However, a poll by the American College of Emergency Physicians showed that 75% of physicians reported seeing an increase in the number of ED visits since January 1, 2014, when the requirement to have health coverage took effect [4].

Patients who use the ED for non-urgent conditions are likely to return for multiple visits. High utilizers make up 4.5-8% of ED patients, but account for a disproportionate 21-28% of visits [5]. If patients can be identified as high-risk for repeat ED visits, then care management interventions have the potential to ensure that patients have access to necessary care without needing to return to the ED. Identifying these patients prospectively will allow such interventions to be deployed.

Previous studies have focused on predicting frequent ED use, and have found that previous ED utilization levels are an important predictor of future utilization [6-8]. Work on predicting revisit after an ED visit has focused on subgroups of patients with specific diseases or disorders, or has aimed to predict the time until the next ED visit [9-14]. Instead of predicting a continuous time until the next ED visit, our study aims to predict whether a patient will have an ED visit within a prespecified period of time following the index visit. Three outcome periods are investigated, namely one month, three months, and six months following the index visit. By accurately stratifying patients according to their risk of a repeat visit to the ED, clinical staff can ensure that interventions specifically target patients for whom they will be most helpful, maximizing benefits while keeping costs low. This approach was successfully used to increase cost-effectiveness of interventions for congestive heart failure patients, and has the potential to generalize across application areas [15]. For this to be possible, it is important that the probability of repeat visit produced by the chosen model corresponds to the risk of the patient returning to the ED within the outcome period. If the model can accurately produce such calibrated probabilities, clinical staff will be able to confidently make decisions regarding the need for additional care management intervention.

The dataset used in this study was provided by the Regenstrief Institute, and contains registration data from the Indiana Public Health Emergency Surveillance System (PHESS) covering a four-year period. Studies have shown that more than 40% of ED visits over a three-year period were for patients with medical data at

multiple institutions [16]. Leveraging PHESS for this study reduces the likelihood that patient visits are missing from the data.

A similar dataset was used by Wu et al. to identify patients who are likely to be high users of EDs [17]. Our study has a similar aim, but involves transformation of the data set and more sophisticated machine learning techniques. The resulting increase in prediction accuracy, might allow for highly targeted interventions for patients deemed to be at high risk of revisit in a prespecified time frame.

Methods

Data

The Regenstrief Institute provided a database of registration data from the Indiana Public Health Emergency Surveillance System for this analysis. This data covers a four year period with a total of 1,125,118 patients are represented in this database. **Error! Reference source not found.** Table 1 lists the characteristics of the patient cohort.

The original data set contained one row per visit, with very few features. Specifically, the features in the data set were:

- Pseudonymized patient identifier
- Patient age at the time of the visit
- Chief complaint (containing up to two main complaints)
- Patient gender
- Zip code
- ED location code
- Date of death (if patient is deceased)
- Date of visit
- Distance from the centroid of the patient's zip code to nearest ED

The data set was transformed so that each row represented one patient rather than a visit. Three data sets were created, one for each outcome period of one month, three months, and six months. For each outcome period, data for each patient was split into non-overlapping sections of a one-year prediction period ending at an emergency department visit and an additional outcome period. Figure 1 shows how data for a single patient would be split into non-overlapping sections.

Additional features were engineered from the original data. The features used for analysis were:

- Patient age
- Patient gender
- Zip code
- Distance from centroid of patient's zip code to nearest ED
- The number of ED visits during the year-long prediction period
- The mean of the times between ED visits during the year-long prediction period
- The standard deviation of the times between ED visits during the year-long prediction period
- The slope of the line fitted to the times between ED visits during the year-long prediction period
- Most common chief complaint during the year-long prediction period
- A measure of the variance in chief complaint (defined as the number of mentions of the most common chief complaint divided by the number of different chief complaints. A high number indicates that a single condition is driving ED visits, while a lower number indicates the patient suffers from several different problems).
- Most common ED location
- A measure of variance in the ED location (defined as the number of mentions of the most common ED location, divided by the number of different ED location mentions).

Table 1: Characteristics of the patient cohort

Total patients	1,125,118
Gender split (% female)	51.9%
Age (years)	Median: 35 Range: 0 - 89
Distance to ED	Median: 6.835 Range: 0.007-12010
Total visits	2675750
Visits per patient:	Median: 1 Mean: 2.378 Range: 1 - 319

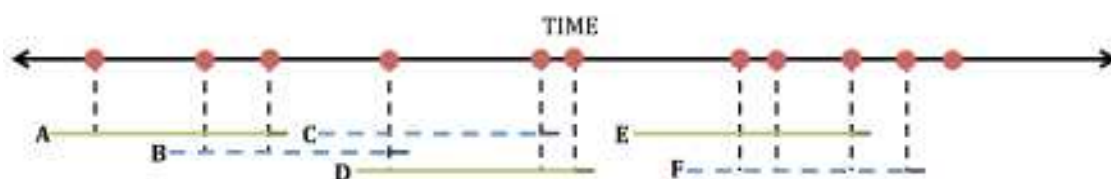


Figure 1: Representation of the emergency department visits for a single patient. The red dots along the timeline indicate ED visits. Both green and dashed blue lines span a year in time before an ED visit, and one month following the ED visit (this outcome period is shown in darker color). Time periods must be non-overlapping. Time periods must contain at least three ED visits. Line B will not be used because it overlaps with line A. Line C will not be used because there are only two ED visits during the time spanned. Line E will be used rather than line F, even though line F spans more ED visits, because line E is chronologically first.

Table 2: Number of rows and patients in training and test sets, along with the proportion of rows with positive outcomes.

	Outcome period:	Six months	Three months	One month
Rows:	<i>Training set:</i>	115674	121536	125244
	<i>Testing set:</i>	34921	35424	35864
Unique patients:	<i>Training set:</i>	104764	106275	107591
	<i>Testing set:</i>	34921	35424	35864
% positive outcomes:	<i>Training set:</i>	15%	10%	3%
	<i>Testing set:</i>	10%	6%	2%

Because multiple visits were combined to find the data for each row, it was possible for features (particularly age or zip code) to take multiple values. When this occurred, the most common value was used in the final data set. Including the slope between ED visits during the first year of data meant that only patients with at least three ED visits during the first year could be included. Since less than 3% of patients with less than three visits during the first year have another visit within three months, applying this constraint results in minimal data loss.

The data were split into training (80%) and testing sets. This split was done at the patient level, so the test set shared no patients with the training set. Although the training set could contain multiple rows of data representing the same patient, the test set contained one randomly selected row per patient. Table 2 shows the number of rows and patients in the training and test sets for each outcome period. Table 2 also gives the proportion of each data set that had the outcomes we attempt to predict (i.e., had a visit within the corresponding outcome period).

The outcome of interest was whether the patient returned to the ED within the outcome period. If the patient died during the outcome period, this was also recorded as a positive outcome.

Predictive Modeling

A range of predictive models were tested, including a logistic regression model with no regularization, models using lasso and ridge regularization, stepwise feature selection models, a support vector machine model (using a linear kernel) and a random forest model. Second order interaction terms were added to the feature set, and a lasso model was trained on this extended feature set. R packages `glmnet`, `e1071` and `randomForest` were used to construct these models [18-20]. Receiver operating characteristic (ROC) curves were constructed to show the performance of each classifier on the held out test set.

The calibration of the models was also investigated, by splitting the patients into ten groups corresponding to deciles of predicted probability of revisit. The mean and standard deviation of the predicted probability of revisit for each group was calculated, and plotted against the actual proportion of patients in each group who had a repeat visit within the outcome period. The Brier score, a measure of the calibration of the model, was calculated.

Results

Figures 2(a), 3(a) and 4(a) show ROC curves for the three outcome periods studied. The performance of the random forest models, lasso models using the original feature set, and lasso models including second order interaction terms are shown. All other models exhibited very similar performance to the original lasso model.

The random forest model had extremely good performance for all outcome periods. Adding interaction terms to the lasso model also improves its performance considerably.

The random forest models all selected the slope of the times between successive ED visits as the most important feature, followed by the number of ED visits in the prediction period, and the mean and standard deviation of the times between ED visits.

Table 3 lists the most important features in the random forest model, for each of the three outcome periods.

Figures 2(b), 3(b) and 4(b) show the calibration of the model for the three outcome periods. Table 4 shows the Brier scores for the three models trained. All three models have good calibration, with the best calibration seen for the 3-month outcome period.

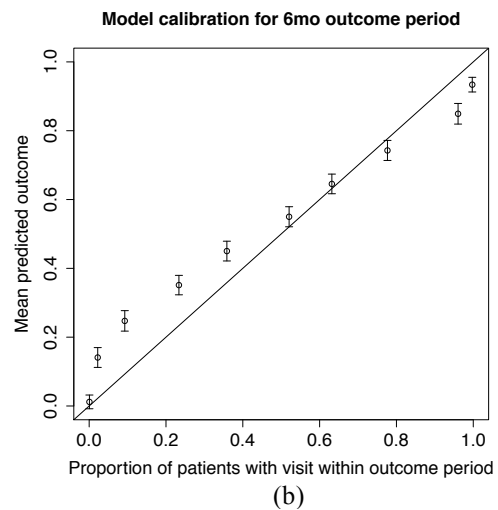
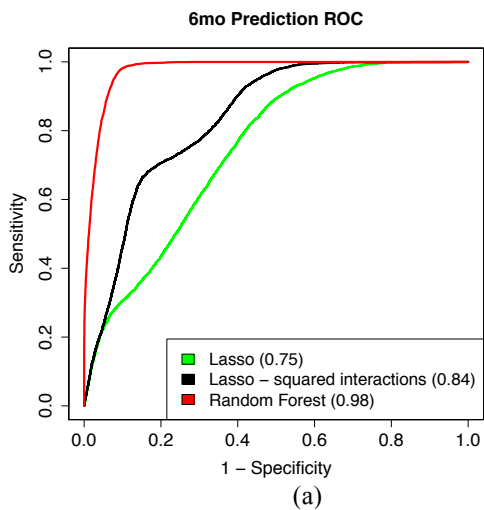


Figure 2: (a) ROC curve for outcome period of 6 months, and (b) calibration of random forest model for outcome period of 6 months. Brier score = 0.0082.

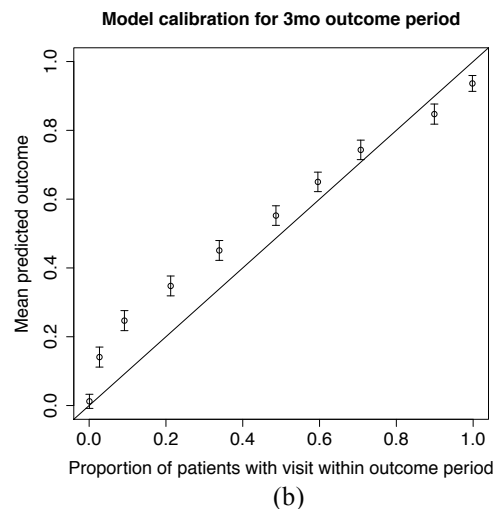
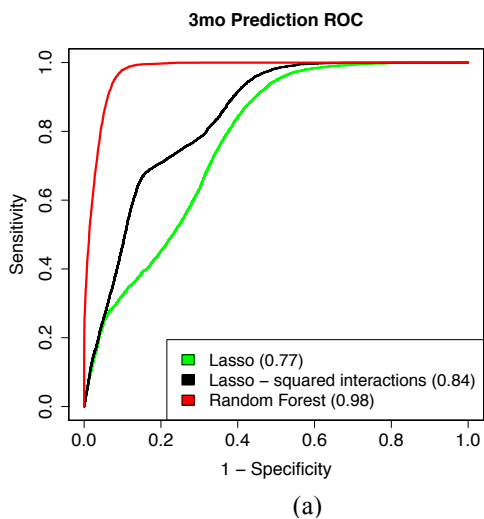


Figure 3: (a) ROC curve for outcome period of 3 months, and (b) calibration of random forest model for outcome period of 3 months. Brier score = 0.0087.

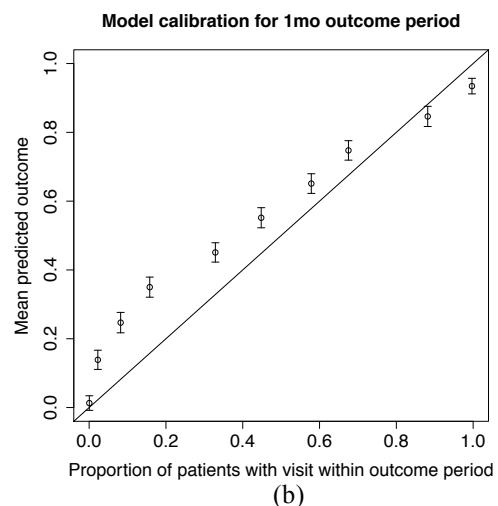
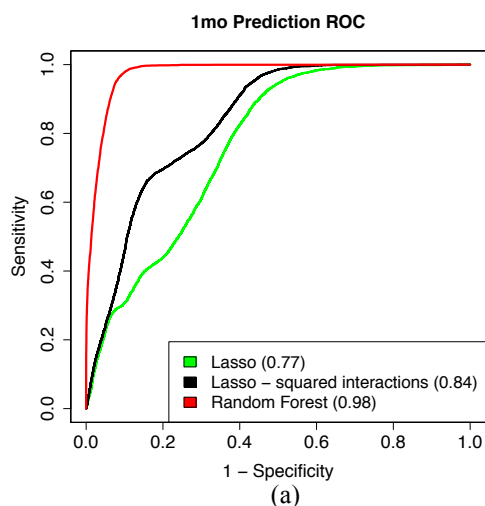


Figure 4: (a) ROC curve for outcome period of 1 month, and (b) calibration of random forest model for outcome period of 1 month. Brier score = 0.012.

Table 3: Five most important features in random forest models (note: this importance ranking is the same for all three outcome periods)

Feature importance rank:	Feature
1	slope(times between visits)
2	stddev(times between visits)
3	Number of visits
4	mean(times between visits)
5	age

Table 4: Brier score (quantifying model calibration) for models for the 3 outcome periods, given to 2 significant figures

Outcome Period	Brier Score (2 sf)
6 months	0.0082
3 months	0.0087
1 month	0.012

Discussion

The random forest model performed exceptionally well for all three outcome periods, with areas under the ROC 0.98. The addition of squared interaction terms to the lasso model improved performance considerably, suggesting the presence of significant non-linear interactions. Random forest models are particularly well suited for exploiting interactions in the data, so the presence of these non-linear interactions may explain the difference in performance between the models.

The random forest model for all three outcome periods shared the five most important features. Three of these top five features were built using the times between ED visits. The slope of the times between visits, which indicates whether the patient is visiting the ED more regularly over time, was the most important feature. The standard deviation of these times between visits was the second most important feature, and the mean was the fourth most important. The raw number of visits during the training period was the third most important feature, and age was the fifth most important. The variance of the chief complaint was consistently a more important feature than the type of chief complaint, suggesting that condition complexity could be a driver of ED usage.

The calibration of the models was calculated, and all three models had good calibration. The best calibration was seen for the 3-month outcome period, which is clinically the most actionable time frame. The poorest calibration was seen in the model for the 1-month outcome period, but this calibration is most variable at the low risk level, so the model will still be reliable for use in clinical decision making.

The high calibration of the random forest model indicates that the risk of a patient having a repeat visit within the outcome period corresponds well to the predicted probability of a repeat visit, meaning that the predicted probability of repeat visit is reliable, and patients can be accurately stratified according to their risk for repeat visits. This ensures that the interventions are given to the group of patients for whom they will have the most benefit, maximizing benefits of the interventions while keeping costs low.

The excellent performance of the random forest model suggests that clinical decision support tools may be useful in the management of emergency room patients at risk for re-admission. The model presented can be run before the time of discharge from an ED visit, and predicts whether the patient will have another ED visit within the prediction window. This allows interventions to be carried out while the patient is in the hospital setting.

Interventions to prevent these repeat visits could include spending additional time with the patient to ensure that they have a primary care physician who is accessible to them, so that they do not need to visit the ED for non-emergency conditions. Further, the patient's primary care provider may be notified of the ED visit and follow-up outpatient care plans can be initiated. Checks should also be made to ensure that the patient is satisfied with their status before they leave the ED, and to identify whether they have appropriate support at home. Additional patients education addressing home self-care could also be provided at this time.

Limitations

The prediction periods used for the three models were different, therefore it is not possible to compare their performance directly. If implemented, the same prediction period would be used for all models, and the results could be used to narrow down the time until the patient is expected to revisit the ED. For example, if the one month and three month models predict no return visit but the six month model predicts a return visit, then the time until the return visit would be expected to be between three and six months. This data could be used to choose the intervention that is given.

Conclusion

Routinely recorded registration data can be used to predict subsequent visits to the ED, with good performance and calibration for outcome periods of one month, three months and six months. The ability to predict future ED use at the index visit can allow interventions to be deployed while the patient is in a hospital setting. Decreasing repeat visits to the ED has the potential to greatly decrease unnecessary ED use, as high users of the ED account for a disproportionately large number of ED visits.

References

1. *Patient Protection and Affordable Care Act, 42 U.S.C., Section 18001 et seq.* . 2010: United States of America.
2. Assistant Secretary for Planning and Evaluation (ASPE), U.S.D.o.H.H.S., *Strategic Plan*. 2015.
3. Billings, J. and M.C. Raven, *Dispelling an urban legend: frequent emergency department users have substantial burden of disease*. Health Aff (Millwood), 2013. **32**(12): p. 2099-108.
4. *2015 ACEP Poll Affordable Care Act Research Results*. 2015, American College of Emergency Physicians.
5. LaCalle, E. and E. Rabin, *Frequent users of emergency departments: the myths, the data, and the policy implications*. Ann Emerg Med, 2010. **56**(1): p. 42-8.
6. Grinspan, Z.M., J.S. Shapiro, E.L. Abramson, et al., *Predicting frequent ED use by people with epilepsy with health information exchange data*. Neurology, 2015.
7. Neufeld, E., K.A. Viau, J.P. Hirdes, and W. Warry, *Predictors of frequent emergency department visits among rural older adults in Ontario using the Resident Assessment Instrument-Home Care*. Aust J Rural Health, 2015.
8. Ko, M., Y. Lee, C. Chen, P. Chou, and D. Chu, *Prevalence of and Predictors for Frequent Utilization of Emergency Department: A Population-Based Study*. Medicine (Baltimore), 2015. **94**(29): p. e1205.
9. Newton, A.S., R.J. Rosychuk, X. Niu, A.D. Radomski, and P.J. McGrath, *Predicting time to emergency department return for anxiety disorders and acute*

- stress reactions in children and adolescents: a cohort study.* Soc Psychiatry Psychiatr Epidemiol, 2015. **50**(8): p. 1199-206.
10. Suffoletto, B., T. Miller, R. Shah, C. Callaway, and D.M. Yealy, *Predicting older adults who return to the hospital or die within 30 days of emergency department care using the ISAR tool: subjective versus objective risk factors.* Emerg Med J, 2015.
 11. Sung, S.F., K.E. Liu, S.C. Chen, et al., *Predicting Factors and Risk Stratification for Return Visits to the Emergency Department Within 72 Hours in Pediatric Patients.* Pediatr Emerg Care, 2015.
 12. Ram, S., W. Zhang, M. Williams, and Y. Pengetnze, *Predicting Asthma-Related Emergency Department Visits Using Big Data.* IEEE J Biomed Health Inform, 2015. **19**(4): p. 1216-23.
 13. Morris, J.N., E.P. Howard, K. Steel, et al., *Predicting risk of hospital and emergency department use for home care elderly persons through a secondary analysis of cross-national data.* BMC Health Serv Res, 2014. **14**: p. 519.
 14. Barrett, T.W., C.A. Jenkins, and W.H. Self, *Validation of the Risk Estimator Decision Aid for Atrial Fibrillation (RED-AF) for predicting 30-day adverse events in emergency department patients with atrial fibrillation.* Ann Emerg Med, 2015. **65**(1): p. 13-21 e3.
 15. Bayati, M., M. Braverman, M. Gillam, et al., *Data-driven decisions for reducing readmissions for heart failure: general methodology and case study.* PLoS One, 2014. **9**(10): p. e109264.
 16. Finnell, J.T., J.M. Overhage, and S. Grannis, *All health care is not local: an evaluation of the distribution of Emergency Department care delivered in Indiana.* AMIA Annu Symp Proc, 2011. **2011**: p. 409-16.
 17. Wu, J., H. Xu, J. Finnel, and S. Grannis, *A Practical Method for Predicting Frequent Use of Emergency Department Care Using Routinely Available Electronic Registration Data.* AMIA Annu Symp Proc, 2013.
 18. Friedman, J., T. Hastie, and R. Tibshirani, *Regularization Paths of Generalized Linear Models via Coordinate DEscent* Journal of Statistical Software, 2010. **33**(1): p. 1-22.
 19. Meyer, D., E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch, *e1071: Misc Functions of the Department of Statistics (e1071), TU Wein. R package version 1.6-4.* 2014.
 20. Liaw, A. and M. Weiner, *Classification and Regression by randomForest.* R News, 2002. **2**(3): p. 18-22.