International Journal of Extreme Automation and Connectivity in Healthcare
Volume 2 • Issue 2 • July-December 2020

Predicting Patient Admission From the Emergency Department Using Administrative and Diagnostic Data

David W. Savage, Northern Ontario School of Medicine, Canada
Douglas G. Woolford, University of Western Ontario, Canada
Mackenzie Simpson, Lakehead University, Canada
David Wood, Thunder Bay Regional Health Sciences Centre, Canada
Robert Ohle, Northern Ontario School of Medicine, Canada

ABSTRACT

Emergency department (ED) overcrowding is a growing problem in Canada. Many interventions have been proposed to increase patient flow. The objective of this study was to predict patient admission early in the visit with the goal of reducing waiting time in ED for admitted patients. ED data for a one-year period from Thunder Bay, Canada was obtained. Initial logistic regression models were developed using age, sex, mode of arrival, and patient acuity as explanatory variables and admission yes or no as the outcome. A second stage prediction was made using the diagnostic tests ordered to further refine the predictive models. Predictive accuracy of the logistic regression model was adequate. The AUC was approximately 81%. By summing the probabilities of patients in the ED, the hourly prediction improved. This study has shown that the number of hospital beds required on an hourly basis can be predicted using triage administrative data.

KEYWORDS

Admission, Area Under the Curve, Diagnostic Investigations, Emergency Department, Hospital, Logistic Regression, Prediction, Triage

INTRODUCTION

Overcrowding in Canadian emergency departments (ED) is common. It occurs when the demand for services exceeds the ability of the ED to provide timely care (Rowe et al. 2006). In the ED, both physicians and patients often experience negative impacts related to overcrowding (Derlet and Richards 2000). Interventions to reduce the impact of overcrowding can be classified into strategies that affect patient input, throughput, and output (Bond et al. 2006).

The majority of interventions to decrease overcrowding and increase patient flow involve the input and throughput stages of the patient visit. Input interventions often involve either patient diversion to other care providers or the modification of triage processes. From a throughput perspective, the use of alternate treatment streams or a fast-track area where high and low acuity patients can be separated have been studied extensively (Considine et al. 2008, Chrusciel et al. 2019). As well, staffing changes are another strategy to increase ED throughput. Some of the output interventions that have been

DOI: 10.4018/IJEACH.2020070101

studied have involved the creation of short stay units or processes to accelerate patient registration and transfer out of the ED (Bond et al. 2006).

Other studies have proposed another output intervention to identify patients early in their ED visit who are likely to require admission to hospital. This intervention would allow the process of resource allocation and bed planning to begin prior to the physician submitting the admission request or order. Examples include Sun et al. (2011), Peck et al. (2012) and Hong et al. (2018), who studied the problem using logistic regression and other statistical techniques to predict patient admission from administrative triage data. These models typically use age, gender, patient acuity, and mode of arrival as explanatory variables in the modelling. Others have attempted to incorporate patient vitals in the prediction model with varying success (LaMantia et al. 2010, Caterino et al. 2017). Non-statistical techniques of admission prediction have also been proposed including predictions made by triage nursing staff (Peck et al. 2012, Cameron et al. 2016).

The objective of this study was to develop statistical models for predicting patient admission to hospital with the goal of early admission planning to decrease patient time in the ED. Two different types of models were developed, an initial stage model that incorporated information known at the time of triage, and a second-stage model that incorporated information about decisions made by the physician in the ED. Specifically, triage administrative data was used to calculate an initial probability of admission with a second stage or updated probability calculated after the physician initial assessment. The second stage probability calculation was intended to improve the prediction accuracy using the presence or absence of certain diagnostic tests (i.e., laboratory and imaging data) being ordered by the physician early in the visit. For patients with a high probability of admission, resource planning (e.g., hospital bed allocation) can begin to reduce delays between admission and transfer to the ward.

METHODS

Study Setting

The Thunder Bay Regional Health Sciences Centre (TBRHSC) is a regional referral centre in Northwestern Ontario, Canada for both pediatric and adult patients. The ED experiences annual patient volumes of approximately 108,000 visits per year. Arriving patients are placed in either an acute care or fast-track queue depending on acuity. Higher acuity patients classified under the Canadian Triage and Acuity Scale (CTAS) as either a level I, II, of III are treated in the acute care area while CTAS IV and V are typically but not always treated in the fast-track area.

Data

Retrospective data for the period May 2016 to April 2017 was obtained. The initial data set consisted of 108,704 anonymized individual patient records with demographic data, treatment area, triage level, presenting complaint, admission to hospital (i.e., yes or no), dates and times for arrival, physician initial assessment, admission or departure from the ED. In addition, laboratory and diagnostic imaging data were obtained for each patient as separate tables and then attached to each individual patient record in the main data set for analysis. Patient records missing data were excluded. Given that the majority of patients admitted to hospital were assessed and treated in the acute care area, all patients in the fast-track stream of the ED were also excluded. As well, pediatric patients less than 18 years of age were excluded since the criteria for admitting these patients is often highly variable. The final data set consisted of 56,060 records. Table 1 describes the data set.

In order to assess the predictive ability of the fitted models and to protect against overfitting, the data was split into training (i.e., model development) and testing (i.e., model validation) data sets. The training data set consisted of 80% of the records and was randomly chosen.

Table 1. Summary of triage administrative variables used in the predictive modelling

Variable	Admitted to Hospital	
	No	Yes
Sex		
Female	56.4%	51.2%
Male	43.6%	48.8%
Age (Mean)	50.0	62.9
Arrival Mode		
Ambulance	66.8%	33.2%
Family vehicle	82.0%	18.0%
Other	62.1%	37.9%
Police	66.3%	33.7%
Walk-in	88.0%	12.0%
Wheelchair	68.7%	31.3%
CTAS		
I	36.8%	63.2%
II	73.5%	26.5%
III	89.5%	10.5%
IV	95.6%	4.4%
V	97.9%	2.1%
Total	45,816 (81.7%)	10,244 (18.3%)

STATISTICAL ANALYSIS

The statistical analysis to predict patient admission was completed in two stages. The initial prediction probability used triage administrative data collected when the patient first arrived in the ED including: age, sex, acuity, and mode of arrival (Table 1), which were used as predictors in a logistic regression model framework that used admission status (i.e., as 0=not admitted and 1=admitted) as the dependent variable.

The second stage of the predictive modelling used the presence or absence (i.e., 1 or 0) of several laboratory and diagnostic imaging tests as the next set of predictors. The process of selecting these variables is described below. Once the laboratory and diagnostic imaging variables were selected, they were coded as either 1 or 0 for each patient depending on whether they were ordered for the patient. The authors did not use the results of the ordered investigations because getting the results can take several hours for some of the tests. Instead, they used the presence or absence of the test being ordered as a proxy for the physician's assessment of severity of illness and how likely the patient was to be admitted to hospital. In the second stage statistical model, the triage variables from the first model and the binary investigation variables were combined in a second logistic regression model. Model evaluation is described below.

The generalized linear model (glm) function in the R statistical computing software package (R Core Team 2015) was used for the logistic regression analyses.

Analysis of Diagnostic Investigation Data

The presence or absence of laboratory and imaging data for each patient was used to refine the probability estimate for admission. Not all laboratory tests were assessed given that some tests are ordered as a panel and will therefore be correlated. For example, we used hemoglobin as a proxy for a complete blood count (CBC). The CBC contains several other hematological measures (e.g., white blood cell count, platelets, etc.) that will have identical pattern of presence and absence to the hemoglobin. The other consideration when selecting the subset of investigations was the frequency in which the tests were ordered. Rare investigations may have high specificity for admission but given how infrequently they were used, they provide little benefit in predicting admission in most patients. Therefore, only tests (i.e., lab and imaging) that were ordered for a minimum of 1% of the patients were included in the analysis.

To find the subset of tests to include in the second stage prediction, the authors used LASSO logistic regression. This technique fits a logistic regression model by minimizing a penalized function of the negative log-likelihood function for a logistic GLM along with a constraint that the sum of the magnitudes of the regression coefficients cannot exceed a specified value (Friedman et al. 2010). By including this constraint, some variable coefficients are forced to zero, thus eliminating them from the model and generating a subset of variables from the overall possible predictors. In other words, LASSO can be used as a variable selection technique for model building. From the available pool of investigations that met the minimum frequency (i.e., at least 1% patients received the test) the subset of laboratory investigations in the initial analysis included: hemoglobin, troponin, INR, lactic acid, magnesium, venous blood gas, albumin, Acetaminophen level. The imaging investigation included: x-ray and computed tomography (CT).

The glmnet function (Friedman et al. 2010) in the R statistical computing software package (R Core Team 2015) was used for the LASSO logistic regression analyses.

Model Evaluation

We first assessed the odds ratios and confidence intervals for the statistical models. The initial logistic regression model was then applied to the testing data set to find the probability of admission for each patient in the testing data set. A receiver operating characteristic (ROC) curve was generated and the area under the curve (AUC) was calculated (Fawcett 2006). Model sensitivity, specificity, positive predictive value and negative predictive value were also calculated.

The ROC and AUC (Fawcett 2006) were calculated using the package ROCR (Sing et al. 2015) in the R statistical computing software package (R Core Team 2015).

Although the objective of this study was to predict individual patient admission to hospital, Peck et al. (2012) had success with aggregating the probabilities of admission and calculating the mean number of beds required on an hourly basis. This was accomplished by calculating the cumulative hourly admission probability for each day to find the mean number of beds. For example, if four patients arrived within an hour and each had probabilities of admission of 0.1, 0.2, 0.3, and 0.4, none would be classified as being admitted given that they are all below the typical threshold probability of 0.5. However, by summing the probabilities to 1.0, on average 1 bed would be required. This analysis was completed for this study; the hourly cumulative sum of admission probabilities was compared to actual admissions over a 10 day period.

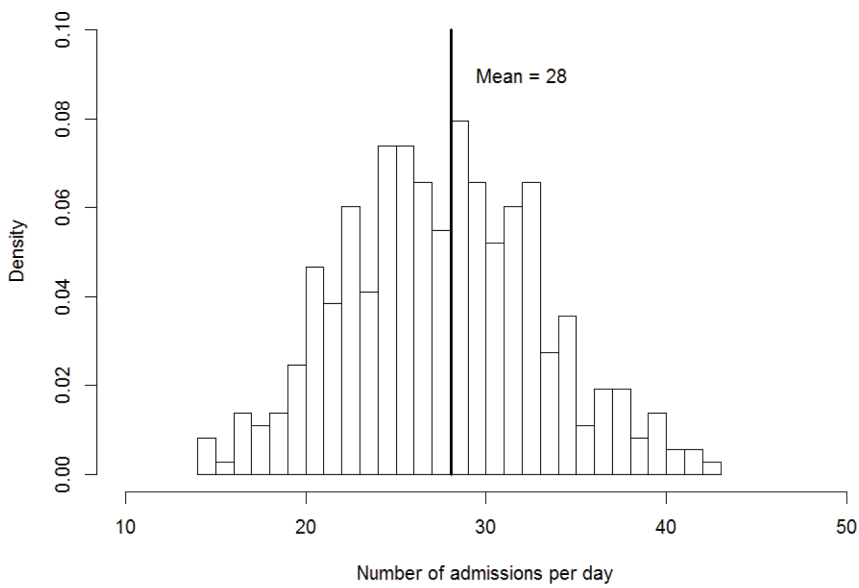
RESULTS

Assessment of Current ED Admission Characteristics

We first assessed the variability in daily admissions (Figure 1). The minimum number of admissions was 14 and the maximum 43. The first quartile, median, and third quartile were 24, 28, and 32, respectively.

The time from patient arrival to physician initial assessment (PIA) and the time from arrival to admission was assessed (Figure 2). There is some overlap between these distributions but the mean difference between PIA time and admission time was on average 216 minutes.

Figure 1. The distribution of daily admissions throughout the year



Predicting Admission to Hospital

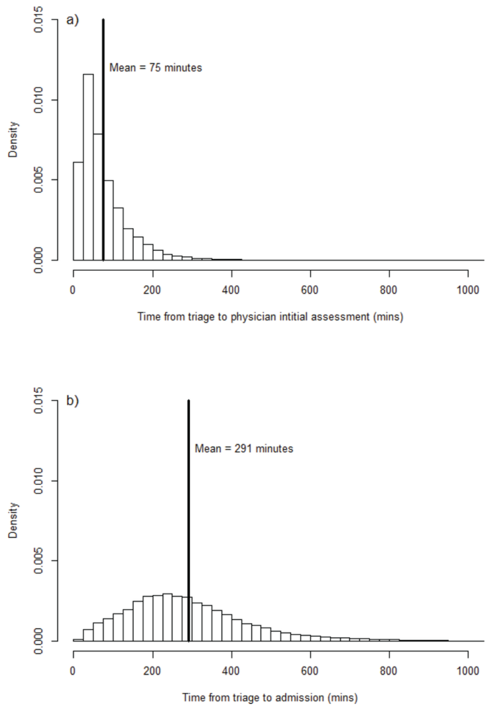
For the logistic regression model, the log odds were transformed using the exponential function to find the odds ratio for each variable as well as the confidence interval (Table 2). Patients who arrived by police car and patients triaged as CTAS I and II had 6, 37 and 10 times greater odds of being admitted, respectively. The 95% confidence intervals for the log odds of patient acuity were quite variable with an upper limit of 123 and lower limit of 15 for CTAS I patients.

An ROC curve was developed for the first and second stage logistic regression models using the triage administrative data and the diagnostic data (Figure 3). The ROC curve shows a small improvement with the addition of the diagnostic data.

For the first and second stage logistic regression models, summary statistics were calculated (Table 3). In general, the specificity and negative predictive value were high with a small improvement in the AUC with the addition of the diagnostic data.

As a further assessment of model performance, the actual hourly cumulative admissions were plotted and compared to the model-based prediction for the expected number of beds (i.e., obtained by aggregating the predicted probabilities) (Figure 4). A total of 10 random days are presented. The top panel shows the model fit for the first stage logistic regression model alone. The model under predicts admission on days 2, 5, 8 and 10 while it over predicts on day 3. When the model was refined with the diagnostic data (bottom panel), the days of under prediction improved, however, on day 7 the model now over predicts admission.

Figure 2. The distribution of times for a) The length of time from patient arrival to physician initial assessment; and b) The time from patient arrival to admission to hospital



DISCUSSION

In this study, we have developed a statistical modelling framework for predicting the initial probability of hospital admission using triage administrative data and then refined that probability using a second stage prediction model that incorporated additional diagnostic test data information. The initial model had satisfactory predictive ability and the diagnostic test data led to a small improvement in the model. Although, there was reasonable predictive ability for individual patients, when the probabilities were pooled (i.e., summed), the model performed better across all patients. The risk pooling of patient probabilities in this study was similar to Peck et al. (2012) and provided a prediction that was informative.

For this predictive modelling to improve patient flow, there were two conditions that were important: 1) the variability in daily admissions must be significant, otherwise the mean admission rate would be adequate for the planning of daily bed needs and 2) if diagnostic testing data were to be used for admission prediction, the time between test ordering and hospital admission must be sufficiently long. These two conditions were true in the ED studied, the number of daily admissions was quite variable and the time between test ordering and admission was sufficiently long that this system would likely provide benefit from an ED patient flow perspective.

The AUC was one of the model evaluation measures used in this study and was comparable to other studies with the same objective and similar predictive variables. In Sun et al. (2011), the AUC was 0.849, while in Peck et al. (2012) their AUC was 0.887. Hong et al. (2018) were able to develop predictive models with AUC of 0.9-0.92 using machine learning techniques. Similar to this study, they also were able to collect variables from the physician's initial assessment (i.e., characteristics of the patient's history) to refine and improve the models. Although the AUC is a common performance measure for evaluating diagnostic tests, it can be misleading (Zhou et al. 2011). Therefore, other

Table 2. Summary of odds ratios and confidence intervals for the logistic regression model

Variables	Odds Ratio	95% Confidence Interval
Gender		
Male	1.11	(1.05, 1.17)
[Female]	1.00	
Age	1.02	(1.02, 1.03)
Arrival Mode		
Ambulance	2.47	(2.33, 2.61)
Family vehicle	1.25	(0.56, 2.52)
Other	4.66	(1.69, 11.96)
Police car	6.21	(5.17, 7.45)
Wheelchair	2.17	(1.79, 2.62)
[Walk-in]	1.00	
CTAS		
I	37.24	(15.40, 122.64)
II	9.97	(4.15, 32.75)
III	3.67	(1.53, 12.06)
IV	1.53	(0.61, 5.15)
[V]	1.00	
[ ] = Reference group		

Figure 3. The receiver operating characteristic curves comparing the logistic regression model alone to the logistic regression model plus diagnostic data

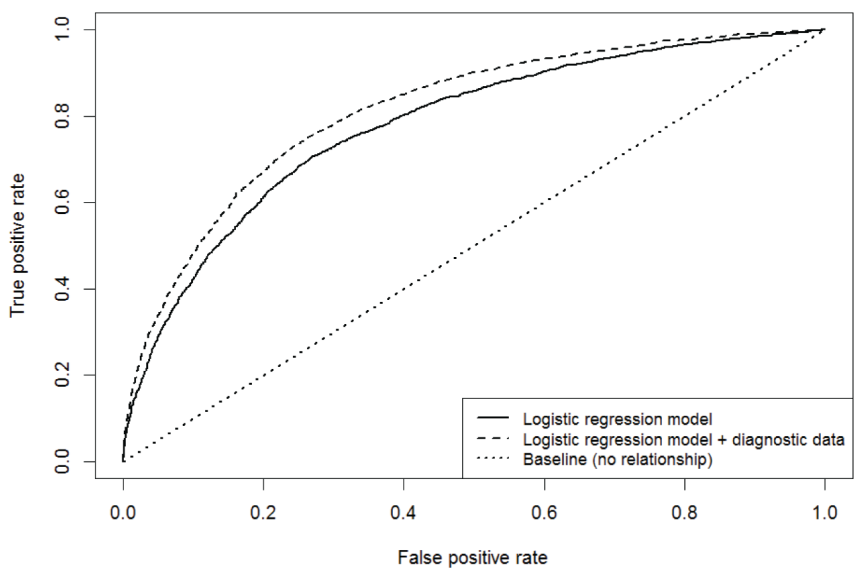
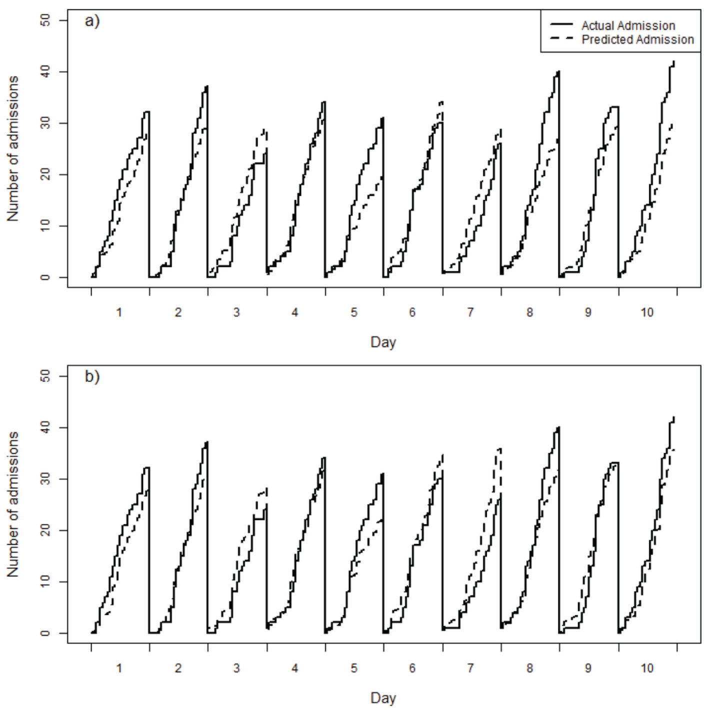


Table 3. Summary statistics comparing first and second stage logistic regression models

Statistic	Logistic Regression Model Alone	Logistic Regression Model + Diagnostic Data
Sensitivity	0.23	0.30
Specificity	0.97	0.96
Positive Predictive Value	0.60	0.63
Negative Predictive Value	0.85	0.86
Area Under the Curve*	0.78	0.81

*Associated with receiver operating characteristic curve

Figure 4. A plot comparing the cumulative hourly probability of admission to the actual hourly admission for a) The triage administrative data alone; and b) The triage administrative data plus the diagnostic data



quantitative performance measures were also investigated including sensitivity, specificity, PPV, and NPV. In this study, the sensitivity and specificity of the model was also comparable to Sun et al. (2011), however, the PPV in this study was 20% lower and the NPV was 15% higher. These results indicate that the model performs better than Sun et al. (2011) at predicting which patients will not be admitted.

As an intervention to decrease ED length of stay, early resource planning for admitted patients may be one option to complement a set of strategies related to patient input, throughout, and output. Bed planners could use an electronic tracking system to follow the hourly predicted admissions and allocate open beds to those patients. Given that the average time between triage and admission is 216 minutes, this lead time would provide sufficient time to find a bed, prepare the room and schedule the necessary allied health professionals and staff (e.g., porters, housekeeping staff, and ward nurses).

In the TBRHSC, the hospital is often over-capacity in terms of admitted patients. On a daily basis the admission and discharge rates are approximately equal, otherwise the hospital would continue to fill with patients. If this system were used and the number of predicted admissions exceeded planned discharges, bed planners could initiate protocols that identified patients close to discharge and able to succeed at home to be discharged with homecare services. In addition, patients that require transfer to either a rehabilitation facility or repatriation to their home community could be expedited to provide additional beds.

Study Limitations

The logistic regression model could be improved if several other variables were available. This study used variables similar to those used by Peck et al. (2012), however, they also had chief complaint which were limited to a pre-set number of complaints. Triage vitals may improve the quality of prediction. Triage vitals are collected in the TBRHSC ED for many patients, however, in this data set approximately 30 to 40% of the data was missing.

CONCLUSION

This study has investigated the use of triage administrative data and diagnostic test data for the prediction of patient admission. The goal of this study was to predict which patients would be admitted so that bed planners could begin room preparation to reduce length of stay or boarding in the ED. The model performed reasonably well in predicting the average number of beds required, however, it was less accurate in identifying which patients would be admitted. If a system of predicted admission could be implemented in the ED, the amount of time that admitted patients spend in the ED could be reduced, thus increasing ED patient flow.

ACKNOWLEDGMENT

This study was funded by a Clinical Innovation Opportunities Fund (Grant No: C-14-2-18) through the Northern Ontario Academic Medicine Association. The authors would also like to thank Mr. Bruce Weaver for editing the manuscript.

REFERENCES

Bond, K., Ospina, M., & Blitz, S. (2006). *Interventions to reduce overcrowding in emergency departments*. Canadian Agency for Drugs and Technologies in Health.

Cameron, A., Ireland, A.J., & McKay, G.A. (2016). Predicting admission at triage: are nurses better than a simple objective score? *Emerg Med J*.

Caterino, J. M., Hoover, E. M., & Moseley, M. G. (2013). Effect of advanced age and vital signs on admission from an ED observation unit. *The American Journal of Emergency Medicine*, *31*(1), 1–7. doi:10.1016/j.ajem.2012.01.002 PMID:22386358

Chrusciel, J., Fontaine, X., Devillard, A., Cordonnier, A., Kanagaratnam, L., Laplanche, D., & Sanchez, S. (2019). Impact of the implementation of a fast-track on emergency department length of stay and quality of care indicators in the Champagne-Ardenne region: A before–after study. *BMJ Open*, *9*(6), 1–8. doi:10.1136/bmjopen-2018-026200 PMID:31221873

Considine, J., Kropman, M., Kelly, E., & Winter, C. (2008). Effect of emergency department fast track onemergency department length of stay: A case–control study. *Emergency Medicine Journal*, *25*(12), 815–819. doi:10.1136/emj.2008.057919 PMID:19033498

Derlet, R. W., & Richards, J. R. (2000). Overcrowding in the nation's emergency departments: Complex causes and disturbing effects. *Annals of Emergency Medicine*, *35*(1), 63–68. doi:10.1016/S0196-0644(00)70105-3 PMID:10613941

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. doi:10.1016/j.patrec.2005.10.010

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, *33*(1), 1–22. doi:10.18637/jss.v033.i01 PMID:20808728

Hong, W. S., Haimovich, A. D., & Taylor, R. A. (2018). Predicting hospital admission at emergency department triage using machine learning. *PLoS One*, *13*(7), 1–13. doi:10.1371/journal.pone.0201016 PMID:30028888

LaMantia, M. A., Platts-Mills, T. F., Biese, K., Khandelwal, C., Forbach, C., Cairns, C. B., Busby-Whitehead, J., & Kizer, J. S. (2010). Predicting Hospital Admission and Returns to the Emergency Department for Elderly Patients. *Academic Emergency Medicine*, *17*(3), 252–259. doi:10.1111/j.1553-2712.2009.00675.x PMID:20370757

Peck, J. S., Benneyan, J. C., & Nightingale, D. J. (2012). Predicting emergency department inpatient admissions to improve same-day patient flow. *Annals of Emergency Medicine*, *19*, 1045–1054. PMID:22978731

R Core Team. (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org/

Rowe, B.H., Bond, K., & Ospina, M.B. (2006). *Frequency, determinants, and impact of emergency department overcrowding in Canada.* Technology Report No. 67.3. Ottawa: Canadian Agency for Drugs and Technologies in Health.

Sing, T., Sander, O., & Beerenwinkel, N. (n.d.). *ROCR: Visualizing the performance of scoring classifiers. R package version 1.0-7*. http://rocr.bioinf.mpi-sb.mpg.de/

Sun, Y., Heng, B. H., & Tay, S. Y. (2011). Predicting hospital admission at emergency department triage using routine administrative data. *Annals of Emergency Medicine*, *18*, 844–850. PMID:21843220

Zhou, X.-H., Obuchowski, N. A., & McClish, D. K. (2011). Statistical Methods in Diagnostic Medicine. John Wiley and Sons Ltd. doi:10.1002/9780470906514

David Savage is an emergency physician in Thunder Bay and an Assistant Professor at the Northern Ontario School of Medicine.

Douglas Woolford is an Associate Professor of Environmetrics and Director of the Masters of Data Analytics program in the Department of Statistical and Actuarial Sciences at the University of Western Ontario.

Mackenzie Simpson is a Master's student in the Department of Computer Science at Lakehead University.

David Wood is an emergency physician, Medical Program Director, Trauma/Service Director, Emergency & Critical Care, and an Assistant Professor at the Northern Ontario School of Medicine.

Robert Ohle is an emergency physician at Health Sciences North in Sudbury and an Assistant Professor at the Northern Ontario School of Medicine.