Real-time prediction of inpatient length of stay for discharge prioritization

RECEIVED 2 February 2015
REVISED 18 May 2015
ACCEPTED 31 May 2015
PUBLISHED ONLINE FIRST 7 August 2015

Sean Barnes[1], Eric Hamrock[2], Matthew Toerper[3], Sauleh Siddiqui[4], Scott Levin[5]

ABSTRACT

Objective Hospitals are challenged to provide timely patient care while maintaining high resource utilization. This has prompted hospital initiatives to increase patient flow and minimize nonvalue added care time. Real-time demand capacity management (RTDC) is one such initiative whereby clinicians convene each morning to predict patients able to leave the same day and prioritize their remaining tasks for early discharge. Our objective is to automate and improve these discharge predictions by applying supervised machine learning methods to readily available health information.

Materials and Methods The authors use supervised machine learning methods to predict patients' likelihood of discharge by 2 p.m. and by midnight each day for an inpatient medical unit. Using data collected over 8000 patient stays and 20 000 patient days, the predictive performance of the model is compared to clinicians using sensitivity, specificity, Youden's Index (i.e., sensitivity + specificity – 1), and aggregate accuracy measures.

Results The model compared to clinician predictions demonstrated significantly higher sensitivity ($P < .01$), lower specificity ($P < .01$), and a comparable Youden Index ($P > .10$). Early discharges were less predictable than midnight discharges. The model was more accurate than clinicians in predicting the total number of daily discharges and capable of ranking patients closest to future discharge.

Conclusions There is potential to use readily available health information to predict daily patient discharges with accuracies comparable to clinician predictions. This approach may be used to automate and support daily RTDC predictions aimed at improving patient flow.

Keywords: length of stay, patient flow, machine learning, operational forecasting

RESEARCH AND APPLICATIONS

BACKGROUND AND SIGNIFICANCE

Hospitals and clinician objectives require balance in treating each patient's condition effectively while efficiently distributing healthcare resources to patient populations over time.[1] A key determinant of hospital capacity and resource management is linked to patient flow, a common indicator of patient safety, satisfaction, and access.[2–5] Optimal patient flow facilitates beneficial treatment, minimal waiting, minimal exposure to risks associated with hospitalization, and efficient use of resources (e.g., of beds, clinical staff, and medical equipment). Patient flow is also a determinant of access to specialized inpatient services. These patients request admission from external sources (e.g., other health services) and internal sources such as the emergency department (ED), procedural areas, or peri-anesthesia care units (PACUs).[6,7]

Evidence supporting the benefits of improved flow has been mounting.[8] For patients, prolonged hospital stays increase the risk of adverse events, such as hospital-acquired infections, adverse drug events, poor nutritional levels, and other complications.[9–12] Extended stays have also been associated with poor patient satisfaction.[11,13–17] For hospitals, economic pressures to deliver efficient and accessible care are at unprecedented highs. Healthcare costs as a percentage of gross domestic product (GDP) (17.9% in 2012) have been rising faster than anticipated,[18] and approximately 30–40% of these expenditures have been attributed to "overuse, underuse, misuse, duplication, system failures, unnecessary repetition, poor communication, and inefficiency."[19] These factors impede patient flow, prolong patient stays, and increase the cost of care per patient.[14,20–22]

Patient flow, and by association, bed and capacity management, is a common focus area for operations management methods applied to healthcare. Discrete-event simulation, optimization, and Lean Six Sigma approaches have been applied successfully in various settings to improve patient flow by either redesigning care delivery processes or more efficiently matching staff and other resources (e.g., operating rooms, medical equipment) to demand.[23–35] These patient flow evaluations are data-driven and inform long-term operational decision-making. Outcomes of these studies include improved patient and staff scheduling strategies, new bed management policies, and reduction in care process variability.

A more recent advancement in patient flow management alternatively focuses on short-term operational decisions. Real-time demand capacity management (RTDC) is a new method developed by the Institute for Healthcare Improvement that has shown promising but variable results when pilot tested in hospitals. The RTDC process involves 4 steps: 1) predicting capacity, 2) predicting demand, 3) developing a plan, and 4) evaluating the plan.[36] The RTDC process centers around a morning clinician huddle to predict which and how many patients will be discharged that same day. Given daily predictions for demand, the group then attempts to match their supply of beds to the demand from new patients (i.e., admissions) by prioritizing current patients able to be discharged. RTDC implementation signifies a culture change as the hospital staff dedicates time each morning to coordinate and focus on patient flow. The developers of RTDC have demonstrated that this approach, after an initial learning period, may reduce

Correspondence to Department of Decision, Operations & Information Technologies, Robert H. Smith School of Business, 4352 Van Munching Hall, University of Maryland, College Park, MD 20742, USA; sbarnes@rhsmith.umd.edu; Tel: (301) 405-9679

For numbered affiliations see end of article.

key indicators of patient flow across hospitals. These indicators include ED boarding time, patients who left without being seen, and overnight holds in the PACU.

RTDC does have limitations that we strive to address in this work. First, RTDC requires a daily clinician huddle that demands dedicated time that may be expedited or even eliminated through automated prediction. Second, the prediction process is subjective and thus susceptible to high variability. In addition, the most successful implementations of RTDC have occurred in surgical units, where patient conditions and clinical pathways are commonly well prescribed.[37,38] Thus, previous results on RTDC predictive performance for surgical patients may not be generalized to patients in medical units where clinical pathways are less defined and length of stay (LOS) is more variable.

Our objective for this study is to support automation of the RTDC process by producing daily predictions of patient discharge times for a single inpatient medical unit. We hypothesize that a predictive model using readily available health information may perform comparable or better than predictions made by clinicians during their daily huddle. We apply supervised machine learning methods to predict the probability of patient discharge by 2 p.m. and by the end of each day (i.e., midnight). By discharging a patient earlier in the day (e.g., prior to 2 p.m.), there is an increased likelihood of admitting a new patient to the same bed, thereby facilitating increased bed utilization and patient flow.[36]

Operational Forecasting Literature

Forecasting models have been leveraged to predict hospital occupancy, patient arrivals and discharges, and other unit-specific operational metrics. These models have been derived from variations of autoregressive moving average approaches,[39–44] exponential smoothing,[39,40,42] Poisson regression,[45] neural networks,[42] and discrete-event simulation methods.[46–48] More recent studies have employed logistic regression[49] and survival analysis[50] methods to predict patient LOS. Through accurate prediction of patients' LOS, clinical staff may efficiently schedule future appointments or admissions to avoid backlog or denials.[44,51–53] Our approach builds on this research by focusing on short-term predictions (i.e., daily discharges) for patients.

Supervised Machine Learning Literature

Tree-based supervised machine learning algorithms have been applied in healthcare to 1) predict a continuous-valued outcome or 2) classify patients into one or more clinical subgroups. Example applications for the former include using linear regression or regression trees to predict the range of motion for orthopedic patients,[54] costs,[55,56] and utilization.[57] An example of the latter typically involves logistic regression or classification trees, and has been used to differentiate between benign and malignant tumors,[58] identify patients most likely to benefit from screening procedures,[59,60] identify high risk patients,[61] and predict specific clinical outcomes.[62–64] Recently, more advanced machine learning techniques such as bagging, boosting, random forests, and support vector machines have been applied to healthcare problems such as classifying heart failure patients[65] and predicting healthcare costs.[66] We build upon this research by adding a new application area for these methods and leveraging the benefits of ensemble learning techniques such as bagging and boosting.

MATERIALS AND METHODS

Setting and Data Sources

The study was conducted within a single, 36-bed medical unit in a large, mid-Atlantic academic medical center serving an urban population. The unit is staffed with hospitalist physicians with no teaching responsibilities. Patient flow data (i.e., admission and discharge times), demographics, and basic admission diagnoses data was collected for 9636 patient visits over a 34-month study period from January 1, 2011 to November 1, 2013. After excluding incomplete and erroneous records, we retained data for 8852 patient visits and converted these visits to $N = 20\,243$ individual patient days. These data are summarized in Table 1.

These data are standardized and readily accessible in most hospital information systems, thus reproducible in other hospitals. The demographic and clinical predictors are static model inputs that are known at the time of admission and do not change during a patient's stay. Other predictors such as patient census, day of the week, and elapsed length of stay are dynamic and are continuously updated during a patient's stay. Elapsed length of stay, age, and patient census are numerical variables, whereas the remaining predictors are modeled using binary indicator variables (i.e., 0 or 1 indicating the absence or presence of a specific category).

The reason for visit is determined according to the International Classification of Diseases-9 diagnoses structure.[67] A physician identifies this condition (via an electronic pick list) at the time of admission to the unit, thereby documenting the primary reason for hospitalization. Observation status is assigned to patients who are cared for within the unit and must be monitored and evaluated before they are eligible for safe discharge.[68] Administratively, hospitals may use this designation to bill Medicare for the patient under the outpatient service category. However, a large study by the Department of Health and Human Services found that observation patients have the same health conditions as those who are fully admitted.[69]

In addition to the patient data listed in Table 1, a novel aspect of this study is our collection of clinician predictions of patients to be discharged at 2 p.m. and midnight. We recorded predictions from daily morning huddles for 8 overlapping months between March 18, 2013 and November 1, 2013. The prediction team was comprised of a charge nurse, case manager, and physicians. Members of this team were directly involved with the administrative or clinical aspects of a subset of patients in the unit each day. These team members had access to substantially more information for their predictions than what was accessible to our prediction models. This unique data facilitated a comparative study between the machine-learning and clinician predictions. In order to compare performance directly to the RTDC process, the machine-learning models were designed to produce predictions based on data available at 7 a.m. each day, which is analogous to the clinician huddle times.

Analytic Methods

We applied and evaluated several supervised machine learning algorithms to predict patient discharge and compare with clinician predictions. These algorithms are considered 'supervised' because they are fit to labeled training data (i.e., the outcomes are known in retrospect) and then independently evaluated on separate labeled test data to estimate their performance in practice. For each patient day, these algorithms produced two predictions representing the probability of a patient being discharged by either 2 p.m. or midnight that same day. These predictions were generated for every patient in the unit for each day of their stay, which simulated the clinician-based RTDC prediction process. This model is designed to support real-time predictions of expected bed capacity and could be adapted to predict patient discharges for any time interval (e.g., hourly instead of specifically 2 p.m. and the end of the day).

Systematic experiments applying common supervised machine-learning methods to predict individual patient discharges were performed. These methods included logistic regression (i.e., reference method), classification and regression trees, and tree-based ensemble

Table 1: Patient flow and prediction data.

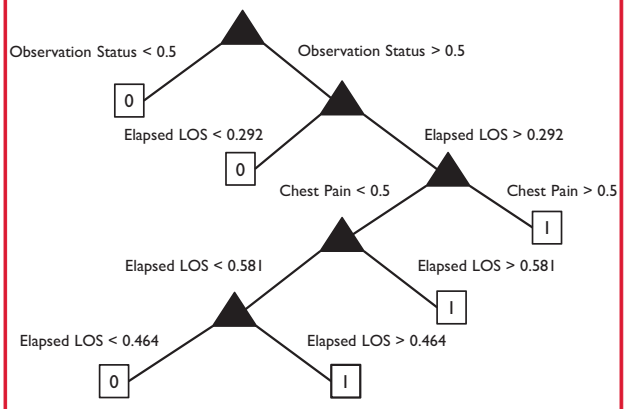
Patient Flow (Outcomes)	
Length of stay: mean, median (IQR)	Mean: 52 h, median: 37 h (42.40 h)
Discharge timing	27.4% of discharges by 2 p.m., 88.8% by 7 p.m.
Demographic predictors	
Gender	57.1% female
Ethnicity	68.9% Caucasian, 26.7% African American, 4.4% Other
Age: mean, median (IQR)	58.4 years, 57.0 years (24 years)
Insurance	45.3% Private, 39.5% Medicare, 9.4% Uninsured, 5.9% Medicaid
Clinical predictors	
Reason for visit	Chest pain 39.4%, syncope 11.4%, abdominal pain 3.8%, chronic obstructive pulmonary disorder (COPD) 3.6%, congestive heart failure 1.8%
Observation status	52.5% of patients
Pending discharge location	Home 85%, other healthcare facilities: 15%
Unit workload predictors	
Patient census: mean, median (IQR)	20.87, 21 (7)
Timing predictors	
Day of week	Monday 14.7%, Tuesday 15.4%, Wednesday 16.0%, Thursday 16.2%, Friday 16.8%, Saturday 11.4%, Sunday 9.5%
Elapsed length of stay	Changes dynamically

learning methods. Model parameters and thresholds were tuned for optimal performance with respect to the predictive measures described in the following section.[70] We compared the results among these predictive methods, and then selected the best method to compare to the clinician predictions.

Logistic regression is a commonly used classification method for clinical applications, and it is an effective approach for providing a baseline for how effectively the available data can be leveraged to predict the primary outcome. However, logistic regression is sometimes limited for predictive applications, especially with large, highly dimensional data where interactions may exist.[70] Despite this, logistic regression is a well-established method that will serve as a baseline for comparison with more robust supervised machine learning algorithms.

Next, we applied tree-based methods to predict patient discharge outcomes, which iteratively partition the data into groups with similar characteristics and outcomes.[71–73] These methods are adept at recognizing predictor variable interactions and identifying sub-cohorts of patients that are more likely to have a positive outcome.[71] Tree-based models have the distinct advantage of providing more practical insight than regression-based methods. The highest levels of these trees can be translated into effective decision rules that may be interpreted

Figure 1: Classification tree for end-of-day discharge predictions.



by clinicians, and embedded within existing clinical information technology infrastructure to facilitate implementation and uptake in practice.[68]

Figure 1 shows a visualization of the classification tree for end-of-day discharge predictions, which provides an example of how our results could be shared with clinicians. In this figure, we observe that observation status is the most critical predictor relative to predicting the outcome. Patients who are not on observation status (i.e., an intermittent care status to evaluate whether they need to be admitted) tend to stay, regardless of any of their other conditions. Patients who are on observation status are more likely to be discharged only when their elapsed LOS exceeds approximately 12 h or if they reported chest pain as their chief complaint.

Despite these benefits, simple tree-based learning methods have the potential to over-fit training data – causing the in-sample performance to exceed the out-of-sample predictive performance, which occurs when the model is overly complex and mistakes noise for key underlying relationships.[72,74] Tree-based ensemble learning methods, such as bagging and boosting, address over-fitting by training large numbers of "weak learners" and leveraging the diversity across learners (i.e., individual trees) to produce stable out-of-sample predictions. Diverse trees are aggregated in some form (e.g., by selecting majority votes or averaging) to classify or estimate risk for an outcome.[72,74–76] Bagging utilizes a bootstrapping process (i.e., sampling with replacement) to train numerous trees to produce model-averaged predictions.[74] The random forest is a common bagging method that increases diversity across each individual tree.[72,75,76] The random forest approach also facilitates evaluation of weak learners by selecting a random subset of predictors at each candidate split in the learning process for predictor variable importance in classifying outcomes. Boosting implements optimization algorithms to re-weight misclassified observations in an attempt to improve overall prediction accuracy.[72]

Training and Evaluation

Our ultimate goal was to implement the most parsimonious model with high accuracy that utilizes data that can be automatically extracted from electronic medical record systems. Predictor variables (see Table 1) for each patient day (collected at 7 a.m.) were linked to binary outcomes indicating whether the patient was discharged by 2 p.m. or the end of day, respectively. We trained each model on data collected over 26.6 months from the start of the study (January 1, 2011) until the date when the clinician predictions began (18 March, 2013). Model

predictions were generated for the following 9.4 months (19 March 2013 to 31 December, 2013) out-of-sample; purposefully overlapping with the same period clinician predictions were collected. This facilitated cross-validation (78% training set and 22% testing set) of the model and direct comparison to the clinician predictions. Predictive performance was then estimated for model-based and clinician predictions using binary classification statistics. Measures from the confusion matrix (i.e., true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN)) were used to calculate:

- Sensitivity $=$ TP / (TP $+$ FN)
- Specificity $=$ TN / (TN $+$ FP)
- Youden's Index $J =$ Sensitivity $+$ Specificity $-$ 1

These measures, along with the number of positive predictions (i.e., discharges), captured different aspects of model performance. Youden's Index[77] is a global accuracy measure whereas sensitivity and specificity capture how well the predictions perform with respect to a specific outcome (i.e., discharge or stay). Youden's Index is a common metric used to evaluate the performance of diagnostic tests,[78–80] and similar to these studies, we used it to optimize the cut-off thresholds (i.e., to determine a discharge or stay prediction) for each algorithm. We also calculated these metrics for near future outcomes (i.e., outcomes for the next time period). For example, how many patients predicted to be discharged by 2 p.m. were discharged by the end of the day? Similarly, how many patients predicted to be discharged by the end of the day were discharged by the end of the next day? These measures include correct predictions but also incorrect predictions that were correct within one day, which provides some insight as to the magnitude of predictive errors.

We conducted hypothesis testing ($\alpha = 0.05$) to detect statistically significant differences between the predictive capabilities for the model and the clinicians. Each of these tests can be applied readily to paired samples, which is required for comparing predictions for the same patient day(s) We used McNemar's test with Yates correction for continuity[81–82] to test the null hypothesis H_0: $p_{model} = p_{clinicians}$, where p represents the proportion of correct predictions relative to the total number of positive (sensitivity) or negative (specificity) outcomes. For Youden's Index J, we apply the method devised in Chen et al.[83] for paired samples to test the null hypothesis H_0: $J_{model} = J_{clinicians}$. For both tests, we used two-sided alternative hypotheses.

In addition, we evaluated the results in two additional measures that are relevant to model implementation. First, we aggregated (i.e., summed) the number of expected daily discharges from the model and clinician predictions, and then compared these results to the actual number of discharged patients for each day using paired hypothesis tests at the $\alpha = 0.05$ level. This aggregate measure is useful for proactive patient flow management by anticipating available (i.e., unoccupied) beds in advance and facilitating accept/reject admission decisions related to capacity constraints. Separately, we used model predictions to rank the patients in order of their likelihood for discharge each day. Spearman rank correlation coefficients were computed to evaluate the accuracy of these rankings with observed patients remaining length of stay.[84] This latter approach provided some indication as to whether this model could be used to accurately prioritize patients for discharge.

RESULTS

We applied tree-based supervised machine learning methods to predict discharge by 2 p.m. and the end of the day for each patient day. The regression random forest (RRF) method proved most accurate from systematic experiments of several tree-based algorithms with parameters tuned for optimal predictive performance. Figure 2, summarizing variable importance from the training of the RRF models, provides evidence that elapsed LOS and observation status are substantially more important than all of the other predictors, followed by Sunday, chest pain, disposition, age, and syncope. Predictors such as gender, ethnicity, weekdays (Monday through Thursday), and less common reasons for visit (e.g., abdominal pain, COPD, congestive heart failure) had little to no predictive power for this patient population.

Individual Predictions

We compared the RRF model predictions to the clinician predictions in order to evaluate the potential of this approach in practice. We have also included comparisons to logistic regression as a reference to better understand the efficacy of the ensemble learning approach. Results for both 2 p.m. and end-of-day outcomes are summarized in Table 2 for the specific days when clinician predictions were available (19 March, 2013 to 31 December, 2013), which included 4833 patient days.

Overall, the logistic regression and RRF model were more aggressive in predicting discharge than the clinicians for both the 2 p.m. and end-of-day outcomes. This resulted in the automated models predicting discharges with higher sensitivity ($P < .01$) and lower specificity ($P < .01$) compared to the clinicians. Thus the models predicted a higher proportion of discharges, but at the cost of producing more false positives. The logistic regression baseline model consistently demonstrated more sensitive behavior than the RRF model.

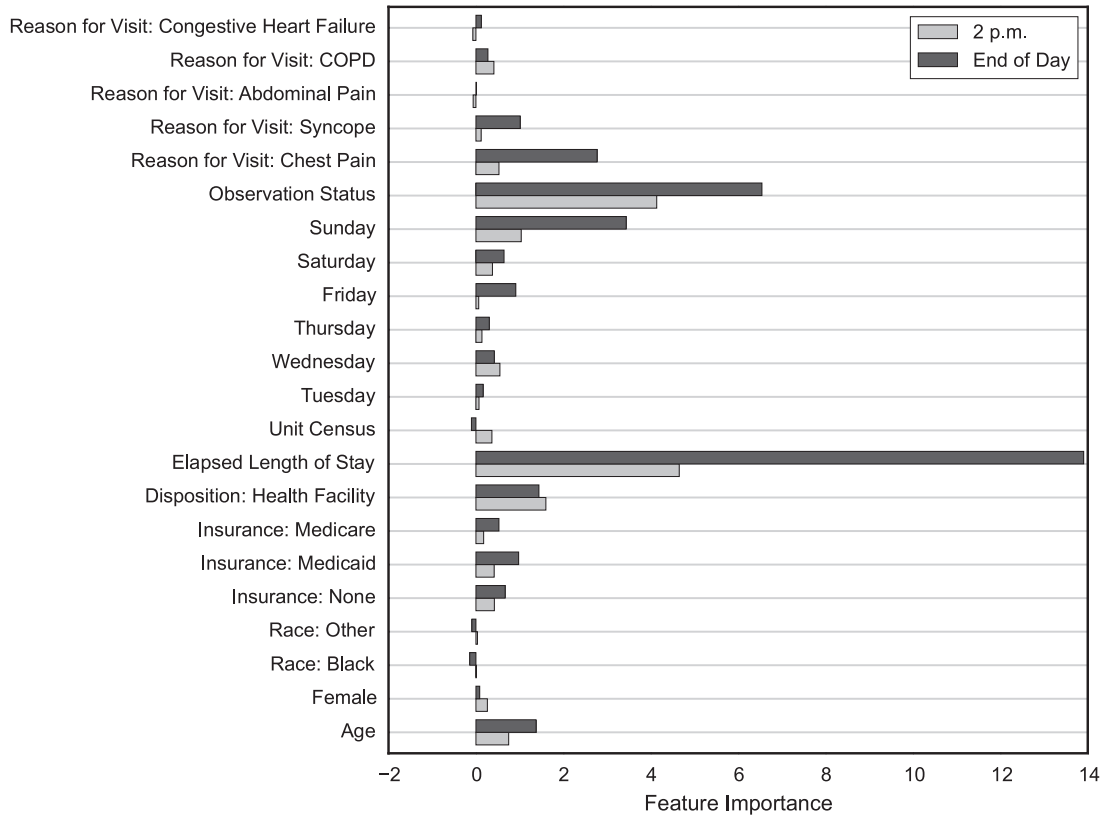
Despite differences in sensitivity and specificity, the RRF model and clinicians predictive performance were comparable for Youden's Index, our global accuracy measure. There were no significant differences for same-day predictions (2 p.m. RRF model: 26.0%, clinicians 26.5%, $P = .81$; end-of-day RRF model: 34.0%, clinicians 34.0%, $P = .84$) or the near-future end-of-day outcome (RRF model: 26.6%, clinicians 25.2%, $P = .62$). However, the RRF model did perform significantly better for the near-future 2 p.m. outcome (RRF model: 31.3%, clinicians 19.0%, $P < .01$). Predictions across models and clinicians were more accurate for the end of the day than for 2 p.m.

Daily Aggregate Predictions

We also compared the aggregated RRF model and clinician predictions for the total number of discharges each day. For the model predictions, we summed the predicted probabilities across all patients to calculate the expected value of discharged patients. The results for 2 p.m. and the end of day for the overlapping date range are summarized in Figures 3 and 4, which show that the model outperforms clinician estimates of the average number of patients to be discharged early and by the end of the day.

In Figure 3, we plot the distribution of the actual number of discharges per day against the distributions for the RRF model and clinician predictions for each outcome. The actual number of discharges per day averaged 2.36 by 2 p.m. and 8.29 patients by the end of the day. Discharges predicted each day by the RRF model were 2.45 ($P = .37$) and 8.51 ($P = .13$), respectively, demonstrating no statistically significant difference from the actual number of discharges. In comparison, clinician predictions were accurate for 2 p.m. discharges (2.16, $P = .19$), but significantly deviated from the actual number of discharges for the end of the day (6.54, $P < .01$). The model predicted the total number of discharges within 2 patients 82% of days for 2 p.m. and 105 63% of days for the end of the day, whereas the

Figure 2: Variable importance summary from training regression random forest models for the 2 p.m. and end-of-day discharge predictions.



clinicians predicted the same for 52% and 32% of days, respectively (see Figure 4). In addition, large residuals (i.e., errors of 5 discharges or more) were much more frequent for the clinician predictions.

Daily Rank Predictions

In practice, our prediction model could be used to rank patients daily—based on their likelihood of being discharged—in order to prioritize the remaining tasks for the most likely patients. We computed Spearman's rank correlation between the remaining LOS for each patient and their respective RRF model predicted probabilities for each day (trained on the full data set). Figure 5 shows a histogram of these correlations (in intervals of 0.05) for the 2 p.m. and end-of-day models. The mean rank correlation for 2 p.m. was 0.4816 and for the end of the day was 0.4489. Both plots show that the correlations are almost exclusively positive and moderately large, which suggests that the rank of the model predictions was moderately correlated with the actual discharge order. For some days, the model nearly predicted the exact order in which the patients were discharged.

DISCUSSION

Improving patient flow continues to be a top priority in the acute-care setting, where patients with longer lengths of stay are less satisfied and exposed to the risk of adverse events (e.g., hospital-acquired infections, complications). Hospitals have aligned incentives to improve patient flow because of the rising demand for services and the economic pressures to reduce costs and improve resource utilization. Our approach is designed to empower clinicians and hospital administrators with analytical tools to increase their collective efficiency.

Our approach could be operationalized in three distinct ways. First, it could be used to identify individual patients who are most likely to be discharged on a given day. Hospital staff could prioritize these patients in order to discharge them as early as possible—without negatively affecting their care—so that other patients can be admitted in their place. Supervised machine learning methods may be used to rank patients concurrently in a hospital (or specific unit) according to their discharge probabilities. We have shown that our models perform well for prediction or ranking. Alternatively, patients who are most likely to be discharged may not be impacted significantly by an effort to prioritize their remaining tasks. Therefore, another approach could be to identify a second tier of patients who are only mildly to moderately likely to be discharged. Prioritizing the remaining tasks for these patients may have a more significant global impact on the number of patients discharged over a given time period. Setting a range of predicted probabilities immediately below the classification threshold would identify these patients. The last approach is to aggregate the predicted discharge probabilities into daily discharge predictions. This approach does not facilitate prioritization for patients likely to be discharged, but it supports bed capacity planning for the unit. We have shown that these types of predictions are very accurate and would provide the staff with a good idea of how many in-use beds are likely to be available by a specific time of day.

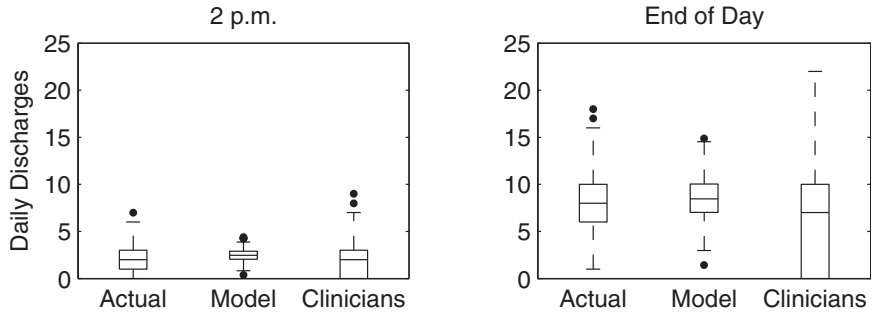
There are three important limitations to highlight for this study. First, the performance of the models may improve with a

Table 2: Performance comparison summary between logistic regression, regression random forest, and clinician predictions performance comparison.

Performance Measure	2 p.m.			End of Day		
	Logistic Regression	Regression Random Forest	Clinicians	Logistic Regression	Regression Random Forest	Clinicians
Positive Predictions (Discharge)	1749	1781	485	1870	2175	1471
Sensitivity (%)	65.9	60.0	33.6	71.5	66.1	51.3
	$P < .01$	$P < .01$		$P = .02$	$P < .01$	
Specificity (%)	52.8	66.0	92.9	54.9	68.3	82.7
	$P < .01$	$P < .01$		$P < .01$	$P < .01$	
Youden's Index (%)	18.7	26.0	26.5	26.4	34.0	34.0
	$P < .01$	$P = .81$		$P < .01$	$P = .84$	
Near Future Sensitivity (%)	73.4	56.1	21.7	81.9	54.1	39.0
	$P < .01$	$P < .01$		$P < .01$	$P < .01$	
Near Future Specificity (%)	51.5	75.2	97.3	49.3	72.5	86.1
	$P < .01$	$P < .01$		$P < .01$	$P < .01$	
Near Future Youden Index (%)	24.9	31.3	19.0	31.2	26.6	25.2
	$P = .06$	$P < .01$		$P = .02$	$P = .62$	

The *P*-value listed beneath each performance measure represents the results of McNemar's test for sensitivity and specificity or the method described in[83] for Youden's index for the estimated difference between each model and the clinicians. In each case, the null hypothesis of equality is tested against the two-sided alternative.

Figure 3: Comparison of actual, model prediction, and clinician prediction distributions of the average number of patients discharged from the unit each day.



larger data set, in terms of the number of patient days used to train the models and also with respect to being collected from multiple sites. Increasing the size of the training data set would improve our ability to detect more complex patterns in patient length of stay. Similarly, the patterns detected in our training data set may not be generalizable to other hospitals or units. Our intention is for each hospital and/or unit to replicate our prediction model and train it on its own data. Finally, we compared the performance of continuous (i.e., probability-based) predictions from our models to binary predictions (i.e., exit or stay) from the clinicians. Ideally, the fairest comparison would be between continuous predictions from both methods, however these predictions would be difficult to generate and collect in practice.

CONCLUSIONS

We applied supervised machine learning algorithms to readily available health information to predict daily discharge outcomes as part of the RTDC process. We directly compared model predictions to clinician predictions using several performance metrics. The model predicted discharges with higher sensitivity and lower specificity compared to the clinicians, and the two methods were comparable (i.e., not statistically significantly different) for our global accuracy measure (Youden's Index). However, the model did outperform the clinicians for some near-future and aggregate prediction metrics. Thus there is high potential for these models to automate and expedite the RTDC prediction process, thereby eliminating the need for daily clinician huddles or supporting more accurate clinician predictions. Furthermore, these

Figure 4: Histogram summary of differences between model, clinician, and actual daily discharges.

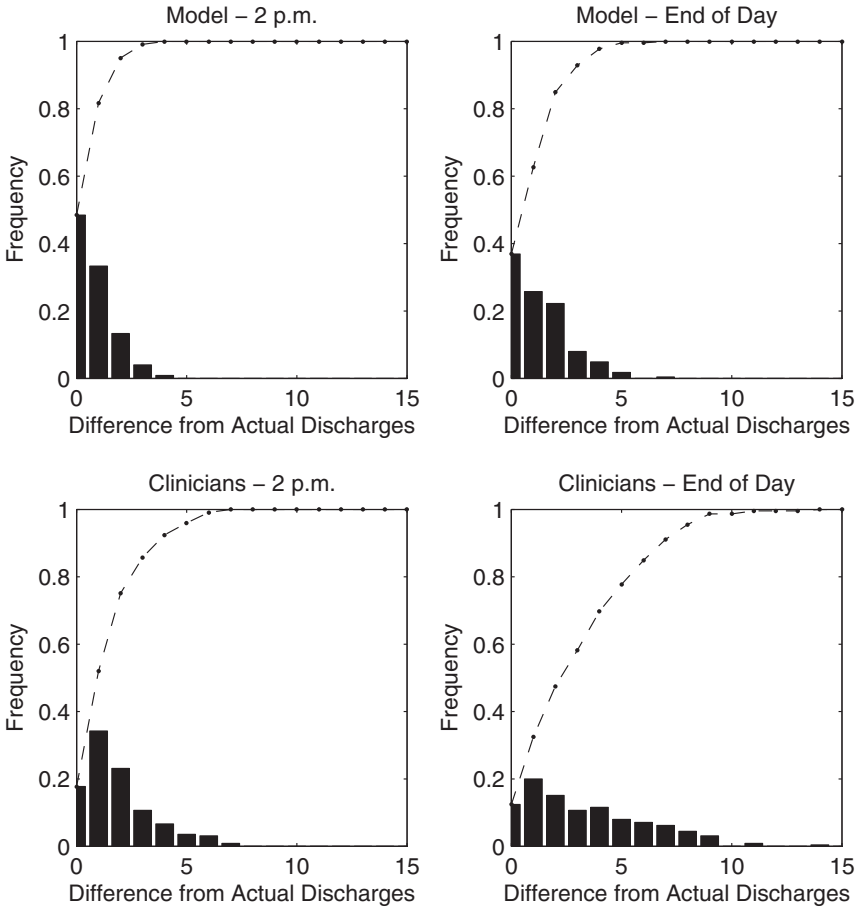
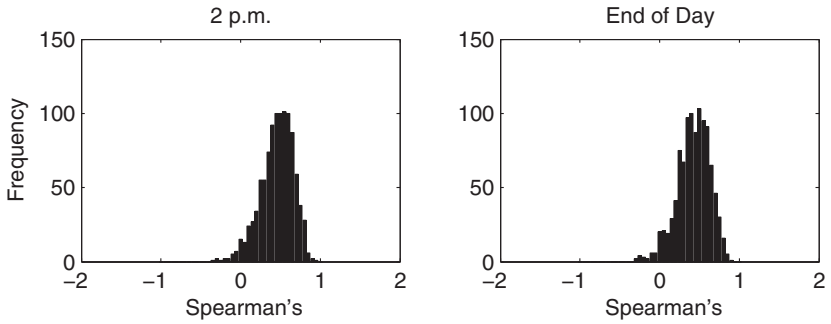


Figure 5: Histogram summary of daily Spearman's rank correlation between remaining LOS and model prediction scores for 2 p.m. and the end of the day.



models were applied to simple and readily accessible information that can be quite powerful, and easily replicated across acute-care environments using electronic information systems.

FUNDING

This work was supported in part by the National Science Foundation grant number 0927207.

CONTRIBUTORS

S.B., S.L., and E.H. made significant contributions to the conception and design of the work. M.T., E.H., and S.L. made significant contributions to the acquisition and processing of the data used for the analysis. S.B., S.L., and S.S. made significant contributions to the analysis and interpretation of the data for the work. All authors have contributed to either drafting or revising the article, have approved the version

to be published, and agree to be accountable for all aspects of the work.

REFERENCES

1. Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academies Press; 2001.
2. Bernstein, SL, Aronsky D, Duseja R, *et al*. The effect of emergency department crowding on clinically oriented outcomes. *Acad Emerg Med*. 2000;16(1):1–10.
3. Levin S, Dittus R, Aronsky D, *et al*. Evaluating the effects of increasing surgical volume on emergency department patient access. *BMJ Qual Saf*. 2011;20:146–152.
4. Ryckman FC, Yelton PA, Anneken AM, *et al*. Redesigning intensive care unit flow using variability management to improve access and safety. *Jt Comm J Qual Patient Safety*. 2009;35:535–543.
5. Compton DW, Ganjiang G, Reid PP, *et al. Building a Better Delivery System: A New Engineering/Health Care Partnership*. Washington, DC: National Academies Press; 2005.
6. Institute of Medicine. *Building a Better Delivery System: A New Engineering/Health Care Partnership*. Washington, DC: National Academies Press; 2006.
7. Faculty Practice Solutions Center. https://www.facultypractice.org/. Accessed July 31, 2014.
8. Institute for Healthcare Improvement. *Optimizing Patient Flow: Moving Patients Smoothly Through Acute Care Settings. IHI Innovation Series white paper*. Boston, MA: Institute for Healthcare Improvement; 2003.
9. Nguyen, HB, Rivers EP, Havstad S, *et al*. Critical care in the emergency department: a physiologic assessment and outcome evaluation. *Acad Emerg Med*. 2000;7(12):1354–1361.
10. Chalfin DB, Trzeciak S, Likourezos S, *et al*. Impact of delayed transfer of critically ill patients from the emergency department to the intensive care unit. *Crit Care Med*. 2007;35(6):1477–1483.
11. Lucas R, Farley H, Twanmoh H, *et al*. Emergency department patient flow: the influence of hospital census variables on emergency department length of stay. *Acad Emerg Med*. 2009;16(7):597–602.
12. Caccialanza R., Klersy C, Cereda E, *et al*. Nutritional parameters associated with prolonged hospital stay among ambulatory adult patients. *Can Med Assoc J*. 2010;182(17):1843–1849.
13. McGuire F. Using simulation to reduce length of stay in emergency departments. *J Soc Health Sys*. 1997;5(3):81–90.
14. Kainzinger F, Raible C, Pietrek K, *et al*. Optimization of hospital stay through length-of-stay-oriented case management: an empirical study. *J Public Health*. 2009;17(6):395–400.
15. Wiler, JL, Gentle G, Halfpenny JM, *et al*. Optimizing emergency department front-end operations. *Ann Emerg Med*. 2010;55(2):142–160.
16. Yankovic, N, Green LV. Identifying good nursing levels: a queuing approach. *Oper Res. 2011*;59(4):942–955.
17. Carden R, Eggers J, Wrye C, *et al*. The relationship between inpatient length of stay and HCAHPS Scores. *HealthStream Discover Paper*. 2013.
18. Hall R., Belson D, Murali P, *et al*. Modeling patient flows through the healthcare system. In: *Patient Flow: Reducing Delay in Healthcare Delivery*. New York, NY: Springer; 2006:1–44.
19. Reid PP, Compton WD, Grossman JH, *et al*., eds. *Building a Better Delivery System: A New Engineering/Health Care Partnership*. Washington, DC: National Academies Press; 2006.
20. Lagoe RJ, Westert GP, Kendrick K, *et al*. Managing hospital length of stay reduction: a multihospital approach. *Health Care Manage Rev*. 2005;30(2):82–92.
21. Thungjaroenkul P, Cummings GG, Embleton A. The impact of nurse staffing on hospital costs and patient length of stay: a systematic review. *Nurs Econ*. 2007;25(5):255.
22. Chan CW, Farias VF, Bambos N, *et al*. Optimizing intensive care unit discharge decisions with patient readmissions. *Oper Res*. 2012;60(6):1323–1341.
23. Dittus RS, Klein RW, DeBrota DJ, *et al*. Medical resident work schedules: design and evaluation by simulation modeling. *Manage Sci*. 1996;42(6):891–906.
24. Bagust A, Place M, Posnett JW. Dynamics of bed use in accommodating emergency admissions: stochastic simulation model. *BMJ*. 1999;319 (7203):155–158.
25. Jun JB, Jacobson SH, Swisher JR. Applications of discrete-event simulation in health care clinics: a survey. *J Oper Res Soc*. 1999;50:109–123.
26. Azaiez MN, Al Sharif SS. A 0-1 goal programming model for nurse scheduling. *ComputOper Res*. 2005;32(3):491–507.
27. Fairbanks CB. Using six sigma and lean methodologies to improve OR throughput. *AORN J*. 2007;86(1):73–82.
28. Burke EK, Li J, Qu R. A hybrid model of integer programming and variable neighbourhood search for highly-constrained nurse rostering problems. *Eur J Oper Res*. 2010;203(2):484–493.
29. Paul SA, Reddy MC, DeFlitch CJ. A systematic review of simulation studies investigating emergency department overcrowding. *Simulation*. 2010;86(8-9):559–571.
30. Fischman D. Applying lean six sigma methodologies to improve efficiency, timeliness of care, and quality of care in an internal medicine residency clinic. *Qual Manage Health Care*. 2010;19(3):201–210.
31. Chow VS, Puterman ML, Salehirad N, *et al*. Reducing surgical ward congestion through improved surgical scheduling and uncapacitated simulation. *Prod Oper Manag*. 2011;20(3):418–430.
32. Cima RR., Brown MJ, Hebl JR, *et al*. Use of lean and six sigma methodology to improve operating room efficiency in a high-volume tertiary-care academic medical center. *J Am Coll Surgeons*. 2011;213(1):83–92.
33. Zhu Z. Impact of different discharge patterns on bed occupancy rate and bed waiting time: a simulation approach. *J Med Eng Technol*. 2011;35(6-7):338–343.
34. Hamrock E, Paige K, Parks J, *et al*. Discrete event simulation for healthcare organizations: a tool for decision making. *J Healthc Manag*. 2013;58(2):110–124.
35. Schmidt R, Geisler S, Spreckelsen C. Decision support for hospital bed management using adaptable individual length of stay estimations and shared resources. *BMC Med Inform Decis Mak*. 2013;13(3):1–19.
36. Resar R, Nolan K, Kaczynski D, *et al*. Using real-time demand capacity management to improve hospitalwide patient flow. *Jt Comm J Qual Patient Saf*. 2011;37(5):217–227.
37. Ronellenfitsch U, Rössner E, Jakob J, *et al*. Clinical Pathways in surgery—should we introduce them into clinical routine? A review article. *Langenbeck's Arch Surg*. 2008;393(4):449–457.
38. Müller MK, Dedes KJ, Dindo D, *et al*. Impact of clinical pathways in surgery. *Langenbeck's Arch Surg*. 2009;394(1):31–39.
39. Lin WT. Modeling and forecasting hospital patient movements: univariate and multiple time series approaches. *Int J Forecast. 1989*;5(2):195–208.
40. Farmer RD, Emami J. Models for forecasting hospital bed requirements in the acute sector. *J Epidemiol Commun Health. 1990*;44(4):307–312.
41. Jones SA, Joy MP, Pearson J. Forecasting demand of emergency care. *Health Care Manag Sci*. 2002;5(4):297–305.
42. Jones SS, Thomas A, Evans RS, *et al*. Forecasting daily patient volumes in the emergency department. *Acad Emerg Med*. 2008;15(2):159–170.
43. Channouf N, L'Ecuyer P, Ingolfsson A, *et al*. The application of forecasting techniques to modeling emergency medical system calls in Calgary, Alberta. *Health Care Manag Sci*. 2007;10(1):25–45.
44. Littig SJ, Isken MW. Short term hospital occupancy prediction. Health Care Manag Sci. 2007;10(1):47–66.
45. McCarthy ML, Zeger SL, Ding R, *et al*. The challenge of predicting demand for emergency department services. *Acad Emerg Med*. 2008;15:337–346.

46. Hoot NR, LeBlanc LJ, Jones I, *et al*. Forecasting emergency department crowding: a discrete event simulation. *Ann Emerg Med*. 2008;52(2):116–125.
47. Hoot NR, LeBlanc LJ, Jones I, *et al*. Forecasting emergency department crowding: a prospective, real-time evaluation. JAMIA. 2009;16(3):338–345.
48. Hoot NR, Zhou C, Jones I, *et al*. Measuring and forecasting emergency department crowding in real time. *Ann Emerg Med. 2007*;49(6):747–755.
49. Levin SR, Harley ET, Fackler JC, *et al*. Real-time forecasting of pediatric intensive care unit length of stay using computerized provider orders. *Crit Care Med. 2012*;40(11):3058–3064.
50. Clark DE, Ryan LM. Concurrent prediction of hospital mortality and length of stay from risk factors on admission. *Health Services Res. 2002*;37(3):631–645.
51. Peck JS, Benneyan JC, Nightingale DJ, *et al*. Predicting emergency department inpatient admissions to improve same-day patient flow. *Acad Emerg Med*. 2012;19(9):E1045–1054.
52. Sun Y, Heng BH, Tay SY, *et al*. Predicting hospital admissions at emergency department triage using routine administrative data. *Acad Emerg Med*. 2011;18(8):844–850.
53. Peck JS, Gaehde SA, Nightingale DJ, *et al*. Generalizability of a simple approach for predicting hospital admission from an emergency department. *Acad Emerg Med*. 2013;20(11):1156–1163.
54. Ritter MA, Harty LD, Davis KE, *et al*. Predicting range of motion after total knee arthroplasty clustering, log-linear regression, and regression tree analysis. *J Bone Jt Surg. 2003*;85(7):1278–1285.
55. Raebel MA, Malone DC, Conner DA, *et al*. Health services use and health care costs of obese and nonobese individuals. *Arch Int Med. 2004*;164(19):2135–2140.
56. Gregori D, Petrinco M, Bo S, *et al*. Regression models for analyzing costs and their determinants in health care: an introductory review. *Intern J Qual Health Care*. 2011;23(3):331–341.
57. de Boer AG, Wijker W, de Haes HC. Predictors of health care utilization in the chronically ill: a review of the literature. *Health Policy. 1997*;42(2):101–115.
58. Sauerbrei W, Madjar H, Prompeler H. Differentiation of benign and malignant breast tumors by logistic regression and a classification tree using Doppler flow signals. *Methods Inform Med*. 1998;37(3):226–234.
59. Barriga KJ, Hamman RF, Hoag S, *et al*. Population screening for glucose intolerant subjects using decision tree analyses. *Diabet Res Clin Pract*. 1996;34 (Suppl.):S17–S29.
60. McGrath JS, Ponich TP, Gregor JC. Screening for colorectal cancer: the cost to find an advanced adenoma. *Am J Gastroenterol*. 2002;97:2902–2907.
61. Goldman L, Cook EF, Johnson PA, *et al*. Prediction of the need for intensive care in patients who come to the emergency departments with acute chest pain. *New Engl J Med*. 1996;334:1498–1504.
62. Falconer JA, Naughton BJ, Dunlop DD, *et al*. Predicting stroke inpatient rehabilitation outcome using a classification tree approach. *Arch Phys Med Rehabil. 1994*;75(6):619–625.
63. Germanson T, Lanzino G, Kassell NF. CART for prediction of function after head trauma. *J Neurosurg*. 1995;83:941–942.
64. Temkin NR, Holubkov R, Machamer JE, Winn HR, Dikmen SS, *et al*. Classification and regression trees (CART) for prediction of function at 1 year following head trauma. *J Neurosurg*. 1995;82(5):764–771.
65. Austin P, Tu J, Levy D. Using methods from the data-mining and machine-learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes. *J Clin Epidemiol*. 2013;66:398–407.
66. Robinson JW. Regression tree boosting to adjust health care cost predictions for diagnostic mix. *Health Services Res. 2008*;43(2):755–772.
67. Centers for Medicare and Medicaid Services. *Medicare Claims Processing Manual: Chapter 23 – Fee Schedule Administration and Coding Requirements*. 2015. http://www.cms.gov/Regulations-and-Guidance/Guidance/Manuals/downloads/clm104c23.pdf Accessed April 8, 2015.
68. Centers for Medicare and Medicaid Services. *Are You a Hospital Inpatient or Outpatient? CMS Product No. 11435*. 2014. https://www.medicare.gov/Pubs/pdf/11435.pdf Accessed April 8, 2015.
69. Wright, S. *Memorandum Report: Hospitals' Use of Observation Status and Short Inpatient Stays for Medicare Beneficiaries, OEI-02-12-00040*. 2013. https://kaiserhealthnews.files.wordpress.com/2013/07/oei-02-12-00040.pdf Accessed April 8, 2015.
70. Perlich C, Provost F, Simonoff JS. Tree induction vs. logistic regression: a learning-curve analysis. *J Mach Learn Res*. 2003;4:211–255.
71. Breiman L, Friedman J, Olshen R, *et al. Classification and Regression Trees, 1st edn*. Monterey, CA: Chapman and Hall; 1984.
72. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY: Springer; 2001.
73. Barner M. *Principles of Data Mining*. London: Springer; 2007.
74. Breiman L. Bagging predictors. *Mach Learn*. 1996;24(2):123–140.
75. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32.
76. Schapire R. The strength of weak learnability. *Mach Learn*. 1990;5(2):197–227.
77. Youden, WJ. Index for rating diagnostic tests. *Cancer*. 1950;3:32–35.
78. Greiner M, Sohr D, Göbel P. A modified ROC analysis for the selection of cut-off values and the definition of intermediate results of serodiagnostic tests. *J Immunol Methods*. 1995;185(1):123–132.
79. Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its associated cutoff point. *Biometrical J*. 2005;47(4):458–472.
80. Schisterman EF, Perkins NJ, Liu A, Bondell H. Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. *Epidemiology*. 2005;16(1):73–81.
81. McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*. 1947;12(2):153–157.
82. Yates F. Contingency tables. *J Royal Stat Soc*. 1934;1:217–235.
83. Chen F, Xue Y, Tan MT, Chen P. Efficient statistical tests to compare Youden index: accounting for contingency correlation. *Stat Med*. 2015;34:1560–1576.
84. Spearman C. The proof and measurement of association between two things. *Am J Psychol*. 1904;15:72–101.

AUTHOR AFFILIATIONS

[1]Department of Decision, Operations & Information Technologies, Robert H. Smith School of Business, 4352 Van Munching Hall, University of Maryland, College Park, MD 20742, USA
[2]Department of Operations Integration, Johns Hopkins Health System, Baltimore, MD, USA
[3]Department of Emergency Medicine, Johns Hopkins Hospital, Baltimore, MD, USA
[4]Departments of Civil Engineering and Applied Mathematics & Statistics, Johns Hopkins Systems Institute, Johns Hopkins University, Baltimore, MD, USA
[5]Department of Emergency Medicine and Civil Engineering, Johns Hopkins Systems Institute, Johns Hopkins University, Baltimore, MD, USA