Characterizing the value of predictive analytics in facilitating hospital patient flow

JORDAN S. PECK1,∗, JAMES C. BENNEYAN2, DEBORAH J. NIGHTINGALE3
and STEPHAN A. GAEHDE4

1VA Boston Healthcare System, Veterans Engineering Resource Center, 150 S Huntington Ave, Building 9 - VERC, Boston, MA 02151, USA
E-mail: peck.jordan@gmail.com
2Northeastern University, Healthcare Systems Engineering Institute, 334 Snell Engineering Center, Boston, MA 02115, USA
3Massachusetts Institute of Technology, Engineering Systems Division, 77 Massachusetts Ave, Building E38, Cambridge, MA 02139-4307, USA
4VA Boston Healthcare System, Emergency Services, 150 S Huntington Ave, Boston, MA 02151, USA

Received January 2014 and accepted May 2014

We apply discrete event simulation to characterize the patient flow affects of using admission predictions and current state information, generated in an Emergency Department (ED), to influence the prioritization of inpatient unit (IU) physicians between treating and discharging IU patients. Shared information includes crowding levels and total expected bed need (based on the sum of individual patients’ imperfect admission predictions and perfect admission predictions). It is found that sharing prediction and crowding information to influence inpatient staff priorities, using specific information sensitivity schedules, can result in statistically significant (p ≪ 0.05) reductions in boarding time (between 11.69% and 18.38% compared to baseline performance). The range of improvement is dependent on varying simulated hospital configurations.

Keywords: Patient flow, emergency department crowding, inpatient unit, simulation

1. Introduction

Prediction continues to grow as a recommended tool for enabling effective and efficient healthcare. The Institute of Medicine (2001) includes the “anticipation of needs” as one of the “new rules” for redesigning and improving care. A significant part of anticipation is predicting. The use of prediction to improve healthcare delivery is not new. Meehl (1954) employed statistical prediction tools for making diagnoses. There is also a long history of using prediction to improve operational decision making. Robinson et al. (1966) used the expert opinions of doctors and nurses about the length of stay of patients in order encourage staff to focus on patients who are expected to be discharged and thereby reduce artificial variability in patient length of stay. Gustafson (1968) predicted the length of stay of patients in a hospital, to assist future planning of elective admissions and resource scheduling, using five different methodologies for prediction: expert opinion, multiple linear regression, Bayesian conditional probability, historical means, and direct posterior odds estimation. These are just early examples of a large body of literature focused on the application of numerous prediction methodologies to equally numerous health care operational issues.

Emergency Departments (ED) are often studied using predictive analytics. In EDs, many metrics of quality are defined by how quickly a patient gets to and through required treatment (Graff et al., 2002, Bernstein et al., 2009, Horwitz et al., 2010). When a patient is ready to be admitted to the hospital Inpatient Unit (IU), ED and hospital staff will begin to coordinate resources to find an inpatient bed. This process can take hours, preventing another patient from seizing a bed in the ED. This process, known as boarding, is therefore considered one of the greatest causes of reduced quality in the ED (US GAO, 2003, Olshaker and Rathelev, 2006, Falvo et al., 2007, US GAO, 2009).

It has been suggested that flow through the ED/IU system can be improved by predicting whether a patient will need admission earlier in the ED treatment process, then the prediction can be shared with the IU. This information sharing could enable the bed coordination process earlier (Yen and Gorelick, 2007). This suggestion involves making predictions on patients, as they arrive, based on the endogenous variables associated with each individual patient.

∗Corresponding author
Some attempts at individual prediction have had clinical objectives. These studies focused on predicting whether specific categories of patients will be admitted: Neural network for children presenting with bronchiolitis (Walsh et al., 2004), Expert opinion on patients with acute coronary symptoms (Arslanian-Engoren, 2004), Expert opinion on patients arriving by ambulance (Levine et al., 2006, Clesham et al., 2008).

Some studies have focused on the operational benefits to making predictions and focused on the entire ED population. These studies used the following methods: Bayesian network (Leegeon et al., 2005), support vector machines, naive Bayes (Li et al., 2009), and logistic regression (Sun et al., 2011). The studies are valuable for developing prediction models; however they fall short from describing how the models could be used in a practical manner to improve flow. In each of the studies, there is a reliance on the historic conclusion that predictions should be useful, but they do not explore how the models can be applied. Instead, they assume that by providing a yes/no prediction, they simply would pre-empt or supplement an ED provider’s decision.

Recently, Peck et al. (2012) suggested that how the prediction is operationalized influences how the prediction data is shared. They suggested that rather than assign a yes/no to a patient; it is possible to increase accuracy by generating a real time total bed demand measure. This measure was derived by summing the admission probabilities of all patients in the ED. The study explored three methods for making these predictions, expert opinion, naive Bayes, and logistic regression. Logistic regression using a patient’s age, primary complaint, designation (ED for Fast Track) and route them to the IU, but is made now in order to generate a real time total bed demand measure. This measure (Peck et al. 2012).

2. Simulation model

Simulation, particularly discrete event simulation (DES), is a popular tool for exploring “what if” scenarios in the ED/IU system (Baesler et al., 2003, Connelly and Bair, 2004, Jacobson et al., 2006, Kolb et al., 2008, Li and Howard, 2010, Paul et al., 2010, Peck and Kim, 2010). The DES model in this study was built with ARENA DES software, version 13.5. The model is based on the ED/IU flow at the Veterans Health Administration (VHA) Boston Healthcare System (BHS). VHA BHS has a 13-bed ED that received approximately 1200 Veteran and non-Veteran patients per month in 2010. The hospital’s IU has a varying capacity of approximately 170-180 staffed beds reserved for Veterans who arrive through ED admissions, elective admissions, and transfers from VHA as well as non-VHA hospitals.

The logic of the model is shown in Figure 1 and is composed of four primary sub-models: Arrival, Emergency Department, Inpatient Unit and Bed Management.

2.1. Arrival sub-model

Simulated patients enter based on the arrival pattern derived from VHA BHS data (N = 6961). After a patient is created, they are assigned an admission probability based on a beta distribution with the following equation:

\[ P(\text{admission}) = 0.94 \times BETA(0.345, 0.878) \times X^2 \text{ test } p < 0.005 \]

The data used for generating this distribution was a list of admission probabilities generated by Peck et al. (2012) when they applied their logistic regression model to a test set from the VHA BHS data.

Patients then flow through a decision module that determines whether the patient actually requires admission using that patient’s assigned probability value. This decision is used at the end of the patients ED treatment to route them to the IU, but is made now in order to generate a perfect predicted bed demand. Thus, simulated patients have two attributes and the model tracks the sums of these attributes over all patients in the ED at a specific time.
2.2. Emergency department sub-model
Upon receiving their admission predictions, patients enter a 13 bed simulated VHA BHS ED and seize a bed for their treatment duration. This duration is based on the real VHA BHS data, which fit an Erlang distribution as follows:

\[ ED \text{ treatment time} = -0.001 + ERLA(78.2, 2) \]  
\[ \chi^2 \text{ test } p < 0.005 \]

After completing their simulated ED treatment, patients being admitted enter a queue to seize an IU bed. Admitted patients will only release their ED bed after an IU bed is assigned.

2.3. Inpatient unit sub-model
The IU sub-model contains 100 beds based on the assumption that a significant number of VHA BHS' 170–180 beds are reserved for elective admissions. To capture how information can affect decisions, and consequently flow, the model treats IU doctors as the decision maker and limited resource in the IU. All other support services are modelled as a single intermediate treatment delay process which is assumed to include the processing time and waiting time for those services.

Just before entering the inpatient unit, a patient is assigned an IU Length of Stay (LOS). This was calculated by analyzing the total LOS for 32,156 real visits to VHA BHS which fit a log-normal distribution as follows:

\[ Patient \text{ IU LOS} = -0.001 + LOGN(8.89, 17.1) \]  
\[ \chi^2 \text{ test } p < 0.005 \]

Some of this LOS can be attributed to a patient waiting to see their doctor in order to be prescribed their next course of treatment. As will be described below, this waiting is a key part of the model. Therefore it must be deducted from the LOS assigned to a simulated patient in order to isolate the total time needed to treat the patient. This waiting time deduction was not known for VHA BHS. Instead it was estimated using model calibration. When running the model using Scenario 1, it was found that dividing LOS by 2.5 resulted in a required treatment time which, combined with simulated waiting times, generated a total LOS distribution that closely resembled that of the true hospital.

After being assigned a total needed treatment time, patients begin the IU treatment process. Figure 2 is a representation of the IU sub-model logic. As can be seen in the figure, a patient first seizes a doctor for treatment. The patient can only be treated or discharged by this unique doctor from that point on. The patient then releases the doctor and goes through a randomized amount of intermediate treatment. At the end of the cycle, the amount of time the patient spent is deducted from the patient’s remaining total needed treatment time.

The patient goes through the cycle, of doctor treatment and intermediate treatment depicted in Figure 2, until they have depleted their assigned treatment time. At this point, rather than re-enter the treatment queue, the patient enters a queue to seize their assigned doctor for discharge. A patient waiting for discharge is in direct competition for their doctor’s attention with patients who are still receiving treatment. This competition results in the waiting times which were removed from the LOS discussed earlier. Scenarios for controlling this competition, based on the current state of the hospital, are described in Section 2.4 below. Doctors only accept patients from 7am to 8pm, at 8pm the doctors will finish processing patients that are in the queues but all others are held until the next day.

2.4. Bed management sub-model and model scenarios
The bed management sub-model has the ability to shift doctor priority between the treatment queue and the discharge queue. This priority is based on the information being sent
Inpatient sub-model patient treatment cycle.

Scenario 1: This is the baseline scenario where priority is set to discharge beginning at 1pm and ending at midnight.

\[
\text{if } T \geq 13 \text{ then } \text{Prio} = 1 \text{ else } \text{Prio} = 0
\]

Scenario 2: At a predetermined time of day priority is set to discharge for three hours.

\[
\text{for } t = 0 \text{ through } 23 \\
\text{if } T \geq t \text{ or } T < t + 3 \text{ then } \text{Prio} = 1 \text{ else } \text{Prio} = 0
\]

Scenario 3: Priority is set to discharge while a time varying designated difference between the number of occupied ED beds and available IU beds, or Crowding Index, has been reached or exceeded.

\[
\text{F(ED}_{n}\text{)} = \begin{cases} 
0 & \text{if } \text{ED bed } n \text{ is occupied} \\
1 & \text{if } \text{ED bed } n \text{ is empty}
\end{cases}
\]

\[
\text{Crowding Index} = \sum_{n}^{13} \text{F(ED}_{n}\text{)} - \text{IUBeds}
\]

\[
\text{while Crowding Index} \geq S(t), \text{Prio} = 1 \text{ else } \text{Prio} = 0
\]

Scenario 4: Priority is set to discharge while a time varying designated difference between the imperfectly predicted IU bed need and IU bed availability, or Imperfect Index, is reached or exceeded.

\[
\text{P(PT}_{i}\text{)} = \text{Imperfectly predicted probability that patient ‘i’ will be admitted, 1 if patient has completed ED treatment and is awaiting admission.}
\]
**Table 1.** Simulation output data vs. VHA BHS monthly average data for validation

<table>
<thead>
<tr>
<th></th>
<th>West Roxbury</th>
<th>Model</th>
<th>Difference (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Patients</td>
<td>1240.5</td>
<td>1137.7</td>
<td>8.3</td>
</tr>
<tr>
<td>Admitted (%)</td>
<td>34</td>
<td>28</td>
<td>17.6</td>
</tr>
<tr>
<td>ED Wait Time (hours)</td>
<td>0.17</td>
<td>0.11</td>
<td>54.5</td>
</tr>
<tr>
<td>ED LOS Admitted (hours)</td>
<td>4.01</td>
<td>3.05</td>
<td>23.2</td>
</tr>
<tr>
<td>ED LOS No Admit (hours)</td>
<td>2.19</td>
<td>2.36</td>
<td>12.2</td>
</tr>
<tr>
<td>Boarding time (hours)</td>
<td>0.28</td>
<td>0.28</td>
<td>0</td>
</tr>
<tr>
<td>IU LOS (days)</td>
<td>10.2</td>
<td>7.53</td>
<td>26.2</td>
</tr>
</tbody>
</table>

**Imperfect Index** = \( \sum_{i}^{} P(PT_i) - IU\text{Beds} \)

while **Imperfect Index** = \( S(t) \), \( Prio = 1 \) else \( Prio = 0 \)

Scenario 5: Priority is set to discharge while a time varying designated difference between the perfectly predicted IU bed need and IU bed availability, or Perfect Index, is reached or exceeded.

**Scenario 6:** The current best practice of discharge by noon where discharge is prioritized for any time before noon (Vicellio et al. 2008).

\[ \text{for } t = 0 \text{ through } 23 \]

\[ \text{if } T < 12 \text{ then } Prio = 1 \text{ else } Prio = 0 \]

**2.5. Model cases**

Scenario sensitivity was studied by varying hospital characteristics through three cases:

*Case 1* represents the hospital baseline and had 25 IU doctors and no non-value-added (NVA) admission delay.

*Case 2* had 25 IU doctors and a variable NVA delay, between the ED and IU, which is normally distributed with a mean of 30 minutes and a standard deviation of 15 minutes. This delay occurs after an IU bed is assigned, but before the

**Fig. 5.** Average IU boarding time (and 95% confidence intervals) with shifting 3-hour discharge priority start times Case 1 (top), Case 2 (middle), Case 3 (bottom).

**Fig. 6.** Optimized sensitivity schedule using Scenario 3 (Crowding index) for Case 1 (top), Case 2 (middle), and Case 3 (bottom).
ED bed is released. This delay can be interpreted as delay of ED staff in receiving the assignment, delay of hospital bed managers from communicating the assignment, extra cleaning requirements, room set up delay, transportation delay, etc.

Case 3 had 20 IU doctors and no NVA delay.

### 2.6. Calibration, validation, and assumptions

Face validity was established by presenting the model to medical experts. Historical validity is established by looking at the outputs of the model and comparing them to the true VHA BHS data.

The model relies on the abstraction of the IU mechanics depicted in Figure 2. Key validation must therefore be based on IU outputs. All validation was based on Scenario 1, which represents the common practice at many hospitals. Figure 3 shows the normalized hourly IU discharge rates histogram for the real and simulated systems. The fit is not perfect, the simulated system uses clear rules which guide behavior and does not account for deviations which occur in real life. For example, the simulation states that at 8pm, IU providers cease to accept patients to the treatment or discharge queue, but finish processing patients who are currently in queue. This leads to very few patients being discharged between 3am and 9am. In reality, VHA BHS has patients who get discharged during these times due to special circumstances. Also, it should be noted that the baseline scenario has IU providers prioritize treatment in the morning until 1pm. While this is generally the case in the VHA BHS system, providers will occasionally choose to prioritize early discharge, this causes a higher incidence of morning discharge for the real VHA BHS compared to the simulated hospital.

Figure 4 shows the normalized IU LOS histogram for the real and simulated systems. Table 1 compares other key performance and patient data. This data also shows some slight disparities between the simulated and real hospital, such as lower LOS in the simulated hospital. The cyclical system which drives the model (Figure 2) is a simplification of the true system. The validation figures and table show that the pattern for the real hospital and for the simulated hospital are not exactly the same, however, the simulated pattern is not unreasonable for a realistic fictional hospital. Although this means that the model is not a perfect fit for the VHA BHS ED/IU system, the results are close enough to suggest that the real and simulated systems have

**Fig. 7.** Optimized sensitivity schedule using Scenario 4 (Imperfect index) for Case 1 (top), Case 2 (middle), and Case 3 (bottom).

**Fig. 8.** Optimized sensitivity schedule using Scenario 5 (Perfect index) for Case 1 (top), Case 2 (middle), and Case 3 (bottom).
Table 2. Difference in average ED boarding times between scenarios: Case 1 – no delay, full IU physician capacity

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Crowding Index</th>
<th>Imperfect Index</th>
<th>Perfect Index</th>
<th>Time Based</th>
<th>Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>3, Crowding Index</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4, Imperfect Index</td>
<td>NA (0.32)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5, Perfect Index</td>
<td>1.68 (0.02)</td>
<td>2.46 (≪ 0.05)</td>
<td>-2.94 (≪ 0.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6, Time Based</td>
<td>NA (0.10)</td>
<td>NA (0.54)</td>
<td>-4.62 (≪ 0.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1, Baseline</td>
<td>-2.94 (≪ 0.05)</td>
<td>-2.16 (≪ 0.05)</td>
<td>-1.68 (0.03)</td>
<td>-2.04 (≪ 0.05)</td>
<td></td>
</tr>
<tr>
<td>Improvement over Baseline</td>
<td>11.69%</td>
<td>8.59%</td>
<td>18.38%</td>
<td>6.68%</td>
<td></td>
</tr>
</tbody>
</table>

**Hypothesis: µx = µy for IU boarding times with no NVA delay, rejecting p < 0.05**

similar dynamics, therefore simulation scenario results may be applicable to the real system, with adjustment.

3. Results

3.1. Time-based discharge priority

For each of the three cases, Figure 5 shows how IU boarding time changes with the time of day that discharges are emphasized (Scenario 2). The error bars in each figure represent the 95% confidence interval for the data point based on 5000 replications of the simulation. The figure suggests that emphasizing discharge is more valuable early in the day and detrimental later in the day, just as Vicellio et al. (2008) asserted with the popular discharge by noon heuristic. By showing behavior that is commonly accepted as true, this data further validates the model dynamics. This also means that Scenario 6 is a special, best case of Scenario 2 and therefore only Scenario 6 will be used for further comparative scenario analysis.

Each of the graphs in Figure 5 show a pattern of behavior based on the unique dynamics of the model, but may also reflect the true operations of the hospital. For example there are boarding time peaks when discharge priority begins in the afternoon; these may be the result of doctors not seeing treatment patients in the afternoon which causes them not to be seen until the next day. Similarly in a real hospital, encouraging discharges at a specific time may negatively affect flow by interacting with the hospitals emergent schedule based on staffing levels, lunch hours, clinic hours, patient arrival patterns, and educational sessions.

3.2. Information-based discharge priority

To enable Scenarios 3–5 sensitivity to crowding and predicted admissions, S(t), was an independent variable for each hour. Figures 6, 7, and 8 show the optimized sensitivities that minimize ED boarding time when using the ED Crowding Index, Imperfect Index and Perfect Index (Scenarios 3–5 above) in each of the three cases. These were generated using OptQuest for Arena version 4.03 (OptTek Systems Inc. 2006). For example the graphic at the top of Figure 6 depicts that ED boarding was found to be minimized for Scenario 3/Case 1 when treatment is always prioritized between 8pm and 8am, then discharge is prioritized between 8am and 8pm whenever the crowding index exceeds 4.

Table 3. Difference in average ED boarding times between scenarios: Case 2 – NVA delay, full IU physician capacity

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Crowding Index</th>
<th>Imperfect Index</th>
<th>Perfect Index</th>
<th>Time Based</th>
<th>Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>3, Crowding Index</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4, Imperfect Index</td>
<td>NA (0.16)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5, Perfect Index</td>
<td>NA (0.93)</td>
<td>NA (0.19)</td>
<td>-1.98 (≪ 0.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6, Time Based</td>
<td>-2.04 (≪ 0.05)</td>
<td>NA (0.18)</td>
<td>-4.56 (≪ 0.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1, Baseline</td>
<td>-4.62 (≪ 0.05)</td>
<td>-3.60 (≪ 0.05)</td>
<td>-2.58 (≪ 0.05)</td>
<td>-3.28 (≪ 0.05)</td>
<td></td>
</tr>
<tr>
<td>Improvement over Baseline</td>
<td>17.58%</td>
<td>13.70%</td>
<td>17.35%</td>
<td>9.82%</td>
<td></td>
</tr>
</tbody>
</table>
Tables 2, 3, and 4 show the statistically significant difference in ED boarding time between the optimized information sharing scenarios (Scenarios 3–5) and the time based scenarios (Scenarios 1 and 6), for each case respectively.

The optimized schedules control system sensitivity to prediction and crowding information and perform very well using the IU boarding time metrics. Average values and variability for the index scenarios are significantly reduced compared to the baseline and time based scenarios. This means that, when combined, an index and a carefully chosen sensitivity schedule have the potential for greater performance than the discharge by noon heuristic that is currently the industry standard.

Flexibility in sensitivity scheduling, when using predictive and crowding indexes, allowed the system to compensate for the NVA delay and reduced IU doctor capacity. This compensation lead to consistently greater improvements in waiting/boarding times than the discharge by noon scenario, when compared to the baseline. Also, while discharge by noon does seem to have some added benefit in the case of the NVA delay, it was not effective at managing the system when resources were reduced, leaving it statistically similar to the baseline case.

It is worth noting that, using the optimized schedules, all three index types (crowding, perfect, and imperfect) were capable of generating superior performance. However, it is unclear that one index was significantly better than another. This means that using the imperfect prediction model proposed by Peck et al. (2012) may be good enough in the true hospital and investment in a more perfect method of prediction could be a waste of resources. Similarly it means that a crowding metric could be used instead of prediction, however if the system is not guided by a sensitivity schedule and associated definitive actions, it is unlikely to have the same intuitive incentive as the prediction values.

### 4. Conclusion, limitation and future work

Prediction is becoming a more common tool in hospital management. Emergency Department flow is a long studied issue that remains a chief concern to many hospitals. Recent studies have suggested that prediction can be used to drive behavior in the Inpatient Unit and improve flow. This paper showed that, in a simulated hospital, sharing prediction and state information from the ED to the IU does indeed have the ability to improve flow and reduce the effects of non-value-added delays.

Specifically, this study shows that hospital managers can use a summative predictive measure, such as the one derived by Peck et al. (2012), in order to drive hospital flow more effectively than a simple time based heuristic and almost as effectively as a system which uses perfectly predicted information. Achieving this improvement required carefully chosen schedules that dictate hourly sensitivity to ED crowding and admission prediction indexes.

The study presented in this paper has some inherent limitations in that it is simulation based. Despite calibrations made to the system, the validation procedure shows that the simulation does not directly match the true hospital system. To that end, the exact dynamics of the simulation are different than the true hospital and therefore the scenarios that were created to optimize flow, based on these dynamics, are likely not directly transferable to the true hospital.

While there is a clear benefit to finding the optimized schedules in the simulation, it is unlikely that an hourly schedule can ever be truly found in VHA BHS. Instead, more practical value would be derived from applying this simulation to finding semi-optimal, simplified solutions where sensitivity is held at a specific level for 2, 3, or even 4 hours, rather than varying on an hourly basis. The search for practical, slightly sub-optimal schedules is saved for future work. Similarly, implementing realistic schedules in a true hospital setting should be a focus of future work.
Acknowledgments

The authors thank the staff of the VHA Boston Healthcare System for their support and inputs; MIT Professor Stephen Graves for feedback on model conception and interpretation. This research was supported by the Department of Veterans Affairs through the Veterans Integrated Service Network 1 (New England) Veterans Engineering Resource Center and the VA Office of Systems Redesign. Any views expressed herein represent those of the authors and do not necessarily reflect those of the Veterans Health Administration.

References


