Proceedings of the 2014 Winter Simulation Conference
A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, eds.

THE IMPACT OF HOURLY DISCHARGE RATES AND PRIORITIZATION ON TIMELY ACCESS TO INPATIENT BEDS

Asli Ozen
Hari Balasubramanian

University of Massachusetts, Amherst
160 Governors Drive
Amherst, MA 01002, USA

Patricia Samra
Mike Ehresman
Haiping Li
Todd Fairman
Joan Roche

Baystate Medical Center
759 Chestnut Street
Springfield, MA 01199, USA

ABSTRACT

We develop an empirically calibrated hospital-wide simulation model to represent a time-varying, multiserver queuing network with multiple patient classes. The focus is on quantifying the impact of discharge profiles to alleviate inpatient bed congestions. A discharge profile is defined by (a) discharge window, which specifies hours of the day discharges are allowed; and (b) the maximum capacity for discharges in each hour of the window. Results of our simulation model show that a more responsive policy that prioritizes discharges in units with longer admission queues can significantly reduce waiting times (40% reduction in queues). In comparison, an early in the day discharge policy has a lower impact on improving bed congestions; we also find that early in the day discharges are very hard to realize in practice. Further, expanding the discharge windows by only 2 hours in the evening (7-9 PM) creates the same benefit, and is more realistic.

1 INTRODUCTION

Inability to satisfy bed requests in a timely manner causes hospital wide congestions. These effects include but are not limited to: patients waiting long hours in the ED or a surgical area for an inpatient bed; patients not being placed in their primary unit (i.e. off-service placement); urgent patients bumping less critically sick patients from ICUs to “step-down units”; and refusing transfers from other hospitals.

In this paper, we use an empirically calibrated hospital-wide simulation model to quantify the impact of *discharge profiles* on timely access to inpatient beds. We define a discharge profile by (a) discharge window, which specifies the hours of the day discharges are allowed; and (b) maximum discharges a hospital can manage in each hour of the window. The discharge process is typically complex and involves many moving parts; it requires the timing of physician rounds; availability of nurses, case managers and social workers; and coordination with families and post-hospitalization facilities to all come together. Rather than considering these individual aspects, which are quite difficult to estimate, we instead use discharge profiles to model the hospital’s aggregate hourly capacity. Timely access in our model is measured by the average number of patients waiting for an inpatient bed (average queue length) in any hour. This queue size includes all patients waiting in the Emergency Department (ED), Post-Acute Care Unit (PACU) or other locations after the physician has made a request for an inpatient bed.

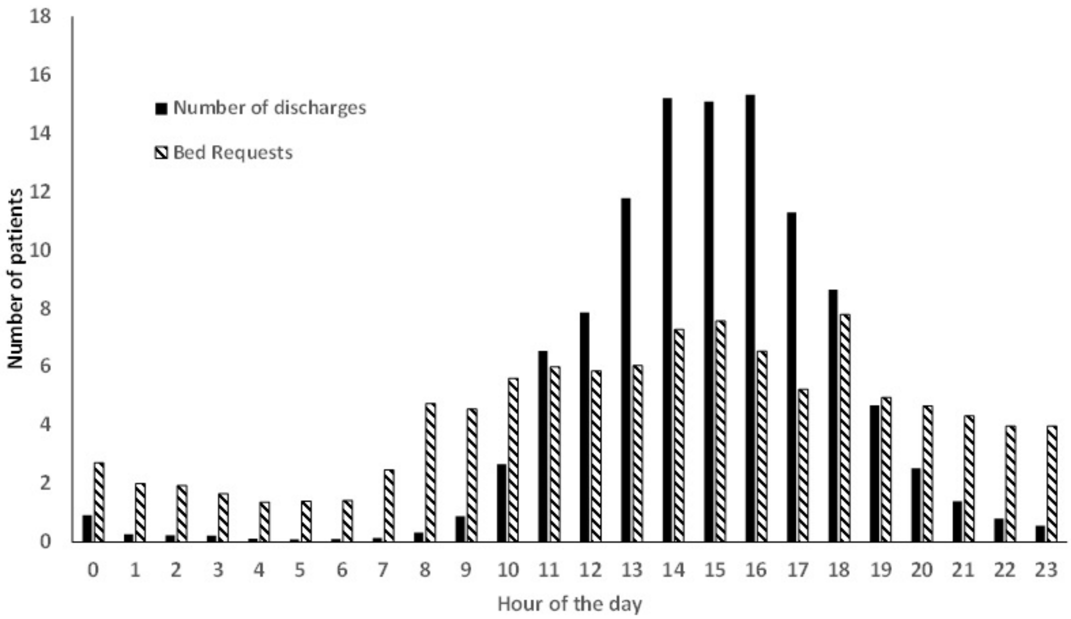


Figure 1: Admission and discharge rates.

Our simulation considers a time-varying multi-server queuing network model with multiple request sources and multiple patient classes. We capture two sets of queues in this network: admission queues for individual hospital units and a discharge queue for the entire hospital. The discharge queue consists of patients whose length of stay is complete but have not left since the hospital does not have the required hourly discharge capacity. Arrival rates, source of arrivals (ED, surgical areas, community physicians etc.), diagnostic categories, length of stays are all randomly sampled based on a year's worth of inpatient flow data from an acute care medical center in the Northeast of the U.S.

On average each day there are around 100 bed requests and discharges at this medical center. Figure 1 shows the mean number of daily inpatient bed requests and patient discharges by hour of the day at our partner hospital. The time-varying nature of the admission request and discharge process can be clearly observed. Notice that the typical window for discharges is 10 AM-7 PM, and the number of discharged patients varies in each hour of the window as shown in Figure 1. Discharges peak in the afternoon between 2-4 PM, producing a bell-curve centered on these afternoon hours.

We investigate in this paper whether discharge profiles different from empirically observed one in Figure 1 (our baseline) can improve timely access to inpatient beds. How much improvement in queue length can be achieved if most discharges happened by noon? Is such an early in the day discharge profile feasible in practice? What if discharge hours were extended in the evening by a couple of hours? What if the hospital is more responsive and prioritizes discharges in units that had longer admission queues? Note that in evaluating all these approaches, we are simply varying the hospital's hourly discharge capacity or changing prioritization; patients are never discharged unless their clinical length of stay is complete.

Results of our simulation show that prioritizing discharges has the greatest impact on queue sizes when compared to the empirical baseline (40% reduction). The early-in-the-day discharge profile has a lower impact on improving bed congestions, and is very hard to realize in practice because it requires the hospital to more than double its current maximum discharge capacity. Finally, expanding the discharge windows by only 2 hours in the evening (7-9 PM) creates the same benefit as early-in-the-day discharge policy, and is more realistic.

A less tangible but equally important contribution is the fact that the entire simulation modeling process - assumptions, data inputs, analysis of outputs, implications for practice, implementation of results – was conducted over a 3-year period with constant input provided by key stakeholder groups at our partner hospital.

2 LITERATURE REVIEW

The literature is reviewed in two parts: first we summarize the literature on hospital-wide flow models and secondly discuss why we used a simulation model as opposed to a queuing model by reviewing a list of queueing models applied to healthcare networks with time-varying arrivals.

2.1 Hospital-Wide Flow Models

Modeling and improving patient flow has been studied extensively in the literature. We are specifically focusing on the hospital-wide optimization rather than unit specific optimization models (see Williams (2006) for a detailed literature survey). These might optimize a specific part of the hospital but will not consider impact on the hospital as a whole (Helm and Van Oyen 2010, Bekker and Koeleman 2010). Thus, we turn our focus to hospital wide optimization models.

Various IEOR techniques have been used including queuing models (Bekker and Koeleman 2010, Shi et al. 2012, Armony et al. 2012, Green and Nyugen 2001, Green 2003), mixed integer programming (Helm and Van Oyen 2010), Markov decision processes (Helm, AhmadBeygi, and Van Oyen 2011, Helm, Lapp, and See 2010) and stochastic optimizations (Best et al. 2012), in order to alleviate the inpatient bed congestion. Simulation models are used frequently to represent the inpatient flows in hospitals since Hancock and Walter (1979) (Helm et al. 2011, Hancock and Walter 1979, Proudlove et al. 2007, Montgomery and Davis 2013, Shi et al. 2012).

2.2 Why Simulation and Not Queuing?

Both queueing and discrete event simulation models have been used extensively in modeling healthcare problems. Queuing models are simpler, require less data, and provide more generic results than simulation (Green 2011). On the other hand, discrete-event simulation, which is more flexible permits modeling the details of complex patient flows. Kolker provides examples that clearly demonstrates why in most cases discrete event simulations are superior and preferred to queuing models, including an example in a healthcare setting with time-varying arrival rates (Kolker 2010). Finally Green (2011) discusses that for these types of queuing systems "using queuing models is inappropriate for estimating the magnitude and timing of delays, and a simulation model will be far more accurate". Also, simulation models are extremely helpful for visualization and for "selling" the model to the stakeholders (Montgomery and Davis 2013).

One of the most extensive research on inpatient flows is conducted by Armony et al. (2012), who analyze the hospital-wide patient flow from a queuing approach, based on Exploratory Data Analysis (EDA) and arrival (birth) and death (departure) processes. Shi et al. (2012) is the first to explore stochastic models to analyze the impact of effective discharge policies. To do so, they develop an analytical model and then use a very rigorous simulation model that mimics the inpatient operations in a Singaporean hospital. They study the effect of early-in-the-day discharge, that was implemented in the hospital, on ED waiting. One of the major findings is that instead of this discharge policy, a hypothetical discharge distribution, which still discharges 26% of patients before noon, but shifts the discharge peak time to 8-9 AM, provides significant improvement in waiting times.

Our problem, on the other hand, consists of multiple patient categories each exhibiting a different time-varying arrival process; length of stay distributions are also dependent on the patient category. We have looked at implementing time-varying arrival rate queuing models to our inpatient flow model. However, the queuing models are unable to tackle the complexity of the problem. As discussed in Shi et al. (2012), time-varying models have been studied extensively for call centers. Unlike call centers our model has extremely long service times and the number of servers (beds) cannot be adjusted in a short time window. Thus, existing approximation methods generated for call center models are not applicable to our hospital model. As a result, like Shi et al. (2012) we turned our focus to simulation models based on sampling historical data collected from a large tertiary care hospital – an acute care medical center in the Northeast of the US in our case.

3 DATA COLLECTION AND ANALYSIS

We analyzed data for all patients who used an inpatient bed at our partner hospital from May 2010 to April 2011. We used anonymous patient records which included patient age and gender, and diagnoses related categorizations. These include the diagnosis related groups (DRGs) and major diagnostic categories (MDCs). This MDC categorization was initially created for the claims and administrative process; each MDC aggregates related DRGs into a single broader category - for example, two such categories are "Respiratory Diseases" and "Circulatory Diseases". There are 25 MDCs and this keeps the model concise and tractable.

Time stamps for each patient were also part of the data we analyzed and in fact form the basis of some key inputs in our simulation model – see Figure 2. A patient may enter the hospital information system by registering through the ED or surgical unit or a physician's office or other sources. After the patient goes though the assessment, consultation and care process, the relevant physician or care provider decides that the patient should be admitted to an inpatient bed in a desired unit. This is the bed request time and in our simulation model it translates to a patient arrival.

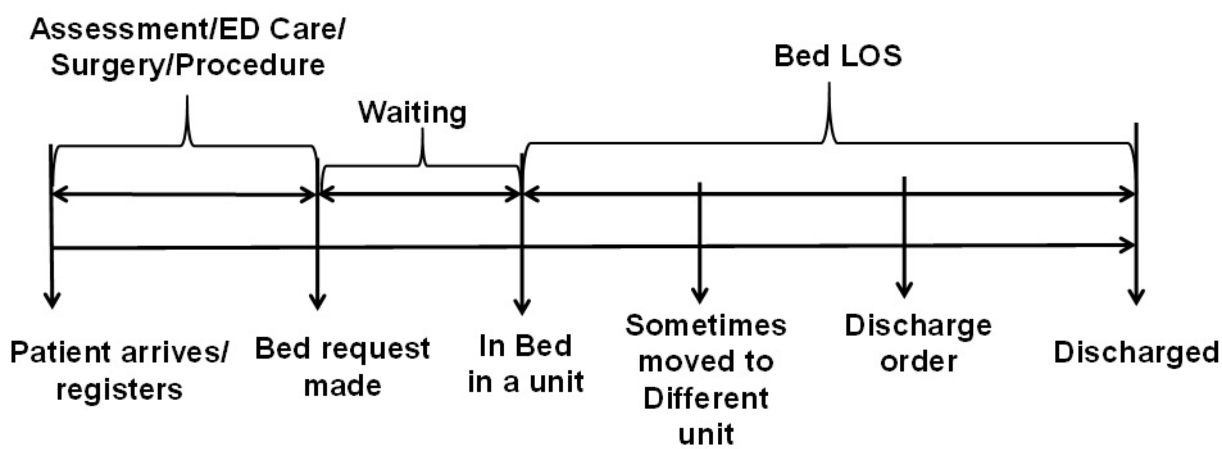


Figure 2: Important time-stamps in the admission and discharge process.

The patient then waits until a bed is available, and is then admitted. After staying for some duration in the inpatient bed, the patient is discharged. The important point here is that by length of stay (LOS) we mean time spent by the patient in the inpatient bed. In Figure 2 this is "the discharge time" minus the "in the bed" time. From the point of view of our simulation model, inpatient bed LOS is the "service time" and number of inpatient beds in a unit are the "servers".

Inpatient bed LOS can vary significantly from patient to patient. In addition to regular inpatients (27,000 in our one year data), there are two separate categories of patients called "day-stay" and "observation patients". Day-stay patients are patients who undergo small procedures like tonsillectomy and stay for 24 hours or less in an inpatient bed (ASCA 2013). Observation patients refer to those patients whose conditions that can be treated within 48 hours, or when the cause for the symptoms has not yet been determined (CMS 2011).

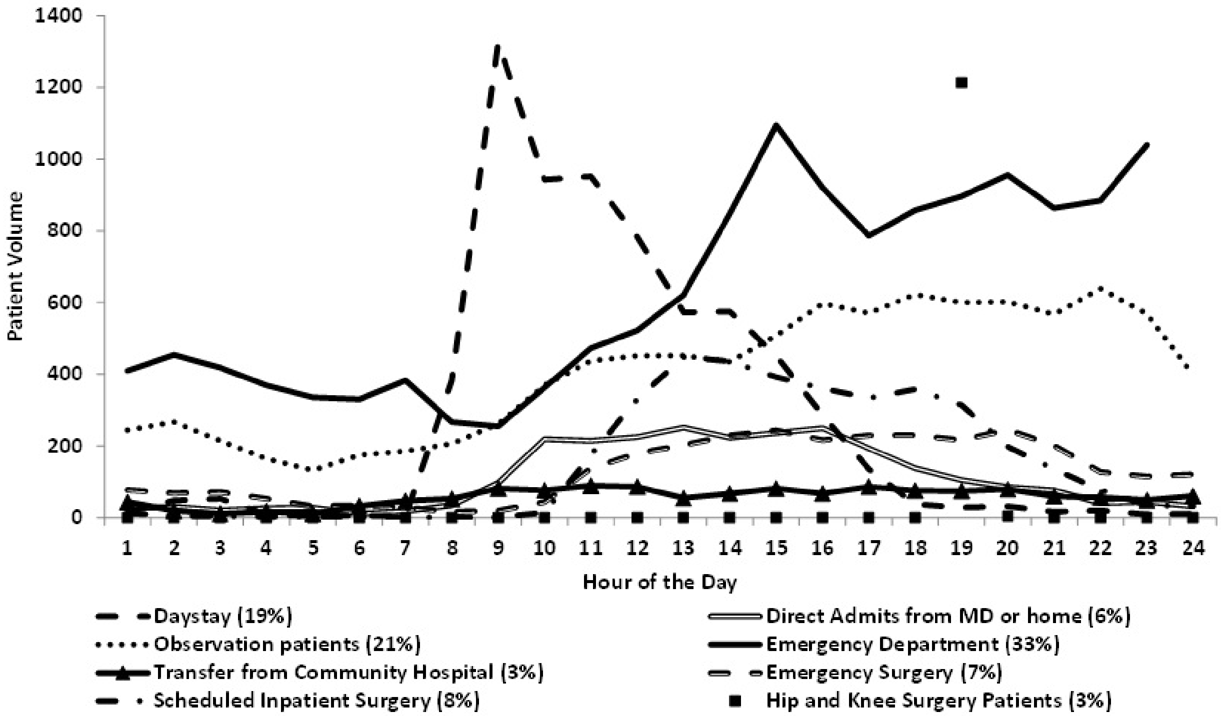


Figure 3: Arrival pattern by patient sources.

Regular inpatient bed requests can get admitted through the emergency department, surgical units (this includes elective surgeries as well as emergency surgeries), from physician offices (direct admits), from other community hospitals (transfers from other hospitals). Categorizing patients by these admit sources and their MDC represents fairly accurately which units they get admitted to and how long they stay in an inpatient bed. Figure 3 shows the total annual bed requests of the major sources with respect to hours of the day. We can clearly see the time-varying arrival process for each source.

There are 25 departments that the patients get admitted to, which total to 575 inpatient beds. The medical specialties include: adult respiratory, oncology, day-stay, medical, observation, congestive heart failure (CHF), surgical, interventional, critical care, women health, renal, neurology, orthopedics and pediatric medical and surgery units. We have analyzed the units in terms of their bed capacity, daily arrival rate, the mean and variance of LOS, utilization and percentage of discharges before noon. Table 1 provides twelve of the most highly utilized units. For simplicity, "utilization" in this table is calculated as daily bed request times the average LOS, divided by the number of beds in the unit.

4 SIMULATION MODEL AND ANALYSIS

Figure 4 shows the main idea behind our simulation model. The simulation updates on an hourly basis. There are M sources of inpatient bed requests. The number of requests from source i in hour t is denoted by the random variable $\lambda_{i,t}$ and is sampled randomly without replacement in order to reflect the time of day and day of week effect. These requests fall into some MDC category and are consequently mapped into demand for N inpatient units. The total bed requests for unit j in hour t is denoted by the random variable $\lambda^{'}_{j,t}$. As an example, for Monday 8 AM, we randomly sample, without replacement, from arrival source, MDC and desired unit requests observed on 52 Mondays at that exact hour. Each unit has B_j beds, and each unit is a time-varying $G/G/B_j$ queue. The arrival rate in each hour follows some general stochastic process; Poisson arrival rates are not a bad assumption (especially for ED admissions), but in our case,

Table 1: Unit specific analysis

Unit	LOS (days)	Std Dev LOS	Daily Rate	Bed capacity	Utilization	Discharge % before noon
Medical Telemetry	5.09	5.85	5.16	26	101%	16%
Cardiac CHF	4.19	4.29	7.58	32	99%	24%
Cardiac interventional	2.64	3.27	11.87	32	98%	23%
Neurological	3.91	5.08	9.94	41	95%	13%
Renal	3.25	4.19	6.92	24	94%	27%
Adolescents	2.6	3.38	2.86	8	93%	27%
Medical Respiratory	5.27	7.35	5.36	31	91%	29%
Surgical/ Orthopedic	4.74	5.02	6.47	34	90%	15%
General Medical	3.26	3.21	11.95	44	89%	23%
Psychiatric	8.68	10.87	2.79	28	87%	21%
Intermediate Surgical	5.41	6.25	6.97	44	86%	9%
Short Stay Surgical	4.25	5.61	6.28	32	83%	12%

we use arrivals sampled from historical data, hence G in the queuing notation. The random variable LOS_j indicates the service time in unit j and follows some general distribution. Right-skewed distributions with high standard deviations such as the lognormal are a good choice for LOS_j but do not hold for all units; we therefore choose to randomly sample from observed values in our one year dataset.

We assume that a patient waits until she is placed in a bed in the desired unit. Therefore, in each unit j, a queue Q_j develops consisting of those patients waiting for a bed in the desired unit to become available. Note that the queue is not a physical waiting line of patients; rather it consists of patients waiting in different parts of the hospital (ED, PACU) and also other hospitals. In each hour, bed requests are fulfilled on a first come first served (FCFS) basis. The patients are ready to be discharged from the hospital after their LOS is completed. They join a discharge queue which has a capacity of D_t discharges in hour t. To start with, patients are discharged on a FCFS basis; later we consider prioritizing discharges in units which have long queues. The bed is available to be assigned after the bed turnover time (a deterministic value) is complete.

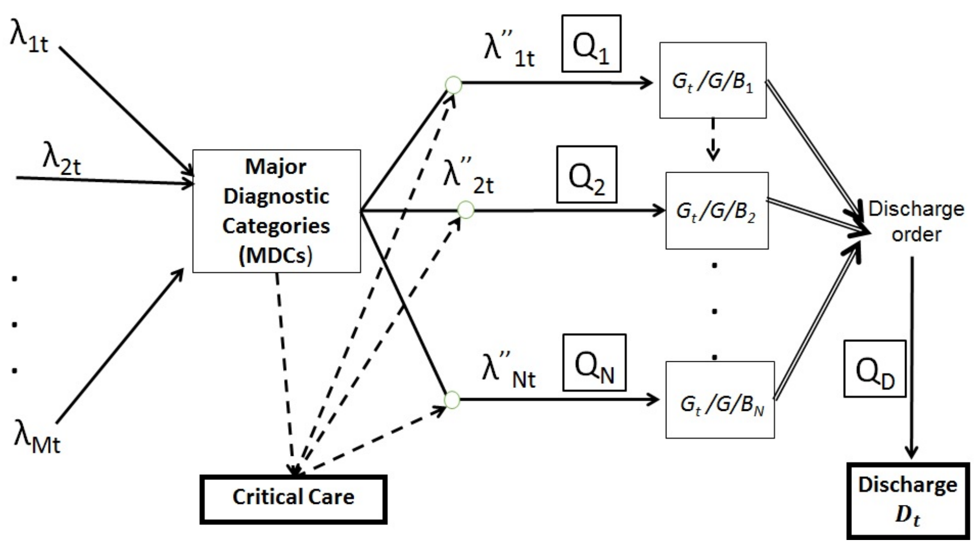


Figure 4: Queuing framework.

The simulation was coded in C#. We ran the model for a year, with hourly increments, and kept the warm-up period as 2 months. We have used 10 replications, following Shi et al. (2012) who also use 10 replications to perform their analysis. More replications will lead to a higher accuracy, however, due to the computational complexity (around 2 hours for each run) we limit the number of replications. In order to have an unbiased comparison while comparing different discharge profiles, we use a common set of random patients for each replication. This is the common random numbers (CRN) approach which serves as a variance reduction technique when comparing different policies.

5 ANALYZING THE IMPACT OF DISCHARGE POLICIES

Recall that the purpose of this paper is to test different discharge profiles by changing the D_t and observing the effect on queue lengths Q_j. We are evaluating 3 components of discharge profiles: (1) Discharge windows which determines the hours of the day when the discharges are allowed; (2) The maximum capacity for discharges in each hour of the discharge window (the D_t values); and (3) The prioritization of discharges in each hour based on admission unit queues (i.e. which patients should have first access to discharge capacity in a given hour). We evaluate different combinations of these 3 components and compare it with the baseline which represents our partner hospital's discharge operations. We now present all the discharge profiles we test in our simulation and also provide the rationale for each.

5.1 Baseline

The baseline discharge profile for our partner hospital was briefly described in the Introduction. The discharge window is currently from 10 AM-7 PM. The hourly maximum capacity for discharges in each hour of this window is set to the average number of discharges achieved by the hospital in the one year period studied. Starting with the hour 10-11 AM, we set D_t equal to 5, 7, 11, 12, 14, 18, 16, 10, 6, 5 until 7 PM. For all other hours D_t is 0. Currently at the hospital, there is no obvious prioritization of discharges, so we assume in the baseline that discharges are done on a FCFS basis.

5.2 DP1: Early-in-the-Day Discharge Policy, 10 AM-7 PM, No Prioritization

In this discharge profile, D_t is only restricted to be less than the remaining number of average daily discharges (i.e. no hourly limit, but a daily limit). Due to this most of the patients leave the hospital in the first 2 hours of the window (10 AM-noon). These patients have already completed their LOS overnight and have been waiting for the hospitalist to discharge them; hence the name early-in-the-day discharge. Discharges are carried out on a FCFS basis (no prioritization).

5.3 DP2: Expanded Discharge Windows (10AM-9PM), No Prioritization

Our partner hospital collaborators also wanted to test the feasibility of expanding discharge hours (an expanded window from 10 AM to 9 PM) as an alternative to early-in-the-day discharges, because they felt that discharges by noon were very difficult to implement in practice. Each hour in the expanded window $D_t = 10$ and 0 otherwise. Our collaborators felt that a maximum of 10 per hour was a reasonable workload for the clinical staff.

The expanded window is also more in line with the hospitalist's natural prioritization rules. They can see recently admitted patients, who need more urgent attention, in the morning, and get to the patients who are ready to be discharged later in the day. The families of patients may be more available to pick up patients in the evening rather than during the day.

5.4 DP3: Prioritization of Discharges

This discharge profile has the exact same D_t values as the Baseline discharge profile above. However, while in the Baseline discharges are prioritized on a FCFS basis, in DP3, the hourly capacity D_t is first used for units whose Q_j values exceed a prespecified threshold (for this paper, we use a value of 2). Physicians and related staff first focus on discharging patients from units whose queue lengths exceed the threshold. However, it is important to point out that these are not hasty discharges (which may cause readmissions); only patients who have completed their clinical length of stay can be discharged.

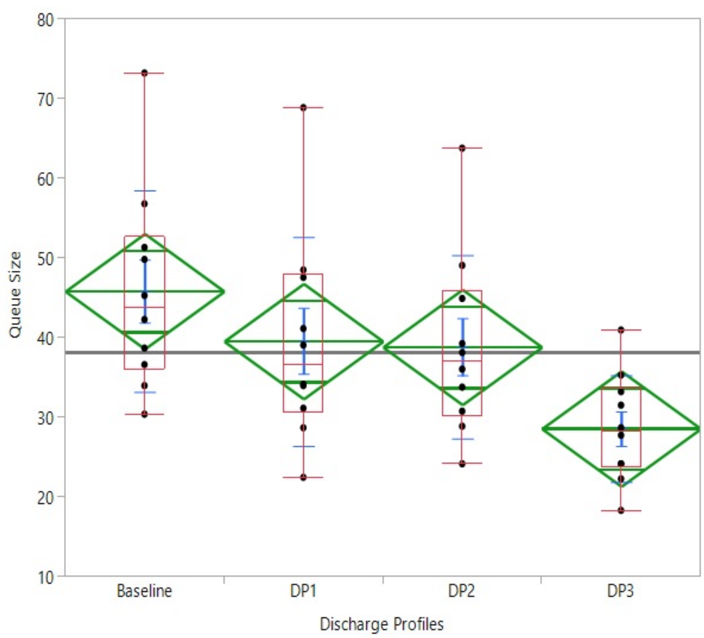
6 RESULTS OF THE SIMULATION

Before trying to improve the existing system, validation was the initial step. We have done the validation in two steps like Montgomery and Davis (2013): stakeholder face validation and comparison of means of the inputs and outputs, including but not limited to; waiting times, queue sizes and utilization levels for units, admission patterns and LOS values for patients. For example, the most utilized units in the hospital were also the most utilized in the simulation model.

6.1 Impact of Discharge Policies

We now present results for the average hourly queue size, with 10 simulation replications for the various discharge profiles. The average hourly queue size represents the average number of people waiting to be admitted to a desired unit waiting in ED or PACU, or in the community. We use one-factor ANOVA to analyze the differences in average queue size between the discharge profiles (Figure 5 (a)), and also the resulting pairwise Tukey test results (Figure 5 (b)).

The red lines in Figure 5 (a) represent the quantiles with the box plot; the top, bottom and middle of the green diamond indicates the upper and lower 95% confidence interval limits and the sample mean respectively. Lastly, the horizontal line across all discharge policies and replications represents the overall mean queue length. The discharge profiles are presented in descending order of the average queue size observed.



(a)

Discharge Profiles			Average Queue Size
Baseline	A		45.85
DP1	A	B	39.57
DP2	A	B	38.84
DP3		B	28.60

(b)

Figure 5: (a) Queue size ANOVA analysis; and (b) Connecting letters report for Queue Size.

The connecting letters report in Figure 5 summarizes the results of the all-pairs Tukey tests. If two discharge profiles are not sharing a letter their means are significantly different. However, statistical significance, while important to acknowledge, should not be confused with clinical significance. Clinical significance has a qualitative component; in our case, it is decided by our clinical collaborators. For example, a 6-person average reduction in queue size is non-trivial even though it may not be statistically significant.

Our results can be summarized as follows:

(1) If the majority of discharges happen before noon, as in the early-in-the-day discharge policy (DP1), then there are 6 less people waiting in the queue compared to Baseline (clinical significance) but there is no statistical difference.

(2) With 2 additional hours of discharge and a steady maximum capacity of 10 discharges in each hour (10 AM-9 PM, DP2), we have 7 fewer patients waiting compared to Baseline (clinically significant). Thus expanding discharge by two hours while limiting the maximum hourly discharge capacity to 10 produces the same effect as performing a large (and impractical) number of early in the day discharges. The reason is that the discharges are performed more evenly throughout the day; expanding the window helps patients, who have finished their clinical length of stay in the evening to be discharged rather than wait until the next morning.

(3) Using the current or empirically observed discharge capacity as the maximum capacity for each hour and a discharge window of 10 AM-7 PM with prioritization (DP3), produces a statistical improvement from the Baseline: it leads to 17 less patients waiting to be admitted. The only difference between Baseline and DP3 is prioritization: the only change in practice is that each hour the hospital staff (physicians, case-managers, nurses, valets and escorts) has to prioritize their discharge activities in units that have the longest front-end (admission) queues. We see also from Figure 5 that the higher percentiles of the average queue size are also reduced drastically.

7 DISCUSSION AND CONCLUSION

In summary, we find that prioritized discharges have the biggest impact on reducing queue sizes. There are a number of other findings that we plan to present in the full paper. For example, the recommended discharge policy in the literature overwhelmingly focuses on early-in-the-day discharges. When we analyze this discharge profile, we find that 70 discharges need to happen before 10 AM-noon. This is more than double what the hospital has achieved even during peak discharge hours (between 2-4 PM). All stakeholders at our partner hospital agreed such a behavioral change is unrealistic.

In the full paper, we will also discuss qualitative aspects of the discharge process in greater detail to provide much needed clinical context. Discharging a patient is a complex process and is not simply about vacating a bed/room: the transition of care of the patient needs to be carefully coordinated. This involves not only physicians and nurses at the hospital, but potentially also case-managers, social workers, pharmacies and special equipment, primary care physicians, insurance companies and eligibility considerations, family members, and skilled nursing facilities. Achieving all this in all units before noon is unrealistic, especially when physicians and nurses have to attend to newly admitted and more urgent patients. Rather, focusing on the units that have the longest admission queues - a state dependent discharge policy - is much more practical. Another option which seems to resonate with clinicians is expanding discharge hours from 7 PM to 9 PM. This allows the clinical team (with shifts slightly staggered) to carry out discharges later in the day. Family members are also more available for pick-ups later in the day. We will also present results at the level of each unit, and provide more insights in the full paper.

The simulation model was developed over a 3-year period with continuous input from all stakeholder groups: ED, surgical suites, bed placement office, nursing, finance, IT, leadership, process improvement etc. As a result, computer simulation and its potential to address key what-if questions has been widely disseminated in our partner organization and has prompted a healthy discussion on hospital-wide patient flow. In May 2013, the number of beds in one of the medicine units was increased, because our simulation model showed that queues for that unit were untenable. The prioritized discharge policy has been well received; to implement it in practice, an effort is underway to create a dashboard on queue sizes by unit, visible to all relevant groups in the hospital, so that the clinical staff can prioritize discharges accordingly. Finally, this research paved the way for an industrial engineering intern dedicated to developing new variations of our simulation model.

REFERENCES

Armony, M. and Israelit, S. and Mandelbaum, A. and Marmor, Y. and Tseytlin, Y. and Yom-Tov, G. 2012, September. "Patient Flow in Hospitals: A Data-based Queueing Perspective". Working paper.

ASCA 2013. "What is an ASC?". Available online.

Bekker, R., and P. M. Koeleman. 2010. "A Model of ICU Bumping". *Operations Research* 14:1564–74.

Best, T., B. Sandikci, D. Eisenstein, and D. Meltzer. 2012. "Managing Hospital Bed Capacity Through Partitioning Care into Focused Wings". *Chicago Booth Research Paper* 12 (64).

CMS 2011, February. "Are You a Hospital Inpatient or Outpatient?". Available online.

Green, L. 2003. "How Many Hospital Beds?". *Inquiry* 39 (4): 400–12.

Green, L. 2011, January. "Queueing Theory and Modeling". In *Handbook of Healthcare Delivery Systems*, edited by Y. Yih. CRC Press.

Green, L. V., and V. Nyugen. 2001. "Strategies for Cutting Hospital Beds: The Impact of Patient Service". *Health Services Research* 36:421–42.

Hancock, W. M., and P. Walter. 1979. "The Use of Computer Simulation to Develop Hospital Systems". *ACM SIGSIM Simulation Digest* 10 (4): 28–32.

Helm, J., M. Lapp, and B. See. 2010. "Characterizing an Effective Hopsital Admission Scheduling and Control Management System: A Genetic Algorithm Approach and Control Management System". *Proceedings 42nd Conf on Winter Sim.*.

Helm, J. E., S. AhmadBeygi, and M. Van Oyen. 2011. "Design and Analysis of Hospital Admission Control for Operational Effectiveness". *POMS* 20 (3): 359–74.

Helm, J. E., and M. Van Oyen. 2010, February. "Design and Optimization Methods for Elective Hospital Admissions". Technical Report 11-01, University of Michigan, Industrial and Operations Engineering.

Kolker, A. 2010. "Queuing Analytic Theory and Discrete Events Simulation for Healthcare: Right Application for the Right Problem". Online.

Montgomery, J. B., and K. Davis. 2013. "The Hospital Patient Flow Model: A Simulation Decision Support Tool".

Proudlove, N., S. Black, and A. Fletcher. 2007. "OR and the Challenge to Improve the NHS: Modelling for Insight and Improvement in Inpatient Flows". *Journal of the Operational Research Society* 58 (2): 145–158.

Shi, P. and Chou, M. C. and Dai,J.G. and Ding, D. and Sim, J. 2012, Sep. "Hospital Inpatient Operations: Mathematical Models and Managerial Insights". Working paper.

Williams, M. 2006. In *Patient Flow: Reducing Delay in Healthcare Delivery*, edited by R. W. Hall, Chapter 2: Hospitals and Clinical Facilities, Processes and Design for Patient Flow, 45–79. Springer.

AUTHOR BIOGRAPHIES

ASLI OZEN has a PhD in Industrial Engineering from the University of Massachusetts, Amherst. Her email address is aslozen@gmail.com.

HARI BALASUBRAMANIAN is an Associate Professor of Industrial Engineering at the University of Massachusetts, Amherst. His email address is hbalasub@admin.umass.edu.

PATRICIA SAMRA is the Director of Clinical Workforce Planning and Finance at Baystate Medical Center. Her email address is patricia.samra@baystatehealth.org.

MIKE EHRESMAN is the Manager of Decision Support at Baystate Medical Center. His email address is mike.ehresman@baystatehealth.org.

HAIPING LI is the Data Analyst of Emergency Services at Baystate Medical Center. His email address is haiping.li@baystatehealth.org.

TODD FAIRMAN is the Business Manager of Perioperative Services at Baystate Medical Center. His email address is todd.fairman@baystatehealth.org.

JOAN ROCHE is an Associate Clinical Professor at the University of Massachusetts, Amherst. She also holds a joint appointment as a Clinical Nurse Specialist at Baystate Medical Center. Her email address is joan.roche@baystatehealth.org.