ABSTRACT

Title of Document:	THE IMPACT OF RESOURCE MANAGEMENT ON HOSPITAL EFFICIENCY AND QUALITY OF CARE
	David Anderson, Doctor of Philosophy, 2013
Directed By:	Bruce Golden, Professor, Decision, Operations, and Information Technologies

Managing scarce resources plays a significant role in hospital operations. Effective use of resources (e.g., operating rooms, specialized doctors, etc.) allows hospitals to efficiently provide high-quality care to patients. In this dissertation, we study how hospitals manage scarce resources to identify systematic ways in which quality of care and efficiency might be improved. We study four different types of hospital resources: post-operative beds, specialist surgeons, resident physicians, and patient information. For each resource type, we show how better utilization could increase the quality of care delivered by the hospital or increase the efficiency of the system. We show that as post-operative bed utilization increases the discharge rate increases as well, meaning that bed shortages impact physician decision making. Further, we show that patients discharged on days with high bed utilization are significantly more likely to be readmitted to the hospital within 72 hours, which implies that poor bed management can lead to

worse health outcomes for surgical patients. We also study how quality of care differs between night and day arrival in trauma centers. Based on a large national dataset, we conclude that a lack of specialized resources at hospitals during the off hours leads to significantly worse patient outcomes, including higher mortality and longer lengths of stay. Further, we exploit a natural experiment to determine the impact that residents have on efficiency in an academic emergency department. Using regression analysis, queueing models, and simulation, we find that when residents are present in the emergency department, treatment times are lowered significantly, especially among high severity patients. Finally, we show two novel uses of medical data to predict patient outcomes. We develop models to predict which patients will require an ICU bed after being transferred from outside hospitals to an internal medicine unit, using only five commonly measured medical characteristics of the patient. We also develop a model using MRI data to classify whether or not prostate cancer is present in an image.

THE IMPACT OF RESOURCE MANAGEMENT ON HOSPTIAL EFFICIENCY AND QUALITY OF CARE

By

David R Anderson

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
Of the requirements for the degree of
Doctor of Philosophy
2013

Advisory Committee:
Professor Bruce Golden, Chair
Professor Guodong Gao
Professor William Rand
Professor Edward Wasil
Professor Paul Schonfeld, *Dean's Representative*

Dedication

To my family

Table of Contents

List of Tables

List of Figures

CHAPTER 1: INTRODUCTION

The dramatic rise in healthcare costs is an increasingly important political and economic issue facing the United States today. Spending on healthcare increased by more than 200% between 1990 and 2007 (Keehan et al., 2008). The Congressional Budget Office projects total healthcare spending will rise from 16.5% of GDP in 2009 to 26% by 2035 (Keehan et al. 2008). According to the Office of Management and Budget, there is as much as $800 billion a year in wasteful medical spending that does not contribute to better health outcomes (Orszag, 2009). Finding a way to slow the growth of medical costs while providing high-quality care is an urgent issue facing operations researchers today.

Growth in healthcare spending threatens to cause massive government budget deficits and significant increases in labor costs. Therefore, identifying systematic sources of inefficiency and studying ways to improve operations has been a focus of healthcare researchers. Interest in the area has been rising over the last decade. In 2012, the *New York Times* published 460 articles on healthcare costs, compared to 252 in 2000. The need to control healthcare costs has been recognized by government officials, the media, and academics alike.

The academic community has been at the forefront of the growing trend of research into healthcare costs and efficiency. From the growth of existing journals, such as *Health Care Management Science* to the introduction of new journals, such as *IIE Transactions on Healthcare Systems Engineering*, to new conferences, to new degree programs in healthcare management and analytics, healthcare has emerged as a significant research area.

Researchers have used a variety of different methods to address efficiency issues in hospitals. Researchers have used traditional operations research methods, such as queueing theory (Green, 2006) and optimization (Chan, et al. 201) to help improve efficiency and outcomes. Others have taken traditional operations management techniques, such as newsvendor models, and applied them to healthcare problems (Green et al. 2010). Recently, data mining and analytics techniques have been applied to large-scale data sets to identify types of inefficiencies, and to predict patients' costs and outcomes (Bertsimas et al. 2008).

Healthcare costs have also been an issue addressed by the federal government. The highest profile effort to reduce costs was the 2010 Patient Protection and Affordable Care Act (H.R. 1-111-148, §2702, 124 Stat. 119, 318-319, 2010). While much of the bill focused on expanding access to care, there were provisions, such as requiring insurance coverage of preventative care and a slow movement away from a fee-for-service model, enacted with the goal of lowering healthcare costs. Additionally, as part of the 2009 stimulus (American Recovery and Reinvestment Act), $25.9 billion dollars were appropriated for the promotion and expansion of health information technology (HIT), such as electronic medical records (H.R. 1--111th Congress: American Recovery and Reinvestment Act, 2009).

In Chapters 2 and 3 of this dissertation, we discuss how poor management of post-operative beds leads to poor patient outcomes, and inefficient care delivery. In order to perform surgery, there are surgical resources required, including an operating room, an anesthesia team, and nurses, and a down-stream recovery bed available for the patient. If there is no post-operative bed available, the surgery must be postponed or cancelled, which is undesirable for the surgeon, hospital, and patient. We show that as post-operative utilization rises and beds become scarce, the discharge rate of patients in the surgical recovery ward increases significantly. This increase in discharge rate then leads to a significantly higher readmission rate for patients who are discharged on days with high utilization. We show how to improve bed management and scheduling of surgeries. This leads to increased throughput and improves the quality of outcomes for patients.

In Chapter 4, we show that the time of day that a patient arrives at a trauma department affects the quality of care the patient will receive. Hospitals are less busy overnight. This should lead to a shorter waiting time for the patient, and better outcomes, as trauma care is typically time sensitive. However, staffing levels differ between night and day. Typically, there are fewer specialists at the hospital overnight, especially at smaller hospitals. This means that patients treated overnight often receive care from less specialized doctors. On average, patients receive care more promptly overnight. We show that patients who arrive overnight have significantly worse outcomes than patients arriving during the day. The biggest difference in outcomes is at small hospitals and among patients with more complex injuries. Based on the differences in outcomes, we conclude that resource management has a significant impact on the quality of care that patients receive. Specifically, the lack of specialized resources available at the hospital overnight leads to higher patient mortality and longer lengths of stay.

Academic hospitals have the dual mandate of treating patients while also educating the next generation of physicians. This is handled by having resident physicians (doctors in their first three to six years out of medical school) treat patients while being supervised by more experienced attending physicians. This presents interesting tradeoffs for hospitals. On one hand, residents treat patients directly and provide care. However, they also require supervision from attending physicians, which reduces the amount of time that attending physicians can spend treating patients. There are also concerns about the quality of care delivered by residents. In Chapters 5 and 6, we present research work that determines the effect residents have on efficiency in the emergency department of a large academic hospital. We show that residents help to increase throughput and decrease treatment times, especially when treating patients determined by the triage nurse to be high severity. This is especially important in the emergency department setting because prompt treatment is essential for many patients. Lowering treatment times help to decrease waiting times as well.

In Chapter 7, we discuss a long-range triage tool developed with internal medicine doctors at the University of Maryland Medical Center (UMMC). Doctors at UMMC noticed that patients arriving by inter-hospital transfer were in worse medical condition and more likely to require a stay in the intensive care unit (ICU) or die than patients arriving via the emergency room or by appointment. The doctors at UMMC did not completely trust the assessments given by the referring doctors. We develop a tool to predict patient risk using objective factors such as blood pressure, heart rate, and blood levels. This tool will help UMMC doctors determine which patients to accept and to anticipate what level of care patients will require.

In Chapter 8, we examine how MRI data can be used to diagnose and locate prostate cancer. Using MRI data from prostates, we develop three classification algorithms. We show that MRI data can be used to predict which prostates have cancer. In addition, we predict where in the prostate cancer will occur. In Chapter 9, we summarize the results of our research and present directions for future research.

CHAPTER 2: EXAMINING THE DISCHARGE DECISIONS OF SURGEONS

2.1 Introduction

Given that the structure of the healthcare system in the United States rewards the volume of specialty services provided, surgical volume tends to be the primary driver of hospital revenues and profits in most large hospitals. Profits from surgical services are used to cross-subsidize less profitable, but vital parts of hospital operations. Surgeons derive a large portion of their personal income from the surgeries they perform, and they make more money by doing more surgeries. Therefore, surgeons and hospitals want to ensure that as many surgeries as possible are performed on a daily basis.

While an operating room is needed to perform surgery, a downstream bed is required for the patient to recover. Immediately following surgery, patients move to the post-anesthesia care unit (PACU) where they spend time (one or more hours) recovering from the anesthesia. These patients might then require time in a specialty unit (one or more days), such as an intensive care unit (ICU), an intermediate care (IMC) unit, or an acute care unit. As a patient's condition improves, he/she transitions to other hospital units where care is tailored to meet his/her changing needs.

If the hospital does not have sufficient downstream bed capacity on the day of surgery, surgical cases are either cancelled or delayed, thereby creating problems for hospital staff and patients. It is in the surgeon's interest and the hospital's interest to ensure that there is capacity available on days when the surgeon is scheduled to perform surgeries. There is already some evidence in the literature to suggest that there are more discharges on days when the surgeons have scheduled surgeries (Price et al. 2007). Because the operating schedule typically does not take into account the future occupancy of the hospital, surgeons could ensure bed availability by adjusting the length of stay for their patients. For example, if the hospital's post-operative beds are full on a given day, a surgeon might discharge patients earlier, in order to make room for the surgeries scheduled on that day. An early discharge from the ICU may result in a patient returning to the ICU after moving to a step-down unit because the patient was not ready for the lower level of care. An early discharge of a patient from the ICU can also lead to an increase in the stress and workload of the individuals who staff the downstream bed units because they are caring for sicker patients who require more intensive care. Similarly, an early discharge from the hospital can lead to the patient being readmitted to the hospital. An early discharge may also lead to a worse overall health outcome for the patient. Therefore, it makes sense to examine the issue of early discharges carefully.

In this chapter, we examine a year's worth of discharge data from 2007 to determine whether the availability of recovery beds has any effect on the discharge rates. We hypothesize that as utilization increases, discharge rates increase. The data set was obtained from a large, academic, tertiary-care medical center located in the United States, with over 300 post-operative beds. Many different types of surgeries are performed at this hospital, ranging from plastic surgery to brain surgery. With the exception of cardiac surgery patients, patients exiting an operating room require a brief stay in the PACU prior to either being discharged home or transferred to an inpatient unit. The destination unit can be designated by level of care (ICU, IMC, acute), specialty (cardiac care, neuro care), or both. Lack of bed availability in these inpatient (downstream) units can lead to patient flow bottlenecks in the PACU. This can result in an inability to move patients out of the surgical arena, which increases the likelihood that surgical cases scheduled later in the day will have to be postponed. We focus our research on the effect of ICU utilization on discharge rates.

In Section 2.2, we review the literature on patient length of stay. In Section 2.3, we examine the data set and the methods used in our analysis. In Section 2.4, we propose survival analysis methods to determine if downstream bed availability influences the discharge rate for patients. Our results are presented in Section 2.4. Additional modeling is described in Section 2.5. In Section 2.6, we discuss the implications of our results and give our conclusions.

2.2 Background

Several papers in the healthcare literature have focused on the problem of detecting and explaining day of week variations in length of stay and the volume of discharges. In addition, researchers have shown a relationship between medical decision making (discharge and admittance) and utilization. For example, Singer et al. (1983) and Strauss et al. (1986) discuss the rationing of intensive care unit beds to maintain the flow of post-operative patients in a hospital. Singer et al. study a situation where there was a lack of beds due to a nurse shortage. In this case, they find that utilization was increased, admissions decreased, and average patient severity was higher. Strauss et al. study normal operations in an ICU, and conclude that when bed utilization is high, there is an increased discharge rate, using standard *t*-tests. The main difference between our work and theirs is that, while their work only examines patients and utilization at the time of discharge, our work observes each patient for their entire stay in the hospital. By doing this, we are able to build survival curves, determine a variety of effects on discharge rates, and control for other confounding variables. Price et al. (2007) use data mining methods and survival analysis to predict cardiac ICU availability a few days in advance. They develop and test their model on historic length of stay data, and find evidence that the availability predictions are systematically worse on days with an above-average volume of scheduled cases. This work implies that external factors, such as the surgical schedule, impact the treatment of individual patients in predictable ways.

There has also been work examining how utilization concerns affect the ability of surgeons to perform surgery. McManus et al. (2003) discover that a decrease in bed availability was caused by scheduled admissions and not emergency arrivals. They find the variability in the hospital census caused by the scheduled cases is much larger than the variability caused by emergency admissions. Thus, patient flow can be improved by considering system-wide efficiency when making scheduling decisions.

Locker and Mason (2005) validate the use of survival analysis techniques for modeling length of stay in a medical setting. We use similar methods to explain variation in discharge practices. Millard et al. (2001) examine bed occupancy levels in geriatric wards. We consider occupancy levels in post-operative beds.

While there has been research on predicting the length of stay for a patient, and modeling utilization over time, we are not aware of any work that examines the interaction between these two processes. In this chapter, we address how utilization levels affect discharge rate and length of stay. In particular, we find that utilization has a significant effect on discharge rates. Examining utilization could help to improve predictions of a patient's length of stay. Furthermore, it may be that an increased discharge rate at high utilization levels leads to higher hospital readmission rates (see Section 2.6).

2.3 Data and Methodology

The hospital provided surgical discharge data from the 2007 fiscal year (July 1st 2006 to June 30th 2007). During this year, there were no major changes in operating room procedures or scheduling. The information included patient age, surgical severity level, and the surgical specialty group which performed the surgery. In addition, we were provided information on the date and time of the surgery, and the dates when the patient was admitted to the recovery ward and discharged. The data contained information on 7,808 patients, of whom 6,470 were admitted for at least one day (i.e., one overnight stay). These 6,470 inpatients stayed a total of 35,478 days in the hospital (this gives us 35,478 observations for our modeling effort). From this data set, we derived the number of recovery beds that were utilized at the start of any day in the ward.

An initial examination of the data indicated that there might be an increase in the discharge rate when utilization is high. We examined how the average discharge rate varied with changes in downstream bed utilization. Figure 2-1 shows a bar chart with the average discharge rates for different ranges of downstream bed utilization. The Pearson correlation is small (r = .004), but, as utilization increases above 93%, the discharge rate also increases. We do not initially see a strong correlation between utilization and discharge for a number of reasons. Discharge rates are affected by numerous factors (e.g., age, severity, surgery type, and health of the patient), and there is a strong cyclical nature to utilization patterns which confounds the relationship between utilization and discharge rates. Looking at the aggregate correlation number masks variation due to patient-specific characteristics. While the discharge rate is not monotonically increasing, we do see an upward trend in the discharge rate as utilization increases. There is some noise, but the chart shows a positive correlation between discharge rate and utilization. While this chart does not prove that there is a relationship between utilization and discharge rate, it does motivate further exploration.

We study the relationship between downstream bed utilization and the rate at which patients are discharged using survival analysis. Survival analysis is a branch of statistics that deals with modeling time to event data. In our case, we are interested in modeling the time a patient spends in the recovery ward before being discharged. Traditional statistical methods do not handle survival data well because there is autocorrelation in the response variable among each subject. If the event (i.e., discharge in our case) happens on the N^{th} day for a given patient, the event could not have happened on any of the previous N-1 days, inducing correlation into the sequence of observations on the same individual. Traditional statistical models assume independence among the observations which can lead to inefficiencies and bias in the model estimates. However, survival analysis methods are able to take the correlation into account and give statistically valid results.

The most common model used for survival analysis is the Cox proportional hazards model (Cox, 1972). The Cox model estimates the rate at which an event will occur as a function of given predictor variables. This rate is called the hazard function. We are interested in how long patients spend in the hospital before being discharged. In particular, we are interested in what effect downstream bed utilization has on the hazard function. One drawback of the Cox model is that it assumes that the time until an event occurs is continuous. In our data set, time is discrete. While this is not a large problem—all data are discrete at some level—assuming continuity is not correct. More importantly, the Cox model assumes that all predictor variables have the same value for the entire length of each observation. The Cox model cannot handle variables that vary with time, such as bed utilization. Our variables of interest change over time, so this is a serious modeling obstacle. Because the Cox model cannot handle variables that change from day to day, we need to use a different model.

Singer and Willet (1993) showed how to handle discrete time survival data. When each time interval has the same length, a modified logistic regression model can be used to estimate the hazard function. Because this model handles discrete time data, the hazard function now measures the probability of an event occurring in a given time period, instead of the rate at which events occur. Logistic regression is used to estimate the odds that an event occurs during any given time period. The dependent variable in the Cox model, time until discharge occurs, is transformed into one observation for each day that the patient was in the hospital. Because we now have an observation for each day, we can handle variables that change from day to day.

We want to determine if increased utilization of recovery beds increases the discharge rate. First, we must define a suitable measure for increased utilization. We define two different measures for our data: a dichotomous measure that is 1 when utilization is above a certain threshold, and a continuous measure which counts the number of filled beds.

To test the conjecture that higher recovery bed utilization leads to a higher discharge rate, we define a variable, denoted by *FULL*, that equals 1 when utilization is over a certain threshold and 0 otherwise. To test the sensitivity of our model to the choice of the threshold, we varied the threshold from 80%

utilization to 95% utilization. In addition to a dichotomous variable, we also investigate the incremental impact of each available bed using a discrete variable, denoted by *BEDS*, that measures the total number of full recovery beds on a specific day. We predict that an increase in *BEDS* will lead to an increase in the discharge rate.

While patients are moved to different ICUs based on surgery type, there is some flexibility in the assignment of patients to ICUs. We use the total number of beds that are full as our utilization measure, instead of the number of beds in use in each unit. While not all of the ICUs are completely interchangeable, there is enough flexibility in where any one specific patient will be placed for our measure to make sense. Each ICU is much smaller than the entire ward and has less than 20 beds, often less than 10 beds. The number of beds in use in a surgical line's specific ICU captures less information than the utilization of the entire ward in general. In addition, each specific ICU's staffing levels are more volatile than the aggregate staffing level, and each individual patient might be assigned to any of several specific ICUs depending on his/her characteristics. For example, a patient undergoing orthopedic surgery might be placed in the trauma surgery's ICU if the orthopedic ICU is full. Furthermore, the overall utilization and the utilization of each specific ICU are highly correlated. By aggregating overall utilization, we are able to account for the flexibility in which ICU the patients is assigned to, and capture more information about the utilization of the whole ICU (i.e., how full is the ICU?). As a result, we feel that overall utilization is a better measure than measuring individual ICU utilization.

We transformed our length-of-stay data into observations that can be tested using the Singer-Willett model by creating a variable denoted by *DISCHARGE* that equals 1 if a patient was discharged on a given day, and 0 otherwise. We control for the urgency of the surgery and sickness level of the patients with a variable, denoted by *ELECTIVE,* that is 1 if the surgery was elective and 0 otherwise. For each day, we calculated *DISCHARGE*, *FULL*, *BEDS,* as well as recorded the patient's age, surgical line, and severity level. Table 2-1 summarizes the variables used in our models.

2.4 Analysis and Results

Using the data set of 35,478 observations, we constructed three different Singer-Willett regression models to determine whether a decreased supply of recovery beds increases the probability that a patient will be discharged. One model tests the range where *FULL* has a statistically significant effect. A second model extends the first model by including controls variables for each surgical group. A third model uses the continuous variable, *BEDS*, instead of *FULL*. By using a continuous variable to measure utilization instead of a threshold, the third model measures the impact of each additional occupied bed.

In the first model, we regressed the *DISCHARGE* on *AGE*, *ELECTIVE,* *FULL,*and 59 daily dummy variables. This model is given by

$$DISCHARGE = AGE + ELECTIVE + FULL + D1 + D2 + \ldots. + D59 + \varepsilon.$$

In order to investigate the sensitivity of our results to the choice of the threshold for *FULL*, we varied the threshold between 80% and 95% bed utilization. Below 90% utilization there was no effect for *FULL*. There was only one day in our sample where utilization at the start of the day was above 96%, so the sample size was too small to perform any meaningful analysis. We ran the same test at each level, and recorded the magnitude of the coefficient for *FULL* and the standard error for the estimate. Table 2-2 shows these results. Figure 2-2 shows how the magnitude of the coefficient for *FULL* varies as the threshold increases. On the *x*-axis, we show the threshold above which *FULL* is defined to be 1. The *y*-axis gives the magnitude of the coefficient for *FULL* in the regression models. The dashed lines show one standard deviation above and below the estimate for each point. The dip at 0.94 is not statistically different than the point at 0.93. The graph stops at 0.94 because the sample size decreases dramatically past 94% utilization, the estimates grow extremely noisy, and the standard errors become very large. We found statistically significant coefficients ($p < .05$) for *FULL* when the threshold is above 91.5%. Ten percent of days had 91.5% utilization or higher, so days in the highest decile of utilization had an increased discharge rate. Table 2-3 shows the output from the regression model, when the threshold for *FULL* is 93%.

We see statistically significant coefficients on *FULL* when the threshold is above 91.5%. This implies that, when there is high utilization in the recovery ward, and hence a chance that some surgeries will have to be rescheduled, the probability that each patient will be discharged increases. The coefficient for

FULL is also large enough to have an observable impact on discharge practices. Figure 2-3 shows the magnitude of this effect for a typical patient (45 year old, elective surgery). The graph plots the percent of patients remaining in the hospital versus the number of days in the recovery wing. The circles are for patients when *FULL* is 1, while the squares are for the same patients when *FULL* is 0. The distance between the two curves is the number of patients who would be discharged when there is high recovery bed utilization but would be in the hospital when utilization is low. The area between the two curves is the total number of bed days freed up by the effect of *FULL*. The maximum difference between the curves is more than 15 beds out of the 320 in the recovery ward. Almost 48% of patients in a recovery ward that is not full (utilization < 93%) will be in the hospital after six days, compared to only 43% in a ward that is full (utilization > 93%).

In the next model, we control for the surgical group performing the surgery. There is some information captured in the surgical line variable that is not contained in *ELECTIVE*. A patient undergoing elective brain surgery might be expected to spend longer in the recovery ward than a plastic surgery patient undergoing a non-elective procedure. We add a dummy variable for each different surgical line to the previous regression model. For example, the variable *CARDIAC SURGERY* is defined to be 1 if the patient underwent cardiac surgery and 0 otherwise. This model is given by

$$\textit{DISCHARGE} = \textit{AGE} + \textit{ELECTIVE} + \textit{FULL} + \textit{CARDIAC SURGERY} + \textit{CARDIOLOGY} + \ldots + \textit{DONOR SERVICE} + D1 + D2 + \ldots. + D59 + \varepsilon.$$

The results from this model are given in Table 2-4. Again, *FULL* has a statistically significant coefficient, with a *p*-value below .01. *FULL* has a statistically significant coefficient regardless of model specification. Adding surgical group control variables had little effect on the magnitude or statistical significance of the *FULL* parameter. This shows that the variable *FULL* has an underlying effect on discharge rate, and is not an artifact of the specific statistical model.

Our third model uses a continuous variable to capture the effect of decreasing recovery bed supply on discharge decisions. Instead of using *FULL* in our model, we use *BEDS*. By treating the data in this way, we now are examining whether each additional filled post-operative bed increases the discharge hazard, instead of there being some threshold above which the discharge rate increases. This model is given by

$$DISCHARGE = AGE + ELECTIVE + BEDS + CARDIAC\ SURGERY + CARDIOLOGY + \ldots + DONOR\ SERVICE + D1 + D2 + \ldots. + D59 + \varepsilon.$$

The results from this model are given in Table 2-5. The coefficient for *BEDS* is positive and statistically significant, with a *p*-value of .02. The magnitude of the coefficient is smaller than the coefficients for *FULL* in models 1 and 2. This is because the range of *BEDS* is in the hundreds, while *FULL* can only be 0 or 1. When accounting for the range of *BEDS*, the total magnitude of the effect is similar to the magnitude of the effect for *FULL*. This shows that as recovery beds fill up, and supply becomes tight, the probability of discharge increases, regardless of how it is measured.

2.5 Additional Modeling

While we have shown that discharges happen at a higher rate when the ICU is full, there is a concern about the cyclical nature of utilization. In Figure 2-4, we see that the ICU tends to fill up over the course of the week and empty out over the weekend. Therefore, *FULL* has a value of one mostly on Thursday and Friday, and a value of zero mostly on Saturday through Wednesday. It could be that, instead of patients being discharged faster because the ICU is full, they are more likely to be discharged on Friday. Because *FULL* is more likely to have a value of one on Friday, we would attribute the increase in discharge rate to *FULL* rather than the day of the week. In other words, it could be that the effect of *FULL* is confounded by the day of the week. To account for this, we add a dummy variable for each day of the week. If there are more discharges on Friday, the dummy variable will capture this situation.

The hospital is not always staffed to full capacity, so the limiting factor in space available for patients is not always the number of physical beds, but the number of staffed beds. To account for differences in staffing levels, we define a new variable, denoted by *FULL2,* that is 1 when the number of beds in use is more than 97% of the most beds ever used on that day of the week. We chose a threshold of 97% because this represents the 10% of days with highest utilization.

We also examine whether or not surgical volume has an effect on the discharge rate. If there are more surgeries on a given day, more space will be needed for those incoming patients, and there might be more discharges. We include two variables in our model, denoted by *TODAY* and *TOMORROW*, that measure the number of surgeries scheduled for the current day and the next day, respectively. By including both the supply of beds (utilization) and the demand for beds (number of surgeries) in our model, we hope to determine what effect each variable has on the discharge rate. When supply is low (high utilization), the discharge rate increases, but we also want to look at how the demand for beds (number of surgeries) impacts discharge rate. If there is an average number of beds available, but a large number is needed due to high surgical volume, then the discharge rate might be increased to make room for incoming patients. This model is given by:

$$DISCHARGE = AGE + ELECTIVE + BEDS + FULL2 + TODAY + TOMORROW + MONDAY + TUESDAY + \ldots + SUNDAY + D1 + D2 + \ldots. + D59 + \varepsilon.$$

The results of this model are given in Table 2-6. First, we observe that the coefficient for *FULL2* has a positive sign and is statistically significant ($p = .041$). After controlling for staffing levels, day of week, and surgical schedule, patients are still discharged at a higher rate when the ICU is fuller. Second, the coefficients for *TODAY* and *TOMORROW* are both positive and statistically significant. This shows that doctors take both the state of the ICU and the future surgical schedule into account when making discharge decisions. Third, none of the day of week variables have a statistically significant effect on the discharge rate. Patients are not more or less likely to be discharged on any given day, after controlling for other variables in the model.

We use two measures of model fit: pseudo *R*-squared and Aikake Information Criterion (AIC). Pseudo *R*-squared is analogous to the standard *R*-squared used in linear regression. AIC measures the amount of information lost by a model, with lower numbers being better. While the model in Table 2-6 has a better model fit (both in terms of pseudo *R*-squared and AIC), the improvement in both values (over the model in Table 2-5) is very small. Controlling for additional factors such as day of the week, staffing levels, and surgical volume does not explain much of the variability in discharge decisions beyond what is captured by the model in Table 2-5.

2.6 Discussion and Conclusions

The results of our models suggest that surgeons discharge patients, when needed, to ensure that their surgeries will not be cancelled due to a lack of recovery beds. We have shown that surgeons discharge patients earlier when there are relatively few downstream beds available. This effect is observable and statistically significant, regardless of model specification. This discharge practice is a source of artificial variability and should be taken into account when predicting patient length of stay and hospital bed capacity. Because the surgical schedule depends on the availability of recovery beds, these practices should also be taken into account when generating the surgical schedule. In essence, the scheduler's job is easier because the system can adjust to make more beds available via early discharge, when necessary. This should not be interpreted as an invitation to over-schedule surgeries, but rather a recognition that the bed management system is robust enough to adjust to occasional imbalances with respect to scheduled arrivals and expected discharges.

While this chapter argues that scarcity in the supply for beds increases the discharge rate, Price (2009) has shown that an increase in the demand for beds has a similar, smaller effect. By looking at a similar data set, Price showed that on days when there were more surgeries scheduled, there were more patients discharged. This demand-side argument nicely complements our supply-side argument that, as supply decreases, the discharge rate increases. Price makes the analogous demand-side argument that as demand increases, so does the discharge rate.

When researchers attempt to model and understand the flow of patients through a hospital, they typically do not take into account how physicians make decisions. Most people believe that the decision to discharge a patient is made independently of the state of the system (e.g., the surgeon's upcoming surgeries or the current number of patients in the ICU). Our research shows that surgeons discharge patients early, based on the impact to their future surgical schedule. This adds a dimension to any study or model that seeks to improve the flow of patients through the hospital. There are many papers in the open literature that use linear and integer programming techniques to improve surgery scheduling (Belien and Demeulemeester 2007, Blake et al. 2002). By omitting the effect of a physician’s discharge practices on the patient length of stay, researchers have overlooked an important factor that may affect hospital utilization.

While we have shown the effect of bed supply on discharge rate, we have not shown that this is a public health concern. It could be that, instead of discharging patients early when the recovery ward is full, surgeons otherwise keep patients an extra day or two to make sure they are fully recovered before being discharged. Our work cannot determine which of these two explanations is more accurate. In future work, we want to examine the outcomes of patients who were discharged early from a full recovery ward and determine whether or not these patients were more likely to be readmitted to the hospital.

Policymakers are increasingly concerned with issues related to the cost and quality of healthcare. Keeping patients longer in the ICU will increase costs. Discharging patients before they are ready to be discharged might lead to incomplete recovery. Furthermore, being discharged too soon increases the stress on the downstream units and may raise the risk of readmission to the ICU both of which could raise cost and decrease quality. Future work could look into the effects of discharge practices on readmission rates. For example, we would like to track individual physicians and monitor their decisions, as well as track the health outcomes of patients discharged from a recovery ward that is full, and compare the outcomes with patients discharged from a less-than-full ward.

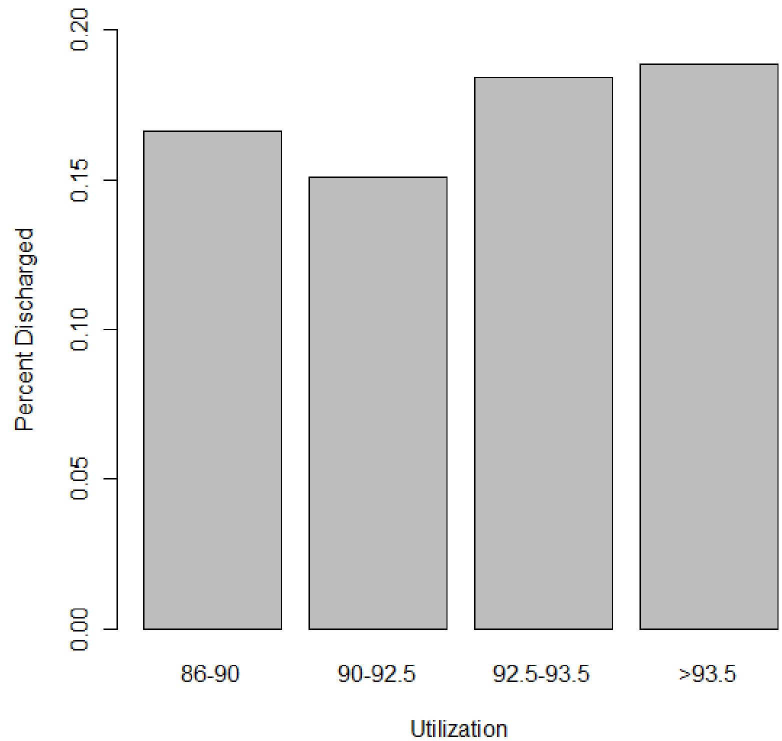

Figure 2-1: Bar chart of discharge rate for different bed utilization ranges.

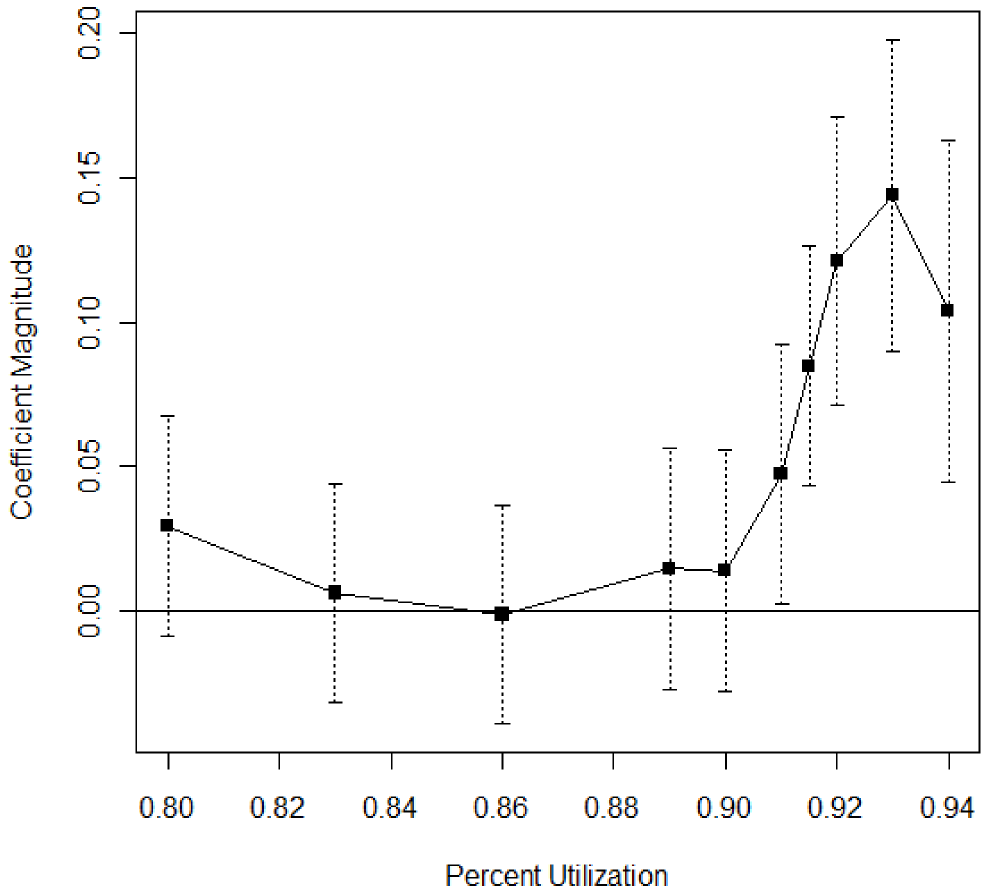

Figure 2-2 Effect of FULL vs. the bed utilization rate.
The *y*-axis gives the magnitude of the coefficient for *FULL* given that *FULL* is defined to be 1 at any bed utilization rate greater than X. Dashed lines show one standard error. We find statistically significant effects when the threshold is above 91.5% utilization.

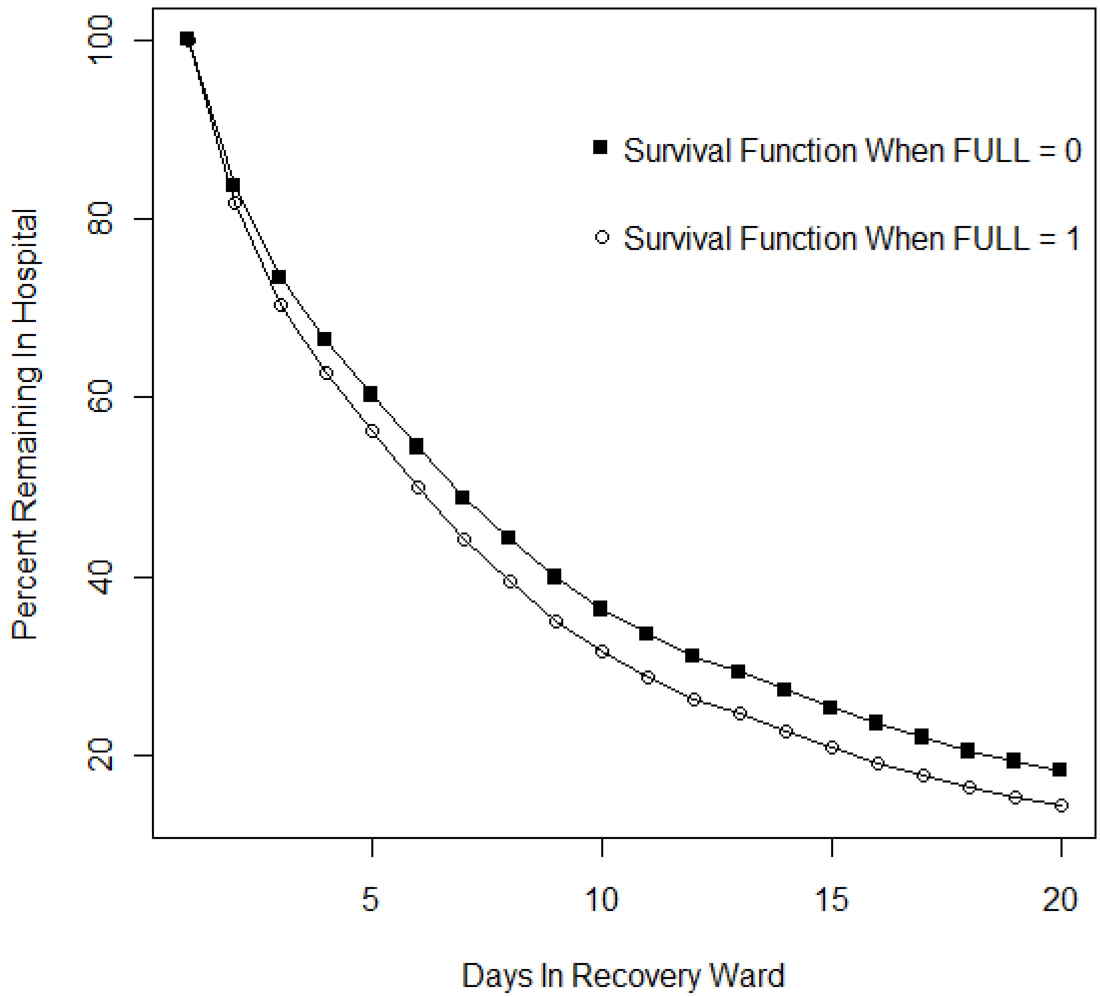

Figure 2-3: Survival Rate vs. Time in Hospital with FULL = 1 and FULL = 0

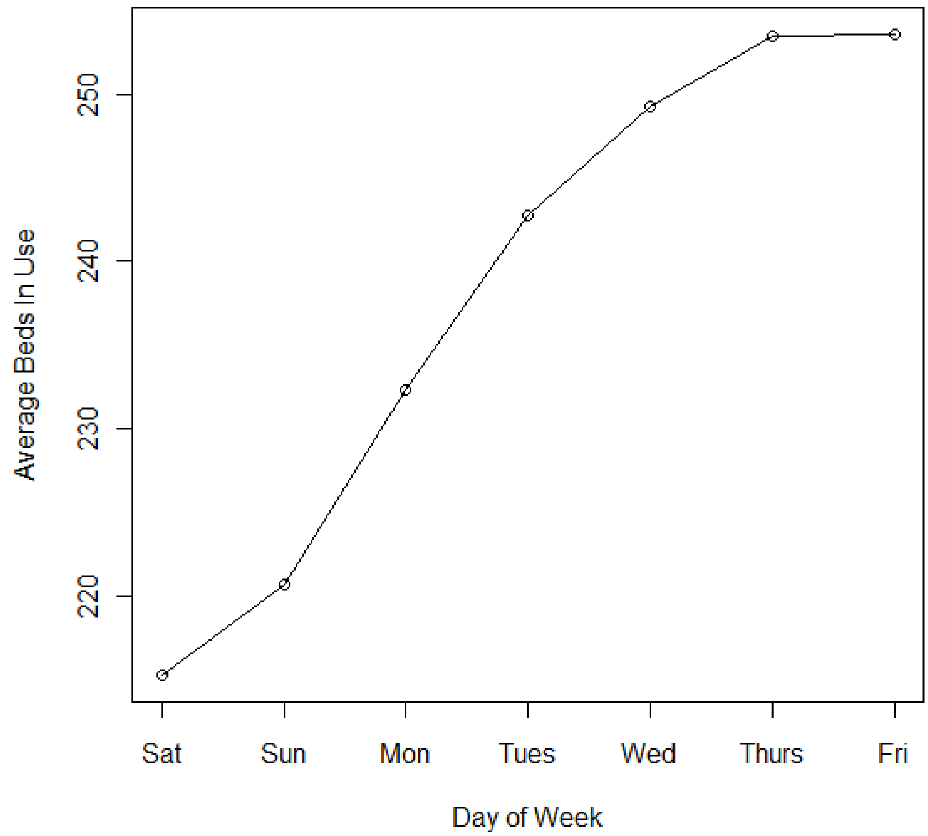

Figure 2-4: Average utilization of beds on each day of the week

Table 2-1: Descriptions of the response and predictor variables

Variable Name	Description	Range
DISCHARGE	The dependent variable. It is 0 for every day that the patient remains in the hospital, and 1 on the day that the patient is discharged. For each observation of Discharge, the six variables below are calculated.	[0,1]
AGE	The age, in years, of the patient on the day of the surgery.	[1,96]
ELECTIVE	A dummy variable that is 1 if the surgery was classified as elective and 0 if it was not.	[0,1]
BEDS	A time-dependent variable that measures the number of recovery beds filled at the start of each day. This variable changes over the course of each patient's stay in the hospital.	[120,320]
FULL	A time-dependent dummy variable that is defined to be 1 on days when the number of filled beds is above a certain threshold, and 0 otherwise. It can change over the course of a patient's stay.	[0,1]
D1, D2, …., D59	Dummy variables for each day. D_N is defined to be 1 if it is the patient's N^{th} day in the hospital, and 0 otherwise. There are 59 variables because the longest stay in our data set was 60 days. One must be omitted to avoid multicollinearity.	[0,1]
CARDIOLOGY, …, DONOR SERVICE	Dummy variables for each service line. They are defined to be 1 if the surgery was performed by that surgical line, and 0 otherwise. There are 23 different service lines in our data set.	[0,1]

Table 2-2: Effect of the threshold definition on the magnitude and significance of the *FULL* parameter

Utilization Threshold for *FULL*	Magnitude of the *FULL* coefficient	Std. Error	*z* value	*p*-value
0.8	0.0295	0.038	0.776316	0.44
0.83	0.00632	0.038	0.166316	0.86
0.86	-0.00078	0.038	-0.02053	0.98
0.89	0.0147	0.042	0.35	0.72
0.9	0.014	0.042	0.333333	0.74
0.91	0.0474	0.045	1.053333	0.29
0.915	0.0849	0.0415	2.045783	0.061
0.92	0.121	0.0499	2.42485	0.024
0.93	0.144	0.054	2.666667	0.0199
0.94	0.104	0.059	1.762712	0.078

Table 2-3: Utilization threshold survival model
(AIC = 32619, pseudo *R*-squared = .3393)

	Estimate	Std. Error	*z* value	*p*-value
AGE	-9.43E-03	6.60E-04	-14.299	< 2e-16
ELECTIVE	4.33E-01	7.68E-02	5.634	1.76E-08
FULL	1.24E-01	5.35E-02	2.327	0.019967
D1	-9.93E-01	8.40E-02	-11.831	< 2e-16
D2	-1.33E+00	8.73E-02	-15.197	< 2e-16
D3	-1.60E+00	9.11E-02	-17.56	< 2e-16
. . .	. . .	. . .	. . .	. . .

Table 2-4: Utilization threshold model with surgical group control
(AIC = 32619, pseudo *R*-squared = .3623)

	Estimate	Std. Error	*z* value	*p*-value
AGE	-9.39E-03	8.15E-04	-11.529	< 2e-16
ELECTIVE	4.04E-01	7.81E-02	5.171	2.33E-07
FULL	1.45E-01	5.49E-02	2.63	0.008527
D1	-7.76E-01	1.04E-01	-7.466	8.28E-14
D2	-9.53E-01	1.06E-01	-8.965	< 2e-16
D3	-1.15E+00	1.10E-01	-10.506	< 2e-16
. . .	. . .	. . .	. . .	. . .
CARDIOLOGY	-7.46E-01	4.81E-01	-1.553	0.120539
CARDIAC SURGERY	-8.73E-01	8.21E-02	-10.642	< 2e-16
. . .	. . .	. . .	. . .	. . .
DONOR SERVICE	2.03E+00	1.16E+00	1.747	0.080553

Table 2-5: Continuous utilization model
(AIC = 32619,   pseudo *R*-squared = .3393)

	Estimate	Std. Error	*z* value	*p*-value
AGE	-9.43E-03	6.60E-04	-14.291	< 2e-16
ELECTIVE	4.34E-01	7.68E-02	5.65	1.61E-08
BEDS	6.73E-04	2.89E-04	2.33	0.01979
D1	-1.16E+00	1.12E-01	-10.329	< 2e-16
D2	-1.49E+00	1.15E-01	-12.958	< 2e-16
D3	-1.76E+00	1.17E-01	-15.01	< 2e-16
. . .	. . .	. . .	. . .	. . .

Table 2-6: Day of week and surgical schedule model results
(AIC = 32202, pseudo R-squared = .3496)

	Estimate	Std. Error	z value	p-value
AGE	-9.71E-03	6.64E-04	-14.628	< 2e-16
ELECTIVE	4.11E-01	7.73E-02	5.32	1.04E-07
FULL2	1.23E-01	6.05E-02	2.037	0.04165
TODAY	5.55E-03	2.01E-03	2.761	0.00576
TOMORROW	4.55E-03	2.26E-03	2.018	0.04363
BEDS	-2.32E-03	3.94E-04	-5.897	3.69E-09
SATURDAY	-1.57E+01	8.83E+02	-0.018	0.98581
SUNDAY	-1.52E+01	8.83E+02	-0.017	0.98631
MONDAY	-1.51E+01	8.83E+02	-0.017	0.98636
TUESDAY	-1.48E+01	8.83E+02	-0.017	0.98665
WEDNESDAY	-1.49E+01	8.83E+02	-0.017	0.98658
THURSDAY	-1.45E+01	8.83E+02	-0.016	0.9869
FRIDAY	-1.48E+01	8.83E+02	-0.017	0.98663
D1	1.39E+01	8.83E+02	0.016	0.98743
D2	1.36E+01	8.83E+02	0.015	0.98769
D3	1.34E+01	8.83E+02	0.015	0.98788
. . .	. . .	. . .	. . .	. . .

CHAPTER 3: THE IMPACT OF HOSPITAL UTILIZATION ON PATIENT READMISSION RATE

3.1 Introduction

In Chapter 2, we showed that both higher post-operative bed utilization and higher demand for beds (more incoming surgeries) lead to increases in the discharge rate. We offered two explanations for this increase: either patients were being held *longer* than needed when beds were available, or they were being discharged *early* when beds were needed for incoming patients. While the discharge rate was increased by higher utilization, no determination between these two explanations could be made.

In this chapter, we examine how the utilization of beds on the post-operative path at the time of discharge affects the readmission rate among surgery patients. By looking at readmission rates, we try to determine which of the two explanations for the higher discharge rate is applicable. If patients are discharged too soon when utilization is high, we would expect to see an increase in the readmission rate. A patient who is still recovering will be more likely to return to the hospital than one who is ready to be discharged. However, if patients are simply being held longer until space is needed, we should not see any effect on readmissions. Figure 3-1 illustrates the hypothesis that we test in this chapter. We are confident in the direction of the causal arrows, because utilization precedes readmissions. When patients are readmitted to the hospital, they are not sent back to the post-operative unit, so they do not affect the utilization rate. Instead, they are either treated in the emergency department, or in a bed elsewhere in the hospital. In this chapter, we examine surgical data from a large urban teaching hospital in the United States.

In Section 3.2, we review the relevant literature with respect to hospital readmission. In Section 3.3, we describe our data, explain the methodology, and present our results. In Section 3.4, we mention the limitations of our work. In Section 3.5, we discuss the implication of our results and provide conclusions.

3.2 Literature Review

Our work focuses on studying post-operative readmission. We seek to show a connection between decreased length of stay due to early discharge and increased likelihood of readmission. While some studies have shown a relationship between length of stay and readmission, others have found no evidence for such a link. Hasan (2001) provides a short survey of work studying hospital readmissions. He concludes that while premature discharge has been proposed as a cause for readmission, no causal link has been shown. We focus on the effect of length of stay on readmission in this section for two reasons. First, we are not aware of any papers that study the effect of utilization on readmission. Secondly, we hypothesize that high utilization causes a decrease in length of stay.

As already mentioned, the literature is split on whether or not length of stay has an effect on readmission rate. It seems that length of stay does not have an effect on readmission rate, unless the length of stay is artificially shortened. Bohmer et al. (2002) study the readmission rate after coronary bypass surgery and find that there is no relationship between length of stay and readmission rate.

However, they show that cost savings from shorter stays are offset by increased use of post-acute services. Cowper et. al (2007) find that patients discharged early after coronary bypass graft surgeries do not have a higher readmission rate, and have an average cumulative savings of over $6,000. Delaney et al. (2001) observe that patients selected for a “fast track” discharge protocol are no more likely to be readmitted. These patients have shorter lengths of stay and lower costs, and, in addition, are not any more likely to be readmitted.

However, there is evidence that early discharge leads to higher rates of readmission. Niehaus et al. (2008) conclude that patients discharged because of bed shortages in psychiatric hospitals are significantly more likely to be readmitted. Campbell et al. (2008) use length of stay as a significant predictor of readmission after discharge from an intensive care unit. Hwang et al. (2003) find that patients who disregard medical advice and leave the hospital early are significantly more likely to be readmitted within 15 days. Dobson et al. (2011) propose a model of ICU bumping. They model physicians’ decisions on which patients to discharge early when bed space in the ICU is limited. They find that the surgical schedule influences physicians’ decisions to bump patients. They conclude that under some circumstances it is optimal, in terms of throughput, for surgeons to discharge patients before it would be medically advisable.

3.3 Data and Analysis

We were provided surgical data by a large urban east coast U.S. academic hospital on every surgery patient during the first half of 2007 (January 1, 2007 to June 1, 2007). During that time period, there were no major changes in operating room procedures or scheduling. The dataset contains information on patient age, surgical schedule type (emergency vs. elective), and the surgical specialty group that performed the surgery. Our data contains a total of 5,265 (adult) patients admitted as inpatients. We also have available the date and time of the surgery, the dates when the patient was admitted to and discharged from the hospital, and the dates when the patient was readmitted, if any. Using this dataset, we calculated the number of post-operative beds that were utilized at the start of each day during the fiscal year. The hospital has a dedicated post-operative unit that has surgical ICU beds, acute care beds, and intermediate care beds.

We focus our analysis on differences in readmission rates within 72 hours because we expect the effect that utilization has on readmission to be most prominent in the first few days. The 72-hour cutoff is important because a patient readmitted within 72 hours of discharge must be treated as part of inpatient services and billed as one claim, thereby reducing the amount the hospital makes from the procedure. If a patient is discharged before he/she is ready to be sent home, we expect that he/she will be more likely to be readmitted sooner rather than later. While an early discharge might lead to more complications down the road, it might also lead to more complications in the short term. We will examine readmission rates in every time frame up to 30 days after discharge, but our main focus is on rates within 72 hours. In this study, we only address readmission after discharge from the hospital, not "bounce backs" when patients move back to an ICU from a step-down unit.

We examine how the percent of patients who were readmitted within 72 hours varies with changes in bed utilization upon discharge. When utilization was high (above 94%), 16.7% of patients were readmitted within 72 hours, compared to just 10% of those discharged when utilization was below 94%. A chi-square test comparing these two rates found that they were statistically different ($p < .0001$). After 30 days, 55% of patients discharged from a full (utilization >94%) post-operative unit were readmitted, while only 50% were readmitted of those discharged when utilization was lower. We initially choose 94% as the cutoff for high utilization because previous work (Anderson et al. (2012)) found an increased discharge rate when utilization crossed that threshold. Table 3-1 shows the average readmission rates within 72 hours for different ranges of downstream bed utilization. As utilization increases, the readmission rate increases, especially at the highest range of utilization. Figure 3-2 shows the relationship between discharge rate and readmission. As utilization increases, the discharge rate (shown by triangles) and the readmission rate (shown by squares) both increase. The discharge rate shown here is the percent of patients discharged each day.

We construct four logistic regression models to study the relationship between readmission rate and occupancy level at the time of discharge. We use a dichotomous dependent variable, i.e., whether or not a patient is readmitted within 72 hours of discharge from the hospital. To determine if increased utilization of post-operative beds at the time of discharge increases the readmission rate, we need to define a suitable measure for increased utilization.

We use two different variables: an indicator variable (called *FULL)* that is 1 if utilization is above a certain threshold when the patient is discharged and 0 otherwise, and a continuous variable (called *BEDS*) that gives the number of filled beds when the patient was discharged. We use four models because we have two different utilization metrics (continuous and discrete), and we examine readmission in the entire post-operative ward as well as just in the trauma surgical line. Using these four models allows us to study the relationship between utilization and readmission rates in more depth. Table 3-2 summarizes the variables used in our models.

In our first model (Model 1), we regress readmission within 72 hours on *FULL*. We control for the patient's age, race, gender, and the type of surgery. By controlling for other determinants of readmission, we are able to isolate the effect of being discharged from a full unit. We are confident that our control variables are relevant, not only practically but also statistically. Each variable lowers the Akaike Information Criterion (AIC) (see Akaike, 1974)) when added to the model, and is included in the final model when a stepwise model selection procedure is run. The equation for this logistic regression model is given by

$$\begin{gathered} logit(READMISSION72) = AGE + BLACK + ASIAN + HISPANIC + FULL + \\ ELECTIVE + TRANSPLANT + TRAUMA + \ldots + NEURO + MALE + \varepsilon \,. \end{gathered}$$

The baseline case, in which all dummy variables are 0, is a white, female patient in the general surgery line. The results from this model are presented in Table 3-3. We do not exclude any insignificant predictors since our sample size is very large and our overall goal is to measure the varying degrees of relationship between response and predictors. The effect of *FULL* is statistically significant at the 1% level and indicates that a patient discharged from a full unit is more likely to be readmitted within 72 hours. When *FULL* is 1, it increases the odds by a factor of 2.341. This implies that for a baseline patient of average age (i.e., 46 years), the probability of readmission increases from 10% to 20%.

We observe expected results from the control variables. For example, trauma and transplant patients are significantly more likely to be readmitted compared to the (baseline) general surgery patients. On the other hand, elective patients are less likely to be readmitted pointing to the lesser severity of elected surgeries. We control for a patient's demographics (age, race and gender). We find that male patients are more likely to be readmitted and that readmissions are significantly higher for black patients.

The second model (Model 2) tests readmission within 72 hours, using the continuous utilization variable, *BEDS*. We regress the readmission variable on the number of beds in use at the time of discharge (*BEDS*). This model allows us to quantify the effect of each additional occupied bed on the likelihood of readmission, as opposed to the effect of crossing a particular threshold. The hypothesized regression equation is given by:

$$logit(READMISSION72) = AGE + BLACK + ASIAN + HISPANIC + BEDS$$
$$+MALE + ELECTIVE + TRANSPLANT + TRAUMA + \ldots + NEURO + \varepsilon.$$

In this regression, *BEDS* measures the number of utilized beds at the time the patient was discharged. The results of this model are given in Table 3-4. The effect of *BEDS* is positive and significant at the 1% level. This tells us that an increase in utilization at discharge leads to an increased readmission rate. The magnitude of the *BEDS* coefficient (.00797) is smaller than that of *FULL* in the previous model (.851) because the range of *BEDS* is over 100. Each additional bed in use at the time of discharge increases the odds of readmission by a factor of 1.008. Table 3-5 shows the effect of increasing the number of beds in use at the time of discharge on the odds of readmission, and on the probability that a baseline patient will be readmitted. The probability of readmission is calculated at the mean age of a patient and with all dummy variables set to 0 (i.e., a 46-year old, white female patient in general surgery).

To further investigate the effect over time, we construct six additional models (all similar to Model 1), with readmission within 5, 10, 15, 20, 25, and 30 days as the dependent variable. For example, if a patient was readmitted after 13 days, the indicator variable would have a value of 0 for 5 and 10 days and a value of 1 for the other day variables. We recorded the magnitude of the coefficient of *FULL* for each model. In Figure 3-3, we show the magnitude of the coefficient of *FULL* as the definition of readmission changes. The error bars for each point show one standard error above and below each value. The utilization at discharge has the strongest effect on readmission within 72 hours, and it slowly diminishes over time. Patients who are discharged too soon are more likely to be readmitted quickly, instead of later in the month.

Next, we were interested in determining the effect when only one surgical line is considered. By isolating the surgical line, we can more precisely measure the effect of utilization. Concentrating on one surgical line only allows us to eliminate any variation between the discharge procedures of different surgical lines, the differences in case severity, and the potential for future complications. The *trauma surgery* line is one of the largest units in our dataset, and has patients who are similar demographically to the general surgery population at the hospital. The percentage of each racial group is nearly the same as the percentage in the general population. The mean and median ages for trauma patients are both within one year of the mean and median ages in the general population, as well. There are 248 patients admitted as inpatients after having trauma surgery. Of those, 18% were readmitted within 72 hours. The results are given in Tables 3-6 and 3-7.

Table 3-6 shows that high utilization upon discharge has a statistically significant effect on readmission rates within 72 hours. While *FULL* is statistically significant, its *p*-value is lower compared to Model 1 (Table 3-3). The reason is the reduced sample size (since we are only focusing on a single surgical line, the sample size is reduced by 95%), which reduces the power of the test. However, the magnitude of the odds ratio is increased dramatically (from 2.3 in the first model to over 25). This implies that for a baseline patient (a 46-year-old, white female) the probability of readmission increases by 63% (from 10% to 73%) – compare this to an increase of only 10% (10% to 20%) when averaging over all surgical lines (Model 1). The insight from this analysis is that the impact of utilization on readmission changes from one surgical line to another. We observe a much stronger effect for trauma patients because surgeries in this line are typically more severe and, thus, more sensitive to variations in length of stay and thoroughness of treatment. Lower acuity patients might be less sensitive to premature discharge, and, therefore, exhibit less of an increased risk for readmission.

The same conclusions hold when measuring utilization on a continuous scale (*BEDS*; see Table 3-7). In fact, we can see that the odds ratio of *BEDS* is 1.027, which is more than three times larger than its effect in Model 2 (which averages over all surgical lines). In other words, regardless of how utilization is measured, increased utilization at the time of discharge increases the readmission rate for trauma surgery patients.

Finally, we use survival analysis to determine what affects the rate at which patients return to the hospital after being discharged. While logistic regression estimates the probability that a patient will be readmitted in a certain time frame, survival analysis models the fraction of patients who have been readmitted over time. In this model, our dependent variable is whether or not the patient was readmitted on a given day. We create one observation for each day, for each patient, until the patient is readmitted, up to 30 days. If the patient is readmitted on that day, the variable is 1, if he/she is not, it is 0. A patient who is readmitted on the 10^{th} day after discharge will have nine observations where the dependent variable is 0. For the 10^{th} observation, the dependent variable will be 1, because the patient was readmitted on that day. We also create 30 dummy variables to account for the baseline hazard on each day (*D1, D2, … , D30*). *DN* is 1 on the N^{th} day after discharge, and 0 otherwise. In the first observation, *D1* is 1, and the remaining *DN* variables are 0. In the 10^{th} observation, *D10* is 1, the remaining *DN* variables are 0. When *FULL* is 1, the odds that a patient will be readmitted on any given day increase by a factor of 1.32. This means that when a patient was discharged from a highly utilized unit, the patient is readmitted at a higher rate. *FULL* has a larger effect on the odds of readmission than age, race, or gender. Figure 3-4 shows the percent of each type of patient who is readmitted as a function of time.

Patients discharged when the post-operative unit was full are more likely to return the first day, and the gap grows as the month progresses. These results are similar to the logistic regression models, in that patients dismissed from a full unit are more likely to return, and that the effect is visible immediately. In this model, the effect that utilization has on readmission rates is statistically significant throughout the entire month. When compared to the logistic regression models, we see a stronger effect later in the month in this model. The difference between the two models comes from a restriction of the survival analysis model. Survival analysis assumes that the effect of each variable is constant over time. However, by varying the readmission window in the logistic regression models, we can examine how the effect changes over time. The survival analysis model gives the effect of *FULL* averaged over the entire month. The coefficients on the *DN* variables are decreasing with time. This means that as the month progresses, patients are less likely to be readmitted. While many readmissions that occur after 72 hours are unavoidable, it is still interesting to note that the effect of utilization at time of discharge has a lingering effect on readmission rates that lasts up to 30 days.

3.4 Limitations

While our results shed new light on hospital readmissions, one should use caution in generalizing from them. Since our data pertain to one particular hospital only, our results don't immediately generalize to all hospitals. While we suspect that similar phenomena occur at other hospitals, since incentives in other U.S. hospitals are essentially the same, our study can only address one hospital. In addition, we also have very little data on patient acuity levels. We only have information on whether or not the surgery was elective or emergency. While this captures some of the variance in patient severity, we cannot control for all aspects of patient acuity levels. Also, we do not have the cause for patient readmission, which would allow us to examine in more depth the effect that ICU utilization has on readmission. While our main result holds, more detailed data would allow us to measure the effect of utilization more precisely.

3.5 Discussion

In previous work, Anderson et al. (2012) show that the discharge rate of patients in the post-operative unit increased when utilization was high. In this chapter, we show that these patients are more likely to be readmitted within 72 hours than patients discharged when utilization is lower. This effect prevails regardless of the utilization measurement or the chosen timeframe for readmission. Our results indicate that an additional day of recovery would help some patients who are being discharged when there are few available beds

The systematic early discharge of patients is problematic because readmissions are costly and could lead to an inefficient use of healthcare resources. In addition, early discharge with readmission is a potential public relations problem for a hospital. We propose four solutions for lowering readmission rates. A first solution is to add more flexibility to the post-operative path for patients. While there is a standard post-operative ICU for each service line, there might be other beds in the hospital that would be able to take a patient and allow the patient to recover more fully. Making it easier to match patients with beds might help to reduce readmission rates. A second solution would be the creation of a discharge checklist with objective criteria. Patients must satisfy the criteria before they can be discharged. This checklist can be used at all times, or only when the unit is operating at high levels of utilization. By standardizing the discharge process, it becomes more likely that each patient is fully ready to be transitioned out of the hospital at the time of discharge. Third, the hospital might consider using transition coaches, especially for patients at high risk of readmission. Hiring social workers to check on patients and to coach them on treatment and rehabilitation has been shown to lower the readmission rate (Coleman et al., 2006). A fourth solution would be to align a surgeon's compensation with a patient's health outcome. Currently, surgeons are paid for performing surgeries and having high operating room utilizations. By incorporating readmission rate into the compensation formula, we might impact the discharge decision process in a way that would lower the readmission rate.

Our work has identified a class of patients for whom the readmission rate is shown to be demonstrably higher. Patients discharged from a highly utilized unit are more likely to be readmitted to the hospital after surgery. Because the discharge rate increases when utilization is high, extra time in the post-operative unit for these patients might help lower the probability that they are readmitted.

In future work, we plan to address questions on the total length of stay for each type of patient. For example, do patients who are discharged from a highly utilized unit come back and stay longer than those who are discharged under normal circumstances? What effect do these early discharges and extra readmissions have on the overall efficiency of the hospital? In terms of throughput, there is a tradeoff to be considered when deciding to discharge a patient a day early. By discharging a patient early, a bed is freed up to allow a surgery to be performed. However, this discharged patient is now more likely to be readmitted. If the patient comes back, he/she might cause future surgeries to be postponed. There is a delicate tradeoff between rushing to discharge patients, which comes with the risk of higher readmission rates, and taking the time to treat patients fully in order to lower future readmission rates.

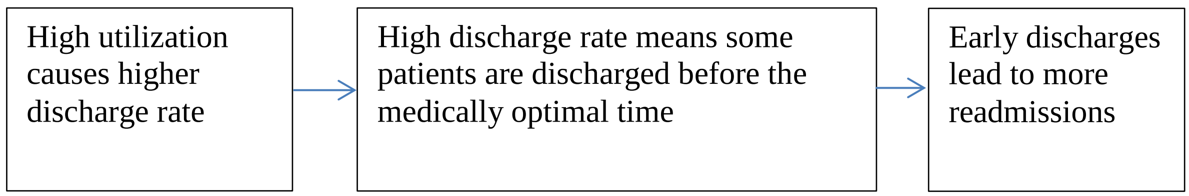

Figure 3-1: Theoretical Model of Hypothesis

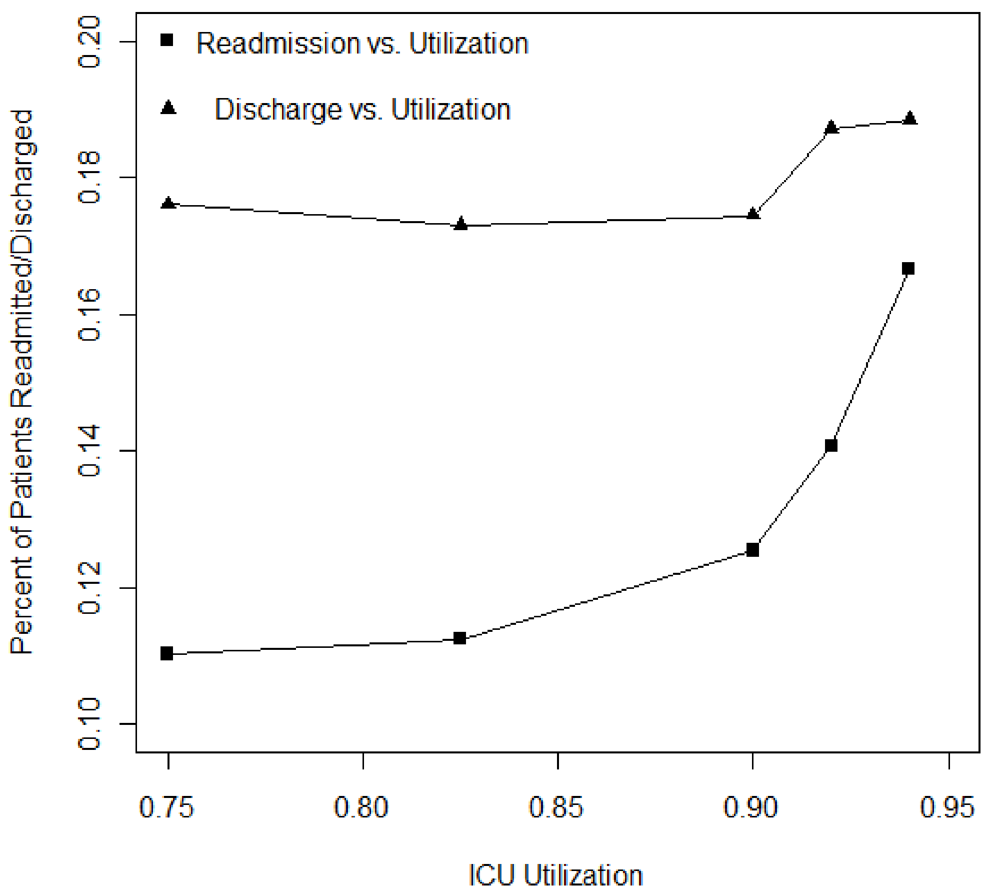

Figure 3-2: Readmission and discharge rates vs. post-operative utilization

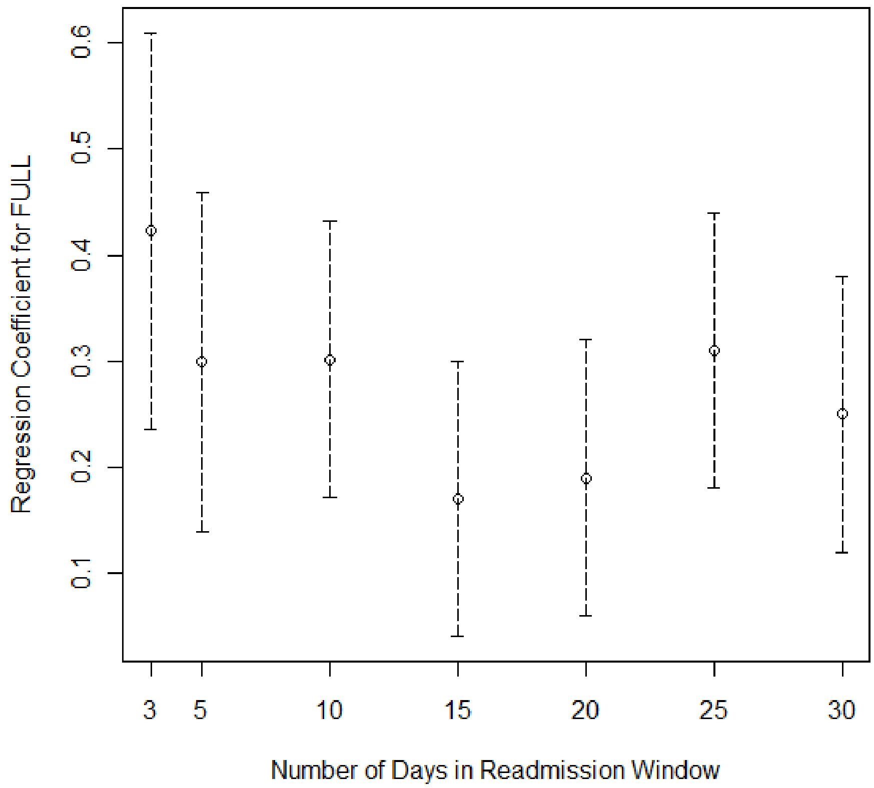

Figure 3-3: Effect of the definition of readmission on the effect of FULL

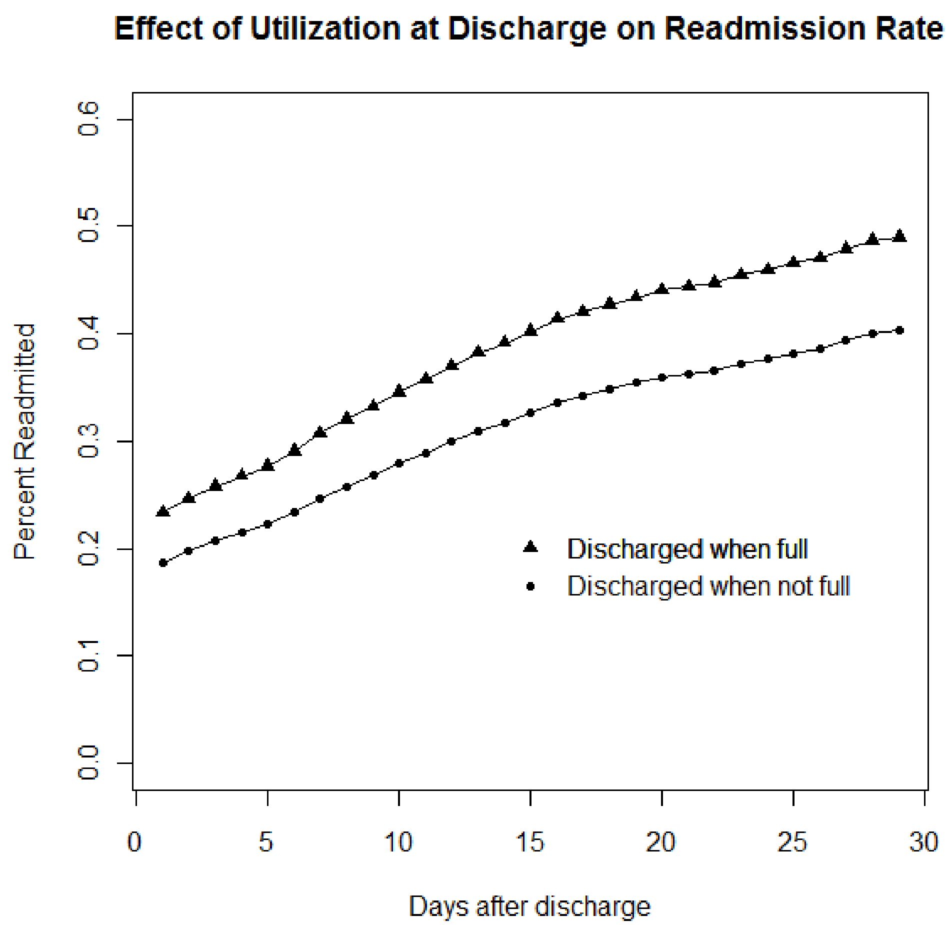

Figure 3-4: Readmission rates for patients discharged when full and not full

Table 3-1: Readmission rates for different utilization ranges

Utilization	Number of Patients	Number Readmitted	Percent Readmitted
<75	911	82	9.0
75-82.5	1124	117	10.4
82.5-90	1795	183	10.2
90-92	625	66	10.6
92-94	528	67	12.7
>94	282	47	16.7

Table 3-2: Descriptions of the response and explanatory variables

Variable Name	Description	Range
READMISSION72	A dummy variable that is 1 if the patient is readmitted within 72 hours of discharge, and 0 otherwise.	0,1
FULL	A dummy variable that is 1 if the unit was above 94% utilization when the patient was discharged, and 0 otherwise.	0,1
BEDS	The number of beds in use when the patient was discharged.	225, …, 323
AGE	The age, in years, of the patient on the day of the surgery.	12, …, 107
ELECTIVE	A dummy variable that is 1 if the surgery is elective, and 0 if the surgery is urgent.	0,1
BLACK, ASIAN, HISPANIC	Dummy variables that indicate the race of each patient. There are four races including white, so we use three indicator variables.	0,1
TRANSPLANT, TRAUMA,…, NEURO	Dummy variables that indicate which service line performed the surgery on the patient. There are 10 different service lines in our dataset, so we use nine indicator variables.	0,1
MALE	A dummy variable that indicates the gender of our patient. There are two genders, so we use one indicator variable.	0,1

Table 3-3: Model 1 results (AIC: 1978.6)

Variable	Odds Ratio	95% Confidence Interval	*p*-value
(Intercept)	0.088	[0.055 , 0.14]	<.001
FULL	2.341	[1.54 , 3.556]	<.001
BLACK	1.359	[1.055 , 1.748]	0.016
HISP	0.969	[0.449 , 2.084]	0.946
ASIAN	1.222	[0.344 , 4.335]	0.534
AGE	0.992	[0.984 , 0.998]	0.023
MALE	1.649	[1.279 , 2.126]	<.001
ELECTIVE	0.812	[0.639 , 1.029]	0.086
TRANS	9.772	[6.97 , 13.7]	<.001
NEURO	0.901	[0.54 , 1.502]	0.77
PLASTIC	1.029	[0.456 , 2.319]	0.791
GYNO	0.586	[0.309 , 1.11]	0.134
URO	5.447	[1.922 , 15.436]	0.001
OPTH	1.745	[0.98 , 3.105]	0.043
CARDIAC	1.545	[0.486 , 4.914]	0.334
TRAUMA	2.249	[1.361 , 3.716]	0.001
THORACIC	0.301	[0.089 , 1.009]	0.095

Table 3-4: Model 2 results (AIC: 1982.1)

Variable	Odds Ratio	95% Confidence Interval	*p*-value
(Intercept)	0.011	[0.002 , 0.044]	<.001
BEDS	1.008	[1.003 , 1.012]	0.001
BLACK	1.332	[1.035 , 1.714]	0.025
HISP	0.984	[0.461 , 2.1]	0.913
ASIAN	1.324	[0.373 , 4.69]	0.452
AGE	0.991	[0.983 , 0.998]	0.015
MALE	1.664	[1.29 , 2.145]	<.001
ELECTIVE	0.828	[0.653 , 1.051]	0.121
TRANS	9.790	[6.979 , 13.733]	<.001
NEURO	0.883	[0.529 , 1.472]	0.713
PLASTIC	1.053	[0.468 , 2.367]	0.748
GYNO	0.609	[0.322 , 1.15]	0.166
URO	6.057	[2.185 , 16.785]	<.001
OPTH	1.669	[0.938 , 2.97]	0.061
CARDIAC	1.624	[0.511 , 5.159]	0.293
TRAUMA	2.220	[1.343 , 3.665]	0.001
THORACIC	0.298	[0.088 , 0.998]	0.091

Table 3-5: Effect of increasing utilization on readmission

Number of Beds	Factor by which Odds are Increased	Percent Readmitted
225	1.00	8.6
250	1.22	10.4
275	1.49	12.4
300	1.82	14.7
325	2.22	17.4

Table 3-6: Trauma Surgery Model – Discrete Variable (AIC: 98.4)

Variable	Odds Ratio	95% Confidence Interval	*p*-value
(Intercept)	1.030	[0.189 , 5.591]	0.945
FULL	25.411	[1.085 , 594.979]	0.043
BLACK	8.234	[2.427 , 27.929]	<.001
HISP	5.655	[0.154 , 206.926]	0.305
AGE	0.943	[0.911 , 0.976]	0.001
MALE	0.782	[0.243 , 2.506]	0.658
ELECTIVE	0.580	[0.182 , 1.843]	0.381

Table 3-7: Trauma Surgery Model – Continuous Variable (AIC: 186)

Variable	Odds Ratio	95% Confidence Interval	*p*-value
(Intercept)	0.002	[0 , 0.857]	0.064
BEDS	1.027	[1.002 , 1.052]	0.047
BLACK	7.525	[2.276 , 24.872]	0.001
HISP	3.069	[0.085 , 109.748]	0.502
AGE	0.934	[0.899 , 0.968]	<.001
MALE	0.556	[0.16 , 1.926]	0.355
ELECTIVE	0.565	[0.178 , 1.784]	0.356

CHAPTER 4: DIFFERENCES IN TREATMENT QUALITY FOR TRAUMA PATIENTS BASED ON HOSPITAL ARRIVAL TIME

4.1 Introduction

The ability of hospitals to consistently deliver high quality care is a matter of significant concern. At hospitals in the United States, patients receive only about 54% of recommended care (McGlynn et al. 2003), and service quality varies considerably (Vandamme and Leunis 1993, Lam 2010). Finding ways to improve the quality and consistency of care delivered by healthcare systems is an important task facing the medical community. A critical step is to identify factors that lead to variations in the quality of care.

Until now, most studies have been focused on the differences in the quality of care across different types hospitals (e.g., academic hospitals and larger hospitals with higher volume tend to provide better quality service (Theokary and Ren 2011). Most hospital quality measures are aggregated at the hospital-year level, such as the HospitalCompare program (Centers for Medicaid and Medicare Services 2013) from the US Department of Health and Human Services or the hospital rankings published by the U.S. News & World Report (U.S. News & World Report 2012). While these quality rankings are useful in highlighting cross-hospital quality variation, within-hospital quality variation has not been studied extensively. For any hospital that aims at providing consistent, high-quality care, within-hospital variation in quality is an important concern. Within-hospital variation also presents a clear managerial challenge to hospitals. More specifically, systematic variations in quality indicate opportunities for a hospital to improve its quality of service.

In this chapter, we study the differences in treatment quality that trauma patients receive based on their arrival time at the emergency department. Using data from a large national database, we use regression analysis to determine the differences in treatment quality based on arrival time. We create a data set matching patients who arrive during the daytime to patients with the same injury severity and primary diagnosis who arrived off-hours. Trauma from unintentional injury is the leading cause of death among Americans age 1 to 44, and the fifth leading cause of death overall, with 121,902 deaths in 2008 (CDC 2008). Therefore, improving the quality and efficiency of trauma care is of national importance. Additionally, trauma care has a relatively short treatment cycle and clear quality metrics. These make it ideal to examine quality variation and consequences by focusing on quantifiable clinical outcomes.

In Section 4.2, we examine the relevant literature and develop our hypotheses. In Section 4.3, we describe the data set in detail. In Section 4.4, we present our empirical analysis and discuss our main results. In Section 4.5, we rule out alternative explanations of the observed differences between night and day and provide further statistical robustness checks. In Section 4.6, we discuss the implications of our findings and provide conclusions.

4.2 Literature Review and Hypothesis Development

4.2.1 Variations in the Quality of Care

The healthcare industry has been striving to consistently deliver quality care. Hospitals have tried many initiatives to increase the quality of healthcare, such as implementing checklists and guidelines for care (Downs and Black 1998, Gawande 2009). These efforts yield mixed results. Checklists have been shown to significantly lower the rates of preventable errors and to improve the quality of care delivered in hospitals. There are also nationwide quality transparency programs from both government agencies and practitioners, such as HHS' HospitalCompare, the U.S. News ranking of best hospitals, and the Leap Frog Group.

One limitation of these quality measures is that they are often constructed based on annually aggregated data and do not reflect how service quality varies within a hospital. Consequently, most existing academic research based on these quality measures has tried to examine cross-hospital quality variation using fixed hospital characteristics, such as size, volume, ownership, or teaching status, with the goal of improving the overall performance of low-quality hospitals. For example, Keeler et al. (1992) and Hughes et al. (1987) find that larger, urban hospitals deliver higher quality care than their smaller, rural counterparts, while McClellan and Staiger (2000) show that not-for-profit hospitals deliver a slightly higher quality of care than do for-profit hospitals.

Our study examines the quality of care issue from a different perspective by analyzing a systematic variation in quality over which hospitals have control:

namely, the differences between daytime and “off-hours” hospital staffing and resource availability. In so doing, we contribute to an emerging stream of research focusing on within-hospital quality variation. For example, Kc and Terwiesch (2009) examines how workload variation affects service rates and outcomes in cardiothoracic surgery. Several studies have analyzed treatment quality variation in hospitals based on the time of day and the day of the week. These studies (e.g., Magid et al. (2005), Saposnik et al. (2007), Bell and Redelmeier (2001), and Reeves et al. (2009)) show that patients outside of the emergency department have worse outcomes when they arrive off-hours, either at night or on the weekends. Specifically, Bell and Redelmeier (2001) find that risk-adjusted mortality rates for patients who arrive at the hospital on the weekends are significantly higher. Saposnik et al. (2007) also find increased mortality risk among patients who have strokes on the weekend. And recent work by Egol et al. (2011) reveals that mortality rates for trauma patients are higher at night, with larger off-hours/daytime variations at lower level trauma centers.

Although such work has begun to shed light on the issue of time-related within-hospital quality variation, they stop at reporting the difference in care quality and fail to reveal what causes the variation in quality of care. Additionally, most time-related studies use a black-box approach to focus on mortality, which makes it difficult to provide a comprehensive analysis on the off-hours/daytime quality difference. Our study aims to fill these gaps in the existing literature by presenting a comprehensive examination of how the timing of patient arrival affects the quality of care received and the outcome of treatment. More importantly, we try to uncover the mechanisms driving this quality variation, rooting our analysis in the literature on how resource availability affects quality of care and developing testable hypotheses for the causes of the quality variation.

Empirically, we advance existing research by using various approaches for more rigorous tests, including fixed effects models, diagnosis matching, as well as sub-category analysis. In so doing, we hope to contribute to a more comprehensive and fundamental understanding of the time effect in care quality variation.

Below we provide a theoretical foundation and a possible explanation for off-hours/daytime differences in quality of care from a resource management perspective, and derive testable hypotheses.

4.2.2 Possible Causes of the Variation in Care Quality between Daytime and Off-hours

There are many factors that lead to differences in the quality of care that hospitals offer. For example, hospitals with higher volume tend to provide higher quality care (Dudley et al. 2000). It has also been shown that, on average, academic hospitals and not-for-profit hospitals provide higher quality care than their nonacademic or for-profit counterparts (Jha et al. 2005).

Scholars in operations management have identified resource availability as one important factor that leads to quality variation in hospitals. One stream of research examines the impact of resource strain. Kc and Terwiesch (2011) show that when the ICU is full, patients are discharged at a higher rate, who are then readmitted at a higher rate. Anderson et al. (2011, 2012) uncover a similar effect when studying post-operative discharge and readmission rates. Cardoen et al. (2009) summarize the effects that operating room availability can have on patient care, with too few staffed ORs potentially leading to long patient waiting times and poorer clinical outcomes. Miro et al. (1999) show that overcrowding in emergency departments decreases the quality of care delivered, and Trzeciak and Rivers (2003) also discuss how overcrowded emergency departments offer lower quality care. These studies all concur on the basic premise that as strain on hospital resources (doctors, nurses, operating rooms) increases, the quality of the treatment provided declines.

Given the importance of resource availability in delivering consistent, high-quality care, the operations management literature has highlighted the critical role of resource management (Roth et al. 1995). A shortage of ICU beds can have a negative impact on those patients who are denied an ICU bed (Chan et al. 2012). Similarly, Price et al. (2011) show that shortages in downstream bed availability can affect post-operative care. Dobson et al. (2011) find that reserving capacity for urgent patients can help providers deliver higher quality of care overall. Soteriou et al. (1999) develop and present a method for optimal resource allocation to increase the perceived and actual quality of a healthcare delivery system.

However, operating room availability and waiting time before surgery are not the only determinants of quality of care. Recent work has discussed the importance of scheduling specialists optimally and the problems that improper utilization can lead to, such as high costs, long wait times, and lowered quality of care (Vissers et al. (2010), de Kreuk et al. (2004), Day et al. (2012)). It has also been shown that increasing specialization in hospitals leads to higher quality care at a lower cost (Eastaugh (2001), Hyer et al. (2008), Capkun et al. (2008), Barro et al. (2005)).

Even though specialization has been shown to be beneficial, there are reasons to believe that hospitals tend to reduce the level of specialization at off-hours. As a result, the question of how to staff service systems with varying arrival rates and different classes of patients has been well studied in the queueing theory literature. Pinker and Shumsky (2000) claim that "it is a well-known fact from the study of queues that, all things being equal, staffing flexible servers is more efficient than using specialists when customers are heterogeneous in the skills they require." This claim is further examined by Chevalier and Tabrodon (2003). Shumsky and Pinker (2003) also study the optimal number of specialists and generalists to have on staff at a call center, based on the arrival rate of patients and the difference in service quality offered by specialists and generalists. They find that as volume decreases, the optimal percentage of specialists decreases as well.

Trauma wards are a similar environment to those discussed in the studies mentioned above; arrival rates vary throughout the day and night, and many different injury types are treated. During off-hours, then, when arrival rates are lower, we would expect to see fewer specialty surgeons and fewer specialized resources available in a trauma ward. Diette et al. (2001) show that patients treated by specialists are more likely to receive high-quality care than those treated by general practitioners. Therefore, the lack of specialized workers (e.g., surgeons and nurses) will lead to a lower quality of care being delivered during off-hours and a higher rate of complications during surgery.

This provides a more complete picture of the impact of arrival time on health outcomes. Shorter waiting times typically lead to shorter ICU stays and lower mortality (Casaletto and Gatt 2003), while more lower quality care typically lead to longer length of stay and higher mortality (Haynes et al. 2009), so there is a conflict and it is not immediately clear which effect will dominate. Nevertheless, we propose the following hypothesis:

H1: Other things being equal, patients arriving during off-hours will receive lower quality care, and have worse outcomes, as measured by a) higher surgical complication rates, b) longer lengths of ICU stay, and c) higher mortality rates.

We hypothesize that a lack of specialized resources will cause hospitals to deliver lower quality care during off-hours. Because volume goes down during off-hours, we expect the availability of specialized surgeons, nurses, and other resources to decrease as well. This should result in lower quality service. Furthermore, the reduction of specialized resources should be more prominent among hospitals that are more resource-constrained (Keeler et al. 1992). Large hospitals have consistent demand, even at night, which should result in a smaller difference in the quality of resources available between daytime and off-hours than at smaller hospitals where there is a much lower off-hour arrival rate. Furthermore, while they vary by state, there are certain standards that trauma centers must meet in order to be certified as a high level trauma center. These restrictions often include having certain surgical staff in the hospital at all times, and having access to diagnostic and life support services. These restrictions guarantee that higher level trauma centers will have quality resources available during off-hours, while lower level trauma centers may not. Therefore, we expect to see greater increases (from day to off-hours) in complication rates, length of ICU stay, and mortality at smaller (measured by number of beds) and less sophisticated (lower trauma center level) hospitals. We also construct a more direct measure of resource strain based on the number of visits to the hospital per surgeon employed, providing a good proxy for the surgeons' workload. This suggests a second hypothesis:

H2: Hospitals that have fewer beds, level trauma centers, and more visits per surgeon will experience greater differences in quality of care between day and off-hours.

As a further test of whether resource constraints cause quality variation, we look at variation in outcomes based on surgery complexity. Not all surgeries are equally complex or require equal surgical specialization. Biondo et al. (2010) showed that in emergency surgeries, proper specialization can significantly improve outcomes and reduce mortality. Similarly, Chowdhury et al. (2007) showed that surgical specialization leads to better patient outcome and is a stronger determinant than hospital quality. This implies that a significant factor in determining quality of care is the proper specialization of the treating physician, when specialization is required.

If a lack of high quality resources is the main determinant of the difference in outcomes between daytime and off-hours, we would expect that if a patient's injury is straightforward to treat and does not require specialized resources, there should be little significant difference in the observed outcomes. However, if a patient has a complex injury that would be best treated by a specialized surgeon, we expect the outcome to be significantly better during the day, because those specialized resources are present during the daytime but not during off-hours.

H3: Patients with more complex injuries will experience greater differences in quality of care than patients with injuries that are relatively simple to treat.

4.3 Data

We use data from the National Trauma Data Bank (NTDB) version 7.2, which is the largest available aggregation of US trauma data. The research dataset includes all patients admitted at 570 trauma centers nationwide between 2002 and 2006, with treatment and outcome measures on over 1.5 million patients. The database contains demographic information on the patients, details of their treatment, injury type and severity, and payment, as well as information on the size, type, region, and trauma level of the hospital where the patient was treated. We restrict our focus to only those patients for whom we have complete information about age, arrival time in the emergency department (ED), injury severity score (ISS), mortality, and length of ICU stay. After excluding patients with significant missing data, and those who did not arrive at the hospital directly, we are left with a sample of 660,937 patients from 477 different hospitals. We only include direct arrivals to the hospital, excluding inter-hospital transfer patients. Patients who arrive via transfer have already been treated and have likely been stabilized. Therefore, the arrival time at the new hospital is not informative or a primary determinant of quality of care, so we exclude these patients from the study. The major variables that we use in the study are defined in Table 4-1. In Table 4-2, we provide summary statistics for each variable. Tables 4-3 and 4-4 provide summary statistics by arrival time and by hospital certification level.

Corresponding to the first four hypotheses, we choose four measures of quality of care: time to surgery, complication rate, length of ICU stay, and mortality. These quality metrics have been justified in the literature (Thomas et al. 1997; Dimick et al. 2003; and Thomas et al. 1993) and have all been shown to be key measures of the quality of care that a hospital provides. Time to surgery (*Hours to Procedure*) is measured as the number of hours between the patient's arrival at the emergency department and the first surgery that the patient receives. Occurrence of a recorded complication (*Complication*) is a dichotomous variable; if a patient has a complication recorded during his/her surgery, the variable is 1, and it is 0 otherwise. We measure length of ICU stay (*ICU LOS*) as the number of days that the patient spends in the ICU. Finally, *mortality* is recorded as 1 if the patient dies in the hospital, and 0 if the patient is discharged alive.

4.4 Empirical Analysis

4.4.1 Empirical Models

Since our goal is to examine the difference in quality of care between day and night, we adopt the following model to test H1-H3:

$$Q_{ik} = \beta_1 * T1_{ik} + \beta_2 * T2_{ik} + \gamma * Z_{ik} + \sum_{i=2}^{570} \theta_i * D_i + \varepsilon_{ik} \ ,$$

where i is the trauma center index and k is the patient index. Q is the quality measure, which varies depending on the specific hypothesis we test (e.g., length of stay, mortality). We measure the night effect in two time slots. $T1$ is the time dummy for late night (6:00 PM to 12:00 AM), and $T2$ is the time dummy for early morning (12:00 AM to 6 AM). Z represents a group of control variables for the heterogeneity in patient mix, including injury severity scale (*ISS*) score of the patient, as well as the patient's comorbidity index, primary ICD-9 diagnosis code, age, race, and gender.

To control for the unobserved heterogeneity across the trauma centers, we include the fixed effect, D, for each trauma center (the intercept is omitted to avoid perfect multicollinearity). Robust standard errors are used to control for potential heteroskedasticity in the sample. We further cluster the standard errors at the facility level to account for possible correlations in the standard errors within the same trauma center. While there are potential sources of endogeneity, particularly due to differences in the composition of the patient populations arriving during the daytime and off-hours, we report these initial results because they expand upon the initial findings by Egol et al. (2011) and motivate further investigation. In Section 4.5, we test alternative explanations and try to control specifically for this unobserved heterogeneity. Unless otherwise noted, every model that we test also uses hospital fixed-effects.

4.4.2 Findings

4.4.2.1 Quality of Surgery*:*

To test Hypothesis 1a, we regress whether or not patients had a complication during surgery, taking into account their arrival time and demographic characteristics.

$$Logit(Complication)_{ik} = \beta_1 * Early\ AM_{ik} + \beta_2 * Night_{ik} + \beta_3 * Age_{ik} + B_4 * Gender_{ik} + \beta_5 * Race_{ik} + \beta_6 * \log(ISS)_{ik} + \beta_7 * Facility_{ik} + \beta_8 * Year + \sum_1^N \theta_i * D_i + \varepsilon_{ik},$$

Since the dependent variable is binary, we apply a logit model. The baseline results are given in the column 1 of Table 4-5. After controlling for severity and demographics, we find that a patient arriving at night (or in the early morning) has, respectively, a 4.8% (or 9.3%) higher odds ($e^{.0467} = 1.048$, $e^{.0887} = 1.093$) of incurring a complication than do patients arriving during the daytime. The probability of death increases from 5.0% for a typical patient arriving during the daytime to 5.2% for patients arriving at night and to 5.5% for patients arriving in the early AM ($e^{-2.996} = .05$, $e^{-2.996 + .0467} = .0524$, $e^{-2.996 + .0887} = .0546$). After including the fixed effect dummies, in column 2, the above findings are essentially unchanged.

As another measure of surgery quality, we examine whether patients require multiple surgeries, which is a common outcome of an incomplete or unsuccessful initial surgery. For this analysis, we only include patients with at least one procedure. These results are shown in column 3 of Table 4-5. We find that the odds of a patient arriving during the early morning (or night) requiring multiple surgeries are 18.2% (or 3.8%) higher than for patients arriving during the day (raising the probability for a typical patient from 81.4% for daytime arrivals to 83.8% for early AM arrivals and 82.0% for night arrivals). Taken together, the higher complication rate during off-hours and the fact that daytime surgeries lead to fewer subsequent surgeries imply that the quality of treatment, especially in surgery, is lower during off-hours than during the day.

4.4.2.2 Length of Stay

We next examine the relationship between arrival time and the length of time a patient spends in the ICU, in order to test Hypothesis 1b, using OLS with hospital fixed effects.

$$Log(ICU\ LOS)_{ik} = \beta_1 * Early\ AM_{ik} + \beta_2 * Night_{ik} + \beta_3 * Age_{ik} + B_4 * Gender_{ik} + \beta_5 * Race_{ik} + \beta_6 * log(ISS)_{ik} + \beta_7 * Facility_{ik} + \beta_8 * Year_{ik} + \sum_1^N \theta_i * D_i + \varepsilon_{ik},$$

These results are shown in columns 4 and 5 of Table 4-5. The data show that patients who arrive at the ED in the early morning or at night have ICU stays that are 16.6% and 10.0% longer ($e^{0.154} = 1.166$, $e^{0.0949} = 1.100$), respectively, than patients who arrive during the day, after controlling for demographics, hospital characteristics, and the patient’s severity. This means that not only do patients who arrive at the ED at night or in the early morning have higher complication rates, but they also have longer recovery times after their treatment, after controlling for the severity of their injuries. This implies the care these patients receive is of lower quality than the care delivered during the day.

We conduct further analysis to ensure that the above finding is not simply an artifact of the way hospitals calculate length of stay. It is possible that some hospitals calculate based on calendar days instead of full 24-hour periods, so patients who arrive just before midnight would have an extra day added to their length of stay. However, we also find that patients who arrive at night or in the early morning are both significantly more likely to be sent to the ICU in the first place, suggesting that our findings are not driven by the way length of stay is calculated. Because patients are more likely to require any time in the ICU at all, we find it likely that they also spend, on average, more time in the ICU as well.

4.4.2.2 Mortality

We use the following logistic regression model to examine the effect of patient arrival time on mortality:

$$logit(Mortality_{ik}) = \beta_1 * Early\ AM_{ik} + \beta_2 * Night_{ik} + \beta_3 * Age_{ik} + B_4 * Gender_{ik} + \beta_5 * Race_{ik}. + \beta_6 * log(ISS)_{ik} + \beta_7 * Facility_{ik} + \beta_8 * Year_{ik} + \sum_1^N \theta_i * D_i + \varepsilon_{ik},$$

The regression results are reported in Table 4-6. Using the baseline model (Column 1), we find that the coefficients of *Night* and *Early AM* are .116 and .112, respectively. They are both statistically significant at the .001 level. These coefficients imply that after controlling for patient characteristics, the odds of death for patients who arrive between 6 PM and midnight and those who arrive between midnight and 6 AM increase by 12.3% and 11.9% respectively, when compared to patients arriving during the day. This raises the odds of death for a typical patient from 5.0% for daytime arrivals to 5.6% for both nighttime and early AM arrivals ($e^{-2.995} = .05$, $e^{-2.995+.112} = .0559$, $e^{-2.995+.116} = .0561$). These results are consistent with the fixed effects model given in the column 2.

4.4.3 Further Analysis and Mechanism Discovery

In the above analysis, we find a significant difference in the quality of care delivered during off-hours and the quality of care delivered during the day, which supports H1. As discussed in Section 2, we suspect that one main cause of the lower treatment quality during off-hours is a reduced breadth of resources available. In this section, we further examine whether this is the case and how the difference is affected by resource availability. We first utilize several proxies for the resource variable across hospitals (H2), and then leverage the difference in resource requirements across clinical conditions (H3).

4.4.3.1 Comparing the day and off-hours difference among hospitals

First, according to Hoetker (2007), we stratify our sample based on the level of the trauma center at which the patient was treated. Level I trauma centers are defined as those possessing a full range of specialists and equipment available 24 hours a day. Level II centers are required to have all essential personnel available 24 hours a day, but not required to have every specialty staffed at all times. Level III and IV centers are not required to have all specialties fully available. Because of these restrictions, we expect the differences in outcomes to be larger at lower-level trauma centers, as they are the most resource-constrained.

Consistent with our prediction, we see that the off-hours/daytime difference in mortality rates is greatest at the hospitals with the most visits per surgeon and smallest at the hospitals with the fewest visits per surgeon. At level I trauma centers, the average difference in the odds of mortality between daytime and off-hours is 10.6%; at level II centers, it is 14.3%; and at level III/IV centers, the difference rises to 22.3% ($e^{(.0928+.110)/2} = 1.106$, $e^{(.138+.130)/2} = 1.143$, $e^{(.231+.171)/2} = 1.226$). This raises the probability of death for a typical patient from an average of 5.7% to 6.3% at level I centers, from an average of 4.7% to 5.4% at level II centers, and from an average of 2.7% to 3.3% at level III/IV centers. These differences are further confirmed by the logistic regression estimations based on each level of trauma centers, as reported in Table 4-7. As before, these models include hospital level fixed effects, in addition to the patient level control variables.

Next, as hypothesized in H2, we suspect that the off-hours/daytime difference in resources available would be greater at smaller, less sophisticated hospitals, due to reduced use of specialists (Pinker and Shumsky 2000). We first segment our sample based on the size of the hospital, measured by the number of beds in the hospital. We then construct the relative load that the surgeons and hospital face and the strain that the patient flow puts on their resources, based on the number of visits to the hospital per number of trauma surgeons employed by the hospital. For each index, sub-sample regressions are conducted using the top and bottom quartile observations based on the above indices. As presented in Table 4-8, the quality difference is greater for hospitals with number of beds in the bottom quartile, compared to those in the top quartile (column 1 vs. column 2). More interestingly, as resource strain increases (hospitals moving from bottom quartile by visits per surgeon to top quartile), the increase in mortality in early AM becomes much more significant (column 3 vs. column 4). These findings are consistent with our conjecture that resource-constrained trauma centers experience a quality drop-off in the early morning/night. Again, these models include hospital level fixed effects, in addition to the patient level control variables.

4.4.3.2 Comparing the day and off-hours difference based on complexity of injury

In this section, we exploit another way to test whether resource availability leads to the difference in care quality. Because complex injuries are more likely to require specialized resources, the difference in quality of care these patients receive during off-hours and during the daytime should be larger, while it should be smaller for those patients who have simpler injuries (we thank the AE and one anonymous reviewer for this suggestion). To test whether the complexity of a patient's injury is associated with the time variation in quality of care, we first sought the expertise of emergency healthcare workers, who in discussions with our research team identified broken femurs as a good example of a low-complexity injury and spinal column/neck injuries as a good example of a high-complexity injury type. While the levels of complexity are different, these two injury types have similar levels of severity and reasonably similar overall mortality rates (5% for femur injuries and 8% for neck/spine injuries), making them ideal for comparison. In addition, the sample size of each population is large enough for meaningful analysis. We then compare the differences in observed outcomes between patients with broken femurs and those who suffered spinal column and neck injuries.

While broken femurs are a serious injury with a 5% mortality rate, treatment is relatively straightforward and an emergency room doctor or generalist trauma surgeon would be qualified to treat such an injury. Rarely would a specialist be called in to treat a patient with a broken femur, regardless of time of day. Because of this, we would expect to see very little difference in treatment quality between the daytime and off-hours. Indeed, regression of mortality on arrival time among patients with broken femurs shows that there is no statistically significant difference in treatment quality between daytime and off-hours (see column 1 of Table 4-9).

Regarding neck and spine injuries, on the other hand, we would anticipate a large difference in treatment quality between daytime and off-hours. This is due to the fact that the neck contains crucial components of several major physical systems, making these injuries far more complex than broken femurs. Specifically, the neck contains the carotid artery and the jugular vein, which carry blood to and from the brain, the spinal cord, which connects the nervous system to the brain, and the esophagus and trachea, which carry food and air from the mouth into the body. The neck also contains salivary glands, thyroid glands, and lymph nodes, all of which play important roles in the human body. Injury to the neck or spine can cause damage to any or all of these organs. Because of the high level of complexity spinal and neck injuries present, they are often treated by specialists, who may not be present in the hospital overnight. When we regress mortality on arrival times, then, we do see a very strong effect. The odds of mortality for patients are 16.0% higher for those arriving at night, and 17.0% higher for those arriving in the early AM than for patients arriving during the daytime. In both cases, this raises the probability of death from 5.0% for patients arriving during the daytime to 5.8% for patients arriving at night or in the early AM). These results are reported in column 2 of Table 4-9. Therefore, we see that the effect of arrival time on mortality is strongly dependent on the complexity of the case. Patients who would typically see a specialist during the day receive much lower quality care when they arrive during off-hours.

More generally, we compare the differences in outcomes for all patients that we consider to have highly complex injuries to those for patients with relatively simple injuries. We classify any patient with an injury to the brain, heart, or spinal cord as complex. Patients with broken bones in the hips, legs, or arms are classified as simple patients. We see that the mortality rate is significantly higher for complex patients off-hours than during the daytime; however there is no difference for simple patients. These results are shown in Table 4-10 (columns 3 and 4).

4.4 Alternative Explanations and Robustness Checks

There are a few possible alternative explanations for the observed results. The first is that the difference in mortality is a consequence of doctors being more tired during off-hours, as suggested by Egol et al. (2011). This may help to explain some of the differences that we see between “night” and “early morning” time periods: quality is typically worst in the early morning, and this difference might be driven by disruptions in doctors' circadian rhythms. However, we find this to be an unconvincing explanation for the overall difference in quality between off-hours and daytime care. While fatigue may play a small part in this difference, if it were the whole explanation, then we would observe a uniform effect across all hospitals and across patient types. But no such uniform effect exists, and it is unlikely that doctors at level I trauma centers are somehow better at functioning at 3 AM than doctors at level IV trauma centers. However, the quality variation trends we identify do fit the pattern predicted when considering resource shortages as a cause of the difference in outcomes.

A second possible explanation is that there is some unobserved difference between the population of patients that arrives during off-hours and those who arrive during the day. We address this possibility in three different ways. First, we add in control variables for each of the 250 most common ICD-9 diagnosis codes. Each patient has between 1 and 10 diagnoses, depending on the extent of his/her injuries. For each patient, we record whether or not he was diagnosed with any of the 250 most common ICD-9 codes. By controlling for specific diagnoses, we are able isolate the effect of differing staffing levels between night and day from the differences in patient case mix. Also, injury severity and comorbidity index may influence mortality in a non-linear manner. To account for this, we treat them as categorical variables, including a dummy variable for each possible ISS and comorbidity index score. After controlling for specific diagnosis, and the exact severity and comorbidity index, we see that there is still a significant increase in mortality off-hours (shown in column 1 of Table 4-10).

Second, we compare the patients arriving between 3:00 PM and 5:00 PM to those who arrive between 8:00 PM and 10:00 PM. We chose these time intervals to give a buffer on either side of the traditional 7:00 PM shift for differing hospital procedures. However, our results are robust to specific time choices. These patients are similar in terms of severity (average ISS = 11.2 vs. 11.4, average comorbidity index = 0.15 vs. 0.18), giving us similar patient cohorts. The only major difference is arrival time and hospital resource availability. We see that patients arriving shortly after shift change (between 8:00 PM and 10:00 PM) have 10.7% higher odds of death than a similar patient arriving just before shift change (3:00 PM to 5:00 PM). This raises the probability of death from 6.0% to 6.6%. These results are shown in Column 2 of Table 4-10. We also examined similar subsets of patients around the morning shift change, but patients arriving between 3:00 AM and 5:00 AM are significantly different in terms of severity and demographics from patients arriving between 8:00 and 10:00 AM; this comparison is, therefore, less useful.

Third, we match patients who arrive during the daytime to patients who arrive off-hours based on their primary diagnosis ICD-9 code. For each patient who arrives during the day, we take the patient arriving off-hours who has the closest severity score to the daytime arrival patient among those who also have the same primary diagnosis. In the overwhelming majority of cases, we are able to match patients who have the same primary injury and exactly the same severity score. The average difference in ISS is less than 1 point, meaning that the matching does a good job of finding patients of similar severity. The resulting dataset consists only of patients who arrived during the day, and patients who match them exactly on primary diagnosis and are very close (or exactly the same) with respect to severity who arrive off-hours. By doing this, we can further isolate the effect of arrival time from the differences in patient populations. We are comparing virtually identical populations now in terms of diagnoses and injury severity. However, we cannot be certain that we have controlled for all external factors. When we regress mortality on arrival time in this sample, we get results consistent with our previous analysis (shown in column 5 of Table 4-10). These results further tell us that the differences in mortality rates are due to hospital factors and not to the differences in the patient population.   While the effect size is somewhat smaller in the matched sample, we still have a consistent effect that is strongly statistically significant.

A fourth possible explanation is that, instead of the quality and specificity of resources being lower at night, there is overall a general lack of resources available, and, therefore, patients must wait longer to be treated. To test this, we look to find evidence that patients who are treated off-hours have longer waiting times or are less likely to receive care in the first hour of arriving to the hospital. The regression model that we test, using OLS with hospital fixed effects, is:

$$\textit{Hours to Procedure}_{ik} = \beta_1 * \textit{Early AM}_{ik} + \beta_2 * \textit{Night}_{ik} + \beta_3 * \textit{Age}_{ik} + B_4 * \textit{Gender}_{ik} + \beta_5 * \textit{Race}_{ik} + \beta_6 * \log(\textit{ISS})_{ik} + \beta_7 * \textit{Facility}_{ik} + \beta_8 * \textit{Year} + \sum_1^N \theta_i * D_i + \varepsilon_{ik}$$

where *Early AM* and *Night* are dummy variables indicating whether the patient arrived during the early morning or night, *ISS* is the patient's Injury Severity Score, and *Trauma Level* is a categorical variable indicating the level of trauma center at which the patient was treated. While a hospital's trauma level can change from year to year, this is a relatively rare event. These infrequent changes are reflected in our data.

The above model is estimated using OLS, and the results are shown in Table 4-11. Column 1 reports the baseline model. We see that the coefficients for the early AM and night dummy variables are negative and significant at the $p < 0.001$ level. This means that patients arriving during the night and early morning periods have shorter waiting times for surgery than patients arriving during the daytime. Specifically, if a patient arrives between 6 PM and 12 AM, the expected waiting time is 14 minutes shorter (-.236 * 60 = -14) than for similar patients who arrive during the day; if a patient arrives between 12 AM and 6 AM, the expected waiting time is 8 minutes shorter (-.126* 60 = -8). The column 2 results in Table 4-11 report the model with fixed effect dummies added for each individual hospital. The results are essentially the same.

As an alternative measure, we test the probability that the patient receives surgery within one hour of arriving at the hospital. This is an important metric in trauma care, where prompt treatment is usually essential. We see that the odds that a patient arriving at night (or during the early morning) will have surgery in the first hour after arriving at a hospital are on average 21.2% (or 21.9%) higher than for patients arriving during the day (raising the probability for an average patient from 21.2% during the daytime to 24.6% at night or 24.7% in the early AM), after controlling for patient and hospital characteristics (shown in column 3 of Table 4-11). Taken together, these results show that patients receive care more promptly during off-hours than do comparable patients who arrive during the day. The fact that patients are treated faster off-hours than they are during the daytime rejects the alternative explanation, and, in fact, strengthens our previous results. Patients who arrive off-hours have worse outcomes, even though they have shorter waiting times. This means that it is not a general lack of resources that causes the decrease in outcome quality.

While the NTDB is generally a high-quality dataset, there are some missing data, and it is possible that the treatment of this missing data may affect our results. In our regressions, we assumed that data were missing completely at random, and any observation from a continuous variable with missing data was simply removed. Missing data from categorical variables are treated as a separate category. To examine whether this assumption biases our results or not, we then imputed the values of any missing continuous variable and estimated the time effect again, using linear imputation for missing values. These results are presented in column 3 of Table 4-6. We find that the impact of replacing missing variables with their imputed values is negligible and the effect of arrival time on mortality is still strongly present.

Similarly, one may question whether the relationship between severity and mortality is truly log-linear. Using the Box-Cox transform of the injury severity score instead of the logged severity (shown in column 4 of Table 4-6) does not dramatically change the earlier findings.

We are also aware that the sample size changes across various regressions. The samples used in the analyses dealing with procedures (time to procedure, complication rate, likelihood of multiple surgeries) include only patients who received surgical treatment. There are also some small variations in the sample size between the fixed effects models, as some observations are dropped due to some individual facilities being perfect predictors. One final robustness check utilized a constant sample across all regressions. The results, reported in Table 4-12, remain much the same, both in magnitude and statistical significance. This gives us further confidence that the findings are robust.

4.6 Discussion

In this study, we explore the fluctuation of healthcare quality with respect to patient arrival time from a resource availability perspective. Using a large collection of national trauma injury data, we find that patients are treated more promptly during off-hours than during the day. However, off-hours patients face significantly higher mortality rates, longer ICU stays, higher surgery complication rates, and an increased risk of needing multiple surgeries. Considered side by side, these two findings suggest that patients arriving during the day have better outcomes not because more care is available, but because the quality of care that they receive is better. We then design various tests to uncover the mechanism that drives the deterioration of care quality. We find that the fluctuation is much larger at smaller hospitals, at lower level trauma centers, and with more complex injuries. All these are consistent with our theoretical explanation that the low availability of high-quality, specialized resources (surgeons, nursing staff, lab availability, etc.) causes worse clinical outcomes.

The above empirical findings are further corroborated by our discussions with medical professionals working in major hospitals. Take, for example, a patient who comes to the hospital during off-hours needing a specialized surgery. At a large, sophisticated hospital, this patient is likely to be seen by a specialized surgeon, an operating room is likely to be available, and the patient should receive the appropriate level of care. If the patient requires specialized lab tests, or especially intensive care, there is a higher chance that the required resources are available during the day than off-hours. During the day, an appropriately specialized surgeon will likely be on duty, as well; however, it is also more likely that all of the operating rooms are full, thus increasing the average wait time for the patient.

Now consider this same patient arriving at a smaller, lower level trauma center. If he arrives at night, there might only be a general trauma surgeon available. This surgeon can perform the surgery, but the likelihood that he will make a mistake, resulting in a complication, rises. He also might perform a temporary “patch” surgery to stabilize the patient until a specialist is available. Since this patient will have received inferior care, he will have a higher mortality rate and will spend more time in the ICU recovering. He will also be more likely to need multiple operations, either to fix problems arising from the complications or because the first surgery did not address all of his needs.

We find an interesting tradeoff between daytime and off-hours. If the patient arrives during the day, all of the operating rooms are more likely to be full, thus increasing the average waiting time for the patient. Still, our results, based on clinical outcomes, clearly suggest that higher quality care is worth the wait. Consistent with work in the service operations management literature regarding the tradeoffs between hiring generalist and specialist workers (Pinker and Shumsky, 2000), during the off-hours, when arrival rates are lower, there are fewer specialized resources available. Therefore, while the hospital is less busy off-hours, the quality and variety of resources available are also much less than during the day. A patient arriving at night or in the early morning will get more prompt treatment, but will get less specialized and sometimes lower quality service. These mechanisms are sufficient to explain the differences that we see in the promptness of care that patients receive as well as the differences that we see in patient outcomes.

This study extends the clinical literature on the effect of timing on quality of care (Magid et al. (2005), Saposnik et al. (2007), Bell and Redelmeier (2001), Reeves et al. (2009), Egol et al. (2011)) in several important ways. These earlier studies do little to explain the cause or to discuss the operational implications of their findings. In this chapter, we develop specific hypotheses grounded in the operations management literature about the causes for quality variation across time and identify specific classes of patients who will be more likely to receive lower quality care during off-hours. While the above-mentioned studies focus on a single outcome measure (mostly mortality rate), we provide a comprehensive examination of several imp-ortant operational and quality measures (e.g., waiting time to surgery, ICU length of stay, surgical complication rate, and likelihood of multiple surgeries). The significant improvement in both scope and depth in our study helps uncover the mechanisms that cause the quality difference. Methodologically, we also advance the rigor of analysis in several ways. Our empirical approach controls for unobserved hospital-specific factors by using fixed effects models, and we also better control for patient heterogeneity by diagnosis matching.

Our study has its limitations, e.g., the quality of data from the NTDB and our lack of direct observation of resource allocation. While the NTDB is a well-maintained data set, response bias ought to be a concern when dealing with any voluntary sample. The possible effects of patient mix on the quality of outcomes is another potential concern. Further, although we have had discussions with emergency and trauma physicians who have confirmed that staffing levels are typically lower overnight, it would be desirable to obtain direct quantitative observations. In the future, we would like to measure how resource availability and patient volume impact quality of care in the context of specific hospitals. Finally, there is still some remaining question as to whether the differences in outcomes are driven by resource availability or if it is simply unobserved heterogeneity in the patient population. The patients who arrive off-hours tend to have more severe and extensive injuries which result in higher mortality rates. While we control for much of the difference in patient population (age, gender, race, severity, exact diagnosis, comorbidities, etc), it is possible that there are further underlying differences in the patient populations that we cannot observe.

In summary, this work explores one important source of within-hospital variation in quality of care, the time of day of patient arrival. We show that there are systematic ways in which hospital quality of care is affected. Roughly half of patients arrive off-hours, and half during the daytime. The average daytime mortality rate is 4.9%, and we calculated an increase in the odds ratio of death of approximately 10% for off-hours arrivals, increasing the mortality rate from 4.9% to 5.4%. Together, these facts indicate that about 4.5% of all deaths in our sample occurred due to the differences in quality of care between daytime and off-hours. This translates to approximately 8,100 extra deaths every year at hospitals that report to the NTDB. Furthermore, the average patient spends 2.0 days in the ICU, and we find an average increase of 13.8% (exp ((.163 + .097)/2) = 1.138) in the length of ICU stay for patients arriving off-hours. This translates to an estimated 67,000 extra patient-days in the ICU for the hospitals in our sample over the five years studied. At an average of $19,642 per day for an ICU stay (Dasta et al. 2005), this translates to an additional economic cost of $1.3 billion. These findings hold practical implications for trauma centers: namely, the drop-off in quality at night and in the early morning could be mitigated by increasing off-hours staffing levels and making an effort to have specialized resources available round the clock. Advances in telemedicine may offer some relatively inexpensive opportunities in this regard. We hope that this work will lead to more research on applying operations management knowledge to reduce healthcare quality fluctuation.

Table 4-1: Summary of the variables

Variable	**Description**
Age	The patient's age, in years
Gender	The patient's gender
ISS	The injury severity score of the patient
Comobidity	The patient's Deyo-Charlson comorbidity index score
Trauma Level	The level of the trauma center at which the patient was treated, used as a categorical variable
Early AM	A dummy variable that is 1 if the patient arrived between midnight and 6 AM, and 0 otherwise
Night	A dummy variable that is 1 if the patient arrived between 6 PM and midnight, and 0 otherwise
Mortality	A dummy variable that is 1 if the patient died, and 0 otherwise
ICU LOS	The number of days the patient spent in the Intensive Care Unit
Complication	A dummy variable that is 1 if the patient had a recorded complication during surgery
Hours To Procedure	The number of hours the patient had to wait until surgery
Facility	The identification key for the hospital at which the patient was treated, used as a categorical variable
Procedure	A dummy variable that is 1 if the patient had a surgery, and 0 otherwise
Multiple	A dummy variable that is 1 if the patient required multiple surgical procedures, and 0 otherwise
Prompt	A dummy variable that is 1 if the patient received treatment within 1 hour of arrival, and 0 otherwise
Year of Discharge	The year that the patient was discharged from the hospital

Table 4-2: Summary Statistics

Variable	**Mean**	**Std. Dev.**	**Min**	**Max**
Age	44.8	20.3	18	89
Male	.65	.48	0	1
ISS	10.2	10.0	0	75
Comorbidity	.14	.63	0	14
Early AM	.19	.39	0	1
Night	.30	.48	0	1
Mortality	.05	.22	0	1
ICU LOS	1.32	5.16	0	300
Complication	.06	.23	0	1
Hours To Procedure	3.51	2.27	0	6
Procedure	.70	.46	0	1
Multiple	.78	.41	0	1
Prompt	.37	.47	0	1

Table 4-3: Summary statistics by arrival time
(Early AM: midnight – 6 AM; Daytime: 6 AM – 6 PM; Night: 6 PM – midnight)

Time of Day	Average ISS	Mortality	Average Days in ICU	Percent Having Surgery	Average Age
Early AM	10.72	5.377%	1.41	71.4%	36.1
Daytime	9.92	4.917%	1.27	70.2%	48.7
Night	10.25	5.389%	1.36	69.7%	43.4
----------------	--------------	-------------------	---------------	--------------------------	-------------------
Time of Day	Number of Patients	Percent Surgery within 1 hour	Percent with ICU Stays	Percent of Surgeries with Complications	Average Hours to Surgery
Early AM	138304	20.9%	26.0%	5.6%	3.44
Daytime	384550	17.0%	21.2%	5.4%	4.09
Night	221487	20.0%	23.6%	5.6%	3.67

Table 4-4: Summary Statistics by Hospital Level

	Number of Hospitals	Number of Patients Treated	Average Surgeons	Average Visits Per Surgeon	Average Beds
Level 1	175	424510	21.5	112.83	454.5
Level 2	188	260241	18.7	74.025	299.2
Level 3-5	97	27805	12.3	23.305	129.9

Table 4-5: Surgery outcome: complications, multiple surgeries, and length of stay

	Complication	Fixed Effects Complication	Multiple	Log (ICU LOS)	Fixed Effects Log (ICU LOS)
	(1)	(2)	(3)	(4)	(5)
Early AM	0.0887***	0.0953***	0.167***	0.154***	0.163***
	(0.0174)	(0.0219)	(0.0118)	(0.00779)	(0.0141)
Night	0.0467***	0.0329*	0.0377***	0.0949***	0.0917***
	(0.0144)	(0.0179)	(0.00875)	(0.00642)	(0.00965)
Age	0.0109***	0.0129***	-0.0161***	-0.00179***	0.000198
	(0.00034)	(0.00068)	(0.000202)	(0.000154)	(0.000529)
Log(ISS)	1.305***	1.313***	0.436***	0.990***	0.982***
	(0.0107)	(0.0303)	(0.00374)	(0.00288)	(0.0276)
Comorbid	0.170***	0.148***	0.144***	0.117***	0.0801***
	(0.00664)	(0.014)	(0.00613)	(0.00491)	(0.0104)
Black	0.115**	0.136**	0.113***	-0.0756***	-0.154***
	(0.0562)	(0.0659)	(0.0436)	(0.021)	(0.0326)
Hispanic	0.175***	0.0261	-0.588***	-0.128***	-0.125***
	(0.0585)	(0.0648)	(0.0444)	(0.0214)	(0.0304)
Other	-0.123*	-0.130*	0.110**	-0.0394*	-0.0905***
	(0.0631)	(0.0784)	(0.0478)	(0.0201)	(0.0273)
White	-0.00087	0.0603	-0.153***	0.261***	0.229***
	(0.0544)	(0.0612)	(0.0423)	(0.00594)	(0.0139)
Male	0.113***	0.111***	0.228***	1.110**	-5.869***
	(0.014)	(0.0219)	(0.00798)	(0.503)	(0.0836)
Level II	-0.268***		-0.282***	-0.196***	
	(0.014)		(0.008)	(0.00597)	
Level III	-0.177***		-1.058***	-0.538***	
	(0.0414)		(0.0183)	(0.0131)	
Level IV	0.113		-0.654***	-1.175***	
	(0.172)		(0.0961)	(0.0361)	
Level NA	0.0119		-0.726***	-0.706***	
	(0.0373)		(0.0195)	(0.0117)	
Constant	-4.922***		2.294***	-6.236***	
	(1.146)		(0.13)	(0.503)	
Observations	366,813	344,497	472,431	660,937	660,937
Pseudo R^2	.1379	.1907	.0726	0.185	0.717

Clustered robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4-6: Outcome Results: Mortality

	Mortality	Fixed Effects Mortality	Missing Imputed Mortality	ISS Box-Cox Transformation Mortality
	(1)	(2)	(3)	(4)
Early AM	0.112***	0.118***	0.107***	0.115***
	(0.0165)	(0.0202)	(0.0160)	(0.0169)
Night	0.116***	0.113***	0.110***	0.116***
	(0.0136)	(0.0154)	(0.0132)	(0.0139)
Age	0.0201***	0.0213***	0.0191***	0.0211***
	(0.000339)	(0.00085)	(0.000327)	(0.000345)
Log(ISS)	1.991***	2.039***	1.862***	1.274***
	(0.0142)	(0.0596)	(0.0129)	(0.00759)
Comorbid	0.0852***	0.0888***	0.0746***	0.0891***
	(0.00791)	(0.0136)	(0.00779)	(0.00801)
Black	0.344***	0.370***	0.355***	0.361***
	(0.0448)	(0.109)	(0.0436)	(0.0459)
Hispanic	0.0793*	0.100	0.0701	0.0888*
	(0.0465)	(0.105)	(0.0453)	(0.0477)
Other	-0.0580	0.0666	-0.0625	-0.0219
	(0.0528)	(0.118)	(0.0998)	(0.104)
White	-0.174***	0.0166	-0.0234	-0.0471
	(0.0430)	(0.107)	(0.0512)	(0.0540)
Male	0.261***	0.244***	0.272***	-0.166***
	(0.0136)	(0.0283)	(0.0132)	(0.0441)
Level II	-0.0323**		-0.0494***	-0.0264**
	(0.0129)		(0.0126)	(0.0132)
Level III	-0.173***		-0.175***	-0.166***
	(0.0412)		(0.0396)	(0.0419)
Level IV	-0.862***		-0.764***	-0.842***
	(0.202)		(0.187)	(0.203)
Level NA	-0.231***		-0.260***	-0.242***
	(0.0344)		(0.0334)	(0.0351)
Constant	-9.860***		-7.595***	-9.060***
	(0.150)		(1.398)	(0.150)
Observations	660,921	656,424	680,489	660,921
Pseudo R^2	.2538	.2056	.2346	.2670

Clustered robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4-7: Mortality regressions by level of trauma center

	Level I Trauma Center	Level II Trauma Center	Level III/IV Trauma Center
	(1)	(2)	(3)
Early AM	0.0928***	0.138***	0.231**
	(0.0205)	(0.0300)	(0.108)
Night	0.110***	0.130***	0.171*
	(0.0173)	(0.0236)	(0.0880)
Age	0.0201***	0.0204***	0.0152***
	(0.000434)	(0.000584)	(0.00213)
Log(ISS)	1.975***	2.004***	1.667***
	(0.0183)	(0.0246)	(0.0893)
Comorbid	0.0542***	0.134***	0.0369
	(0.0105)	(0.0126)	(0.0480)
Black	0.387***	0.213***	0.585
	(0.0554)	(0.0795)	(0.442)
Hispanic	0.111*	0.0215	0.548
	(0.0581)	(0.0798)	(0.449)
Other	0.0581	-0.196**	-0.213
	(0.0671)	(0.0888)	(0.523)
White	-0.165***	-0.188**	-0.0379
	(0.0536)	(0.0741)	(0.429)
Male	0.240***	0.286***	0.291***
	(0.0173)	(0.0236)	(0.0890)
Constant	-9.893***	-8.436***	-8.077***
	(0.191)	(0.847)	(0.788)
Observations	374,577	235,515	23,851
Pseudo R^2	.2535	.2472	.1845

Clustered robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4-8: Mortality by hospital size and resource strain (number of beds, and number of visits per surgeon)

	Top Quartile By Number of Beds Mortality	**Bottom Quartile By Number of Beds Mortality**	**Bottom Quartile by Visits Per Surgeon Mortality**	**Top Quartile by Visits Per Surgeon Mortality**
	(1)	(2)	(3)	(4)
Early AM	0.120***	0.164***	0.0541	0.135***
	(0.0346)	(0.0500)	(0.0438)	(0.0403)
Night	0.121***	0.146***	0.0729**	0.0849**
	(0.0286)	(0.0393)	(0.0349)	(0.0338)
Age	0.0227***	0.0227***	0.0235***	0.0208***
	(0.000688)	(0.000942)	(0.000885)	(0.000855)
Log(ISS)	2.064***	2.364***	2.077***	2.117***
	(0.0211)	(0.0303)	(0.0387)	(0.0376)
Comorbid	0.0901***	0.130***	0.0398	0.0794***
	(0.0134)	(0.0188)	(0.0263)	(0.0239)
Black	0.321***	0.257	0.218	0.332***
	(0.0990)	(0.161)	(0.139)	(0.109)
Hispanic	-0.123	-0.0245	-0.157	-0.104
	(0.109)	(0.162)	(0.146)	(0.112)
Other	0.150	0.0443	-0.230	-0.0682
	(0.118)	(0.168)	(0.179)	(0.130)
White	-0.272***	-0.237	-0.238*	-0.228**
	(0.0969)	(0.153)	(0.134)	(0.105)
Male	0.254***	0.294***	0.261***	0.366***
	(0.0283)	(0.0382)	(0.0344)	(0.0342)
Constant	-590.9***	-11.10***	-10.45***	-9.206***
	(0.268)	(0.201)	(0.381)	(0.174)
Observations	165,028	129,047	95,470	94,203
Pseudo R^2	.2690	.2877	.2360	.2544

Clustered robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 4-9: Comparison of mortality rate between spine and femur patients

	Mortality Femurs	**Mortality Spines**
	(1)	(2)
Early AM	0.0146	0.148***
	(0.0891)	(0.0283)
Night	-0.0296	0.157***
	(0.0634)	(0.0222)
Comorbid	0.195***	0.0943***
	(0.0233)	(0.0129)
Log(ISS)	2.238***	2.551***
	(0.0575)	(0.0260)
Age	0.0372***	0.0343***
	(0.00190)	(0.000592)
Black	-0.000920	0.414***
	(0.271)	(0.0784)
Hispanic	-0.108	0.159*
	(0.287)	(0.0818)
Other	0.594	-0.219
	(0.552)	(0.216)
White	-0.315	0.228***
	(0.323)	(0.0885)
Male	0.277***	0.218***
	(0.0621)	(0.0214)
Constant	-11.34***	-12.22***
	(0.357)	(0.125)
Observations	31,714	154,868
Pseudo R^2	.1427	.2340

Clustered robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4-10: Further robustness checks

Variables	(All) (1)	(Shift Change) (2)	(Simple) (3)	(Complex) (4)	(Matching) (5)
Age	0.0245***	0.0254***	0.0345***	0.0258***	0.0247***
	-0.00047	-0.00102	-0.00105	-0.00046	(0.000564)
Early AM	0.0479**		-0.0182	0.0897***	0.0644**
	-0.0216		-0.0473	-0.0197	(0.0326)
Night	0.0568***	0.102***	0.00207	0.0802***	0.0789***
	-0.0177	-0.0342	-0.036	-0.0169	(0.0259)
ISS 1	-2.672***	-2.474***	-6.533***	-3.998***	-0.798***
…	-0.115	-0.254	-0.59	-1.171	(0.297)
ISS 66	3.147***	3.071***	-0.271	3.979***	4.479***
	-0.207	-0.479	-0.323	-1.076	(0.357)
Comorbidity Index 1	0.0907*	0.169*	0.274***	0.0967*	0.0723
…	-0.0465	-0.0955	-0.0899	-0.0514	(0.0558)
Comorbidity Index 12	1.539	1.851**	3.609***	1.521***	1.998
	-1.294	-0.817	-1.096	-0.573	(1.358)
Black	0.304***	0.0391	0.0815	0.374***	0.498***
	-0.0591	-0.141	-0.125	-0.0589	(0.0769)
Hispanic	-0.0563	-0.253*	-0.181	0.0505	0.192**
	-0.062	-0.147	-0.13	-0.0614	(0.0802)
Other	0.101	-0.152	0.274	-0.209	0.339**
	-0.127	-0.29	-0.242	-0.137	(0.170)
Asian	0.11	0.0321	0.0224	0.297***	0.235***
	-0.0697	-0.16	-0.149	-0.0655	(0.0899)
White	-0.00568	-0.218	-0.111	0.0644	0.116
	-0.0565	-0.136	-0.117	-0.0559	(0.0732)
Male	0.255***	0.277***	0.131***	0.188***	0.263***
	-0.0181	-0.0396	-0.0359	-0.0181	(0.0220)
Diagnosis 1	0.240***	0.274***	0.227***	0.191***	
…	-0.0278	-0.0612	-0.0522	-0.0185	
Diagnosis 250	-0.183	-0.162	-0.133	-0.343**	
	-0.154	-0.316	-0.106	-0.149	
Constant	-1.705***	-3.131***	0.205	-3.232***	-2.934***
	-0.661	-0.283	-0.244	-1.073	(0.747)
Observations	448,977	97,247	125,046	236,352	236,783

Clustered robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4-11: Waiting time

	Hours to Procedure	Fixed Effects Hours to Procedure	Prompt
	(1)	**(2)**	**(3)**
Early AM	-0.126***	-0.0958***	.192***
	(0.0111)	(0.0162)	(0.00906)
Night	-0.236***	-0.253***	0.198***
	(0.00898)	(0.0260)	(0.00750)
Age	0.0170***	0.0131***	-0.0201***
	(0.000217)	(0.000693)	(0.000177)
Log(ISS)	0.0710***	-0.0180	0.0514***
	(0.00415)	(0.0186)	(0.00347)
Comorbid	-0.0877***	0.0730***	0.0772***
	(0.00557)	(0.0122)	(0.00454)
Black	-0.0237	0.0555	.00662***
	(0.0366)	(0.0390)	(0.0325)
Hispanic	0.891***	0.0545	-.848**
	(0.0379)	(0.0387)	(0.0343)
Other	-0.431***	0.0425	-0.234***
	(0.0395)	(0.108)	(0.0352)
White	-0.0272	0.140***	-0.0854***
	(0.0356)	(0.0314)	(0.0317)
Male	-0.253***	-0.223***	0.242***
	(0.00861)	(0.0152)	(0.0317)
Constant	1.826***		1.616
	(0.0374)		(1.351)
Observations	314,844	314,844	314,844
Pseudo R^2	0.058	0.796	.0374
Regression Type	Logistic	Logistic	Logistic

Clustered robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4-12: Constant sample results

	Mortality	Hours to Procedure	Prompt	Fixed Effects Hours to Procedure	Fixed Effects Mortality
	(1)	(2)	(3)	(4)	(5)
Early AM	0.0847***	-0.111***	0.153***	-0.0993***	0.0842***
	(0.0223)	(0.0103)	(0.00971)	(0.0181)	(0.0233)
Night	0.0825***	-0.239***	0.161***	-0.259***	0.0858***
	(0.0180)	(0.00833)	(0.00814)	(0.0250)	(0.0184)
Age	0.0236***	0.0177***	-0.0140***	0.0139***	0.0246***
	(0.000460)	(0.000199)	(0.000199)	(0.000740)	(0.000966)
Comorbid	0.106***	-0.0872***	0.0655***	0.0696***	0.0944***
	(0.00907)	(0.00506)	(0.00489)	(0.0126)	(0.0157)
Log(ISS)	2.021***	0.0223***	0.0255***	-0.0547***	2.064***
	(0.0187)	(0.00390)	(0.00395)	(0.0190)	(0.0704)
Black	0.320***	-0.0220	0.0484	0.0847*	0.203**
	(0.0672)	(0.0343)	(0.0322)	(0.0451)	(0.0868)
Hispanic	-0.110	0.892***	-0.821***	0.0751*	-0.0304
	(0.0714)	(0.0356)	(0.0342)	(0.0426)	(0.0913)
Other	0.0226	-0.366***	0.317***	0.125	-0.0727
	(0.0749)	(0.0371)	(0.0349)	(0.182)	(0.0878)
White	-0.197***	0.0136	-0.0515	0.159***	-0.129
	(0.0652)	(0.0333)	(0.0314)	(0.0398)	(0.0863)
Male	0.255***	-0.278***	0.239***	-0.241***	0.235***
	(0.0181)	(0.00798)	(0.00788)	(0.0147)	(0.0331)
Level II	-0.0313*	0.116***	-0.133***		
	(0.0173)	(0.00775)	(0.00755)		
Level III	-0.260***	0.279***	-0.455***		
	(0.0596)	(0.0202)	(0.0217)		
Level IV	-0.517	1.759***	-2.095***		
	(0.347)	(0.0665)	(0.191)		
Level NA	-0.200***	0.797***	-0.989***		
	(0.0534)	(0.0213)	(0.0274)		
Constant	-10.17***	1.840***	1.188		
	(0.185)	(0.0350)	(1.3664)		
Observations	366,823	366,823	366,823	366,823	366,826
Pseudo R^2	.2584	0.059	.0669	0.797	.2519
Regression Type	Logistic	OLS	Logistic	OLS	Logistic

Clustered robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

CHAPTER 5: THE IMPACT OF RESIDENTS ON EMERGENCY DEPARTMENT EFFICIENCY

5.1 Introduction

One possible source of inefficiency in hospitals is the residency teaching model in hospital emergency departments (EDs). After medical students complete medical school, they become doctors. New doctors must spend three to six years as a resident, treating patients under the supervision of attending physicians. Residents have two roles in the hospital: doctor and student. They treat patients as doctors and are observed and taught by senior attending physicians. Attending physicians teach residents and treat patients.

These dual roles — doctors who are also students, and doctors who are also teachers — obscure the effect that the residency model has on hospital efficiency. Because they treat patients, the use of residents should help to lower treatment and waiting times for patients. However, the time that attending physicians spend teaching and supervising residents takes away from the time they can devote to the direct treatment of patients. In this chapter, we study the tradeoff between the time residents spend treating patients on the one hand, and the time they take from attending physicians on the other hand. In Section 5.2, we review the relevant literature. In Section 5.3, we discuss the data. In Sections 5.4 and 5.5, we present our analysis and discuss the results. In Section 5.6, we discuss the limitations of the work and conclusions are presented in Section 5.7.

5.2 Literature Review

Operations management can help hospitals improve efficiency and consequently provide better service and increase profit (O'Neill and Dexter, 2005; Sarkis and Talluri, 2002; Swisher and Jacobson, 2002). Hollingsworth gives a summary of much of the literature examining hospital efficiency (Hollingsworth, 2003). In this study, we use ED length of stay (LOS) as the primary measure. LOS is a commonly used ED efficiency metric (Chan and Kass, 1999; Fineberg and Stewart, 1977).

We focus on how residents impact efficiency in the ED. This is an important question in the medical community which has serious policy and operational implications. Medicare reimbursement rates consider the direct and indirect costs of training residents (Rosko, 1996). Medicare assumes that having residents present significantly increases the cost of care, and, thus, increases reimbursement rates to hospitals that train residents. It has been argued that Medicare reimbursement rates overcompensate for the costs of training residents (Anderson and Lave, 1986; Custer and Wilke, 1991; Rogowski and Newhouse, 1992; Welch, 1987). The effect residents have on hospital efficiency is an indirect cost (or benefit) to the hospital and should be considered when setting Medicare reimbursement rates.

There are two competing hypotheses about the effect of residents on efficiency. One claim is that the presence of residents increases faculty staffing requirements, as attending physicians are required to spend time supervising and instructing the residents (DeBehnke, 2001). On the other hand, Knickman et al.

(1992) argue that teaching and treatment can occur simultaneously, meaning that residents can help to improve throughput.

Some recent empirical work has tried to quantify the effects that residents have on efficiency in the hospital, but the literature is inconclusive. Harvey et al. (2008) review ED patient waiting times, time until an admission decision was made, and total ED length of stay during periods when residents were on strike versus times of normal resident staffing patterns at a hospital in New Zealand. They find that without residents, the ED has higher throughput and the length of stay is reduced, meaning that residents slow down treatment. However, the total number of hours worked per week by doctors at the hospital during the strike decreased only 10 hours, from 236 to 226. This means that some of the work that residents would have done was performed by more senior doctors during the strike period. Similarly, Salazar et al. (2001) observe the effects of a resident strike on quality and throughput in an ED at a large teaching hospital. They find that replacing residents with staff physicians leads to an increase in throughput and in quality of care. Lammers et al. (2003) examine the effect of adding residents to an ED at a community hospital, and find that there is a weak, positive correlation between ED patient length of stay and the presence of residents, meaning that residents had a detrimental effect on ED efficiency. The authors note that, in addition to supervising residents, attending physicians saw all patients, repeated parts or all of the examinations, reviewed medical histories, and were present for procedures.

Other work has shown that residents have a more positive effect on hospital operations. Theokary et al. (2011) study the effects that residents have on service quality at teaching hospitals. They conclude that residents help to increase the quality of service, especially in small to medium-sized hospitals. Blake and Carter (1996) find that patient waiting times are affected by the amount of time attending physicians spend teaching residents and the amount of time that residents spend treating patients. After the introduction of residents to an anesthesiology ward, Eappen et al. (2004) found no significant adverse effects, either economically or on patient outcomes. Offner et al. (2003) study the addition of residents to a trauma care center and conclude that residents improve efficiency while having no effect on the quality of care. The added residents perform surgeries and contribute to the direct treatment of patients.

Huckman et al. (2005) study how cohort turnover affects hospital operations. They find that the influx of new residents coupled with the graduation of the most experienced residents lead, not surprisingly, to longer treatment times and lower throughput. Dowd et al. (2005) study the efficiency of residents as they gain experience. They find that as residents become more experienced they become more autonomous, are able to provide more care, thus helping to increase throughput.

The literature has identified a clear efficiency tradeoff presented by residents. They provide care to patients, but also require attention from attending physicians. Residents seem to provide a net benefit when they are allowed to provide significant amounts of care to patients. However, when they are mainly being taught by attending physicians, their presence decreases throughput.

5.3 Data

We were motivated by the inconclusive literature to further study the effect that residents have on efficiency in the ED. We observed a natural experiment at the University of Maryland Medical Center (UMMC), in which the residents were required to go to a research seminar every Wednesday morning, and thus were absent from the ED during this time period. Residents were present in the ED at all other times. Typically there are two attending physicians on duty, one senior resident, one first year resident, and two more residents of intermediate experience. There were no other changes made to the ED staffing to compensate for the absence of the residents. No other doctors were assigned to the ED and no additional staff were hired to replace the absent residents. We discussed how resident presence affects operations in the ED with physicians from UMMC. They said that when residents are present in the ED, attending physicians perform in a managerial role, supervising care and instructing the residents, and almost all of the hands-on care to patients is provided by the residents. However, when residents are absent, attending physicians become the primary provider of hands-on care. The physicians also said that there are no other changes in their peripheral duties (paperwork, charting, etc.). The only change between Wednesday morning and the rest of the week is that when the residents are absent, the attendings switch from a supervisory role to one of actively providing care.

By comparing treatment times of patients on Wednesday mornings (when there were no residents) to the rest of the week (when residents were present), we can make inferences about the effect that residents have on treatment times (and consequently throughput), assuming patients who arrive on Wednesday mornings are similar to patients from the rest of the week. Because residents do almost all of the hands-on patient care, we assume that every patient is treated by a resident unless they are first treated when residents are absent.

While treatment times are not the only measure of efficiency, we do not have sufficient outcome data to measure quality of care. We use the difference in average treatment times between Wednesday mornings and other times of the week to measure the impact that residents have on possible ED throughput. The patients who arrived at the ED during the seminars on Wednesday mornings were similar in severity to the patients seen throughout the rest of the week. A Kolmogorov-Smirnov test comparing the distributions of patient severity between Wednesday mornings and the rest of the week fails to reject the hypothesis that the distributions are the same ($p = .206$). On Wednesday mornings, 74% of patients required labs and 67% required radiology tests, compared with 76% and 63% during the rest of the week, respectively. The arrival rate of patients was similar, as well. Figure 5-1 shows a plot of the arrival rates of patients for different days of the week.

We analyzed patient who visited the UMMC ED between October 1, 2009 and January 31, 2010. For each patient, we were given information about treatment characteristics and severity information. From this data, we derived metrics describing the state of the ED, including congestion, and whether or not residents were present. We only analyze patients who were treated in the ED; patients who leave the waiting room before being seen and those who were routed to the ambulatory zone by the triage nurse were excluded. The ambulatory zone was designed to provide a faster service for less severe patients. These patients are seen once, treated, discharged quickly, and typically not seen by residents. Our final data set had 7,935 patients. Table 5-1 gives a summary of the variables that we were given. Each variable is integer-valued.

5.4 Analysis

First, we analyzed the two distributions of treatment times ─ for patients treated by residents and for those not treated by residents. We define treatment time as the time from when a patient is first placed in a bed to when he is either discharged or admitted to the hospital. The distributions of treatment when residents are present and absent are shown in Figure 5-2.   A Kolmogorov-Smirnov test comparing the two distributions shows with a *p*-value of .023 that the distributions are different. We see that the treatment times when residents were absent tend to be slightly higher than those when residents were present. The median treatment time for a patient treated by residents is 6.15 hours, while the median treatment time for those not treated by a resident is 7.11 hours. The standard deviation of treatment times when residents were present was 6.54 hours, compared to 7.35 when residents were absent. An F test showed these two variances to be different at the 1% confidence level (p = .0078).

Based on this comparison between the two treatment time distributions, we construct regression models to test what effect residents have on treatment times in the ED. We regress the natural log of treatment times on the state of the ED (number of people waiting for treatment, weekday vs. weekend), patient characteristics (severity score, labs and radiology tests needed), and if residents were present. Because the treatment times are so heavily skewed, we take the log transform for both distributions when doing the analysis. The hypothesized regression equation is:

$$\ln(\mathit{Treatment\ Time}) = \beta_0 + \beta_1 * \mathit{NoRes} + \beta_2 * \mathit{Line} + \beta_3 * \mathit{Labs} + \beta_4 * \mathit{NumLabs} + \beta_5 * \mathit{Rad} + \beta_6 * \mathit{NumRad} + \beta_7 * \mathit{Weekend} + \beta_8 * \mathit{Admit} + \beta_9 * \mathit{Sev1} + \beta_{10} * \mathit{Sev2} + \beta_{11} * \mathit{Sev3} + \beta_{12} * \mathit{Sev4} + \beta_{13} * \mathit{Sev5} + \varepsilon ,$$

where *SevX* are dummy variables that are 1 if the patient is of severity X, and 0 otherwise. The baseline patient, when all dummies are 0, is a patient treated by residents during the week, of NA severity, with no lab or radiology tests needed. Table 5-2 shows the results of this regression.

These results provide insights into factors affecting the length of stay of patients in the ED. Importantly, we see according to this model that the absence of residents increases treatment times by 7.8% (exp (.075) ≈ 1.078). A patient not treated by residents will, on average, have 7.8% longer treatment times than a patient who is treated by residents, all else equal. This effect is strong and statistically significant. This is evidence that contradicts our original conjecture that residents will slow down treatment in the ED and have a negative effect on efficiency.

We also see that having lab or radiology tests greatly increases the treatment time, by 40% (exp (.335) ≈ 1.40) or 16% (exp (.148) ≈ 1.16), respectively. Each additional lab or radiology test has only a minor (though highly statistically significant) incremental impact on the treatment time; since tests are typically run in parallel, we did not expect a large effect from the number of tests. As expected, low severity patients (severity 4-5) have much shorter treatment times than do high severity patients. Similarly, patients who are admitted to the hospital after their time in the ED stay 9.2% (exp (.088) ≈ 1.092) longer in the ED than those who are discharged and sent home. Patients who are eventually admitted are typically higher severity cases, regardless of the triage score and will take longer to treat. Though the model also found that patients with severity 1 tend to have shorter treatment times, this result is statistically insignificant and likely due to the fact that only 29 patients received this severity score. We also see that the more patients there are in the waiting room, i.e., the more congested the ED is, the longer treatment takes. This increase in treatment time could arise from resource shortages or increased demands on healthcare workers.

Next, we examined how residents affect treatment times for different types of patients. For example, residents might play different roles in treating high severity patients and low severity patients. We split the data set into two groups, high severity and low severity, and ran the regressions on both groups. We include patients with no severity score (severity NA) in the high severity group, although their exclusion does not significantly alter the results. Looking at just high severity patients (severity 1-3 and NA), we see that residents have a similar effect. The results of the regression on high severity patients are given in Table 5-3. Again, we see that residents decrease the treatment time of patients by 7.6% (exp (.073) ≈ 1.076) and that this effect is again statistically significant. The rest of the results are similar. Lab and radiology tests, being admitted upon discharge, and congestion all lengthen treatment time.

However, when looking at low severity patients in Table 5-4, we do not see the same effect. When we run the same regression on the low severity patients (triage score 4-5), the coefficient for NoRes is not statistically significant (p = .562). Therefore, unlike in predictions across the entire patient population or for just high severity patients, where the presence of residents reduces treatment times, residents have no statistically significant effect on treatment times of low severity patients. Patients being admitted upon discharge and radiology tests being performed also lost statistical significance in this regression; because only 33 low-severity patients were admitted after treatment, this variable losing significance is not surprising. In this regression model, the baseline patient is the same as in the previous models, except he has a severity score of 5, because no patients with NA severity are included in this population. The distributions of treatment times, split by resident presence, are shown in Figure 5-3.

The difference in the effects that residents have on high severity patients and low severity patients is interesting. While residents have a strong effect on lowering treatment times for high severity patients, they have no significant effect on low severity patients. It may be that there is more work to be done on high severity patients, so having extra healthcare workers around is advantageous. However, on low severity cases, where treatment is fairly routine, the time taken by residents for instruction is enough to outweigh the extra work that they do.

We also examine the treatment times of patients who begin treatment during the hours of 7:00 a.m. to 1:00 p.m. (the hours of the Wednesday seminar). By looking at just these patients, we are able to limit time-of-day effects on patient types, or on the state of the hospital. If lab tests come back slower in the afternoon because there is more demand from elsewhere in the hospital, this might show up as patients being treated faster when residents are present. By examining just patients treated in the morning, we are better able to isolate the effect that residents have on treatment times. In other words, there might be some difference in the hospital operations between the mornings and the rest of the day. By only including patients who arrived in the morning in the analysis, we are better able to isolate the effect that residents have on treatment times. We ran the regression again on this restricted data set to see if the effect holds when looking just at these "morning" patients. Because there are now a smaller number of observations, instead of separating the patients into the five severity dummies, we group them into high and low severity. The baseline patient is the same as in the first model, except he is a low severity patient in this model. Table 5-5 shows these results. The treatment time distributions are also given in Figure 5-3.

Again, we see that residents have a strong effect. In this model, treatment times are 7.0% ($\exp(.068) \approx 1.070$) longer when residents are absent. The rest of the control variables have effects similar to those in the original model. Lab and radiology tests significantly slow down treatment and higher severity patients take longer to treat. Congestion again has a small effect in increasing treatment times. This model gives us further evidence that residents do reduce treatment times. We have now seen statistically significant evidence across a variety of models that residents lower treatment times, especially among high severity patients.

5.5 Survival Analysis

We first analyzed the two distributions of treatment times ─ for patients treated by residents and for those not treated by residents. We define treatment time as the time from when a patient is first placed in a bed to when he is either discharged or admitted to the hospital. The distributions of treatment when residents are present and absent are shown in Figure 5-2. A Kolmogorov-Smirnov test comparing the two distributions shows with a *p*-value of .023 that the distributions are different. We see that the treatment times when residents were absent tend to be slightly higher than those when residents were present. The median treatment time for a patient treated by residents is 6.15 hours, while the median treatment time for those not treated by a resident is 7.11 hours. The standard deviation of treatment times when residents were present was 6.54 hours, compared to 7.35 when residents were absent. An F test showed these two variances to be different at the 1% confidence level ($p = .0078$).

Based on this comparison between the two treatment time distributions, we construct regression models to test what effect residents have on treatment times in the ED. We regress the natural log of treatment times on the state of the ED (number of people waiting for treatment, weekday vs. weekend), patient characteristics (severity score, labs and radiology tests needed), and if residents were present. Because the treatment times are so heavily skewed, we take the log transform for both distributions when doing the analysis. The hypothesized regression equation is:

$$\ln(\mathit{Treatment\ Time}) = \beta_0 + \beta_1 * \mathit{NoRes} + \beta_2 * \mathit{Line} + \beta_3 * \mathit{Labs} + \beta_4 * \mathit{NumLabs} + \beta_5 * \mathit{Rad} + \beta_6 * \mathit{NumRad} + \beta_7 * \mathit{Weekend} + \beta_8 * \mathit{Admit} + \beta_9 * \mathit{Sev1} + \beta_{10} * \mathit{Sev2} + \beta_{11} * \mathit{Sev3} + \beta_{12} * \mathit{Sev4} + \beta_{13} * \mathit{Sev5} + \varepsilon ,$$

where *SevX* are dummy variables that are 1 if the patient is of severity X, and 0 otherwise. The baseline patient, when all dummies are 0, is a patient treated by residents during the week, of NA severity, with no lab or radiology tests needed. Table 5-2 shows the results of this regression.

These results provide insights into factors affecting the length of stay of patients in the ED. Importantly, we see according to this model that the absence of residents increases treatment times by 7.8% ($\exp(.075) \approx 1.078$). A patient not treated by residents will, on average, have 7.8% longer treatment times than a patient who is treated by residents, all else equal. This effect is strong and statistically significant. This is evidence that contradicts our original conjecture that residents will slow down treatment in the ED and have a negative effect on efficiency.

We also see that having lab or radiology tests greatly increases the treatment time, by 40% ($\exp(.335) \approx 1.40$) or 16% ($\exp(.148) \approx 1.16$), respectively. Each additional lab or radiology test has only a minor (though highly statistically significant) incremental impact on the treatment time; since tests are typically run in parallel, we did not expect a large effect from the number of tests. As expected, low severity patients (severity 4-5) have much shorter treatment times than do high severity patients. Similarly, patients who are admitted to the hospital after their time in the ED stay 9.2% ($\exp(.088) \approx 1.092$) longer in the ED than those who are discharged and sent home. Patients who are eventually admitted are typically higher severity cases, regardless of the triage score and will take longer to treat. Though the model also found that patients with severity 1 tend to have shorter treatment times, this result is statistically insignificant and likely due to the fact that only 29 patients received this severity score. We also see that the more patients there are in the waiting room, i.e., the more congested the ED is, the longer treatment takes. This increase in treatment time could arise from resource shortages or increased demands on healthcare workers.

Next, we examined how residents affect treatment times for different types of patients. For example, residents might play different roles in treating high severity patients and low severity patients. We split the data set into two groups, high severity and low severity, and ran the regressions on both groups. We include patients with no severity score (severity NA) in the high severity group, although their exclusion does not significantly alter the results. Looking at just high severity patients (severity 1-3 and NA), we see that residents have a similar effect. The results of the regression on high severity patients are given in Table 5-

3. Again, we see that residents decrease the treatment time of patients by 7.6% (exp (.073) ≈ 1.076) and that this effect is again statistically significant. The rest of the results are similar. Lab and radiology tests, being admitted upon discharge, and congestion all lengthen treatment time.

However, when looking at low severity patients in Table 5-4, we do not see the same effect. When we run the same regression on the low severity patients (triage score 4-5), the coefficient for NoRes is not statistically significant (p = .562). Therefore, unlike in predictions across the entire patient population or for just high severity patients, where the presence of residents reduces treatment times, residents have no statistically significant effect on treatment times of low severity patients. Patients being admitted upon discharge and radiology tests being performed also lost statistical significance in this regression; because only 33 low-severity patients were admitted after treatment, this variable losing significance is not surprising. In this regression model, the baseline patient is the same as in the previous models, except he has a severity score of 5, because no patients with NA severity are included in this population. The distributions of treatment times, split by resident presence, are shown in Figure 5-3.

The difference in the effects that residents have on high severity patients and low severity patients is interesting. While residents have a strong effect on lowering treatment times for high severity patients, they have no significant effect on low severity patients. It may be that there is more work to be done on high severity patients, so having extra healthcare workers around is advantageous.

However, on low severity cases, where treatment is fairly routine, the time taken by residents for instruction is enough to outweigh the extra work that they do.

We also examine the treatment times of patients who begin treatment during the hours of 7:00 a.m. to 1:00 p.m. (the hours of the Wednesday seminar). By looking at just these patients, we are able to limit time-of-day effects on patient types, or on the state of the hospital. If lab tests come back slower in the afternoon because there is more demand from elsewhere in the hospital, this might show up as patients being treated faster when residents are present. By examining just patients treated in the morning, we are better able to isolate the effect that residents have on treatment times. In other words, there might be some difference in the hospital operations between the mornings and the rest of the day. By only including patients who arrived in the morning in the analysis, we are better able to isolate the effect that residents have on treatment times. We ran the regression again on this restricted data set to see if the effect holds when looking just at these "morning" patients. Because there are now a smaller number of observations, instead of separating the patients into the five severity dummies, we group them into high and low severity. The baseline patient is the same as in the first model, except he is a low severity patient in this model. Table 5-5 shows these results. The treatment time distributions are also given in Figure 5-3.

Again, we see that residents have a strong effect. In this model, treatment times are 7.0% ($\exp(.068) \approx 1.070$) longer when residents are absent. The rest of the control variables have effects similar to those in the original model. Lab and radiology tests significantly slow down treatment and higher severity patients take longer to treat. Congestion again has a small effect in increasing treatment times. This model gives us further evidence that residents do reduce treatment times. We have now seen statistically significant evidence across a variety of models that residents lower treatment times, especially among high severity patients.

5.6 Discussion

We have shown that residents decreased treatment times at the UMMC ED, and that effect is particularly pronounced when treating high severity patients. This is fortunate, because the main reason that residents are in the ED is to learn how to treat patients, and they learn more when working on more complex, higher severity cases. This indicates that the best use of residents, both for ED efficiency and for the education of residents, is to have them treat high severity cases.

With new Accreditation Council of Graduate Medical Education rules restricting residents' maximum weekly working hours to 80, it is becoming more important to prioritize the cases on which residents work (Philibert, 2002). Our work suggests that residents be assigned to the highest acuity cases in the ED, as residents both learn more from these cases and contribute more to the efficiency of the hospital.

After the conclusion of our study, changes in patient routing decisions at UMMC have taken this approach to patient care in the ED. They have started to route more of the lowest severity cases to an ambulatory zone. Because there are typically no residents in the ambulatory zone, this has the effect of raising the severity level of the patients seen by residents, so that they are, on average, treating higher acuity patients. Our results sometimes conflict with those in other papers in the literature. We propose three explanations. First, many of the other hospitals studied replaced residents either with nurses or with more senior physicians. Our work is the only one that has a true *ceteris paribus* experiment, in which residents are removed from the ED and no other changes are made. In the other papers, there are either staffing changes or effects are measured over the course of several years, where other changes in hospital conditions could impact the results. Second, we believe residents have a greater effect on treatment times on patients with more severe problems; in these cases, more things can be done in parallel. Third, residents at UMMC play an active role in treating patients and are somewhat autonomous. By having residents provide substantial amounts of care, they help to increase throughput enough to offset the time that attending physicians must spend supervising and teaching them. Variation in patient severity mixes between hospitals could also play a role.

5.7 Limitations and External Validity

The data imposed a few limitations on this study. We only have data from one department at one hospital over the course of four months. We also do not have outcome data on the patients or any way to measure quality of care. We suspect that our results are applicable to other EDs across the U.S. where residents play a similar role, but we cannot assert this with certainty. Discussions with ED physicians lead us to believe that our results should be applicable to other hospitals, especially large, urban teaching hospitals like UMMC. Though it would have been best to have similar data from multiple hospitals, the unique nature of the natural experiment observed at UMMC prevents us from performing the same sort of analysis at multiple hospitals. Whether our findings hold up for other departments in the same hospital and other hospitals is an open question. We believe that the impact the residents have on treatment times is a function of how much hands-on care they provide to patients. When they are allowed to contribute, especially autonomously (i.e., more experienced residents), they can significantly increase throughput.

5.8 Conclusions and Future Work

In this work, we have shown that residents can help to reduce emergency department treatment times. This occurs when the work residents do treating patients outweighs the time attending physicians spend teaching them, an effect that is pronounced when residents are treating high severity patients. Other studies have found that residents impair efficiency, but we have shown that, in some cases, residents can help to reduce treatment times. We suggest, that to maximize efficiency in an ED, residents should be allowed to provide as much hands-on care as they are capable of, especially to high-severity patients. In future work, we hope to examine similar data from other major hospitals that have residents in the ED. With more detailed data, we could examine how residents affect treatment times in greater detail. For example, if we knew which residents treated which patients, we could study the difference in effect between younger and more experienced residents.

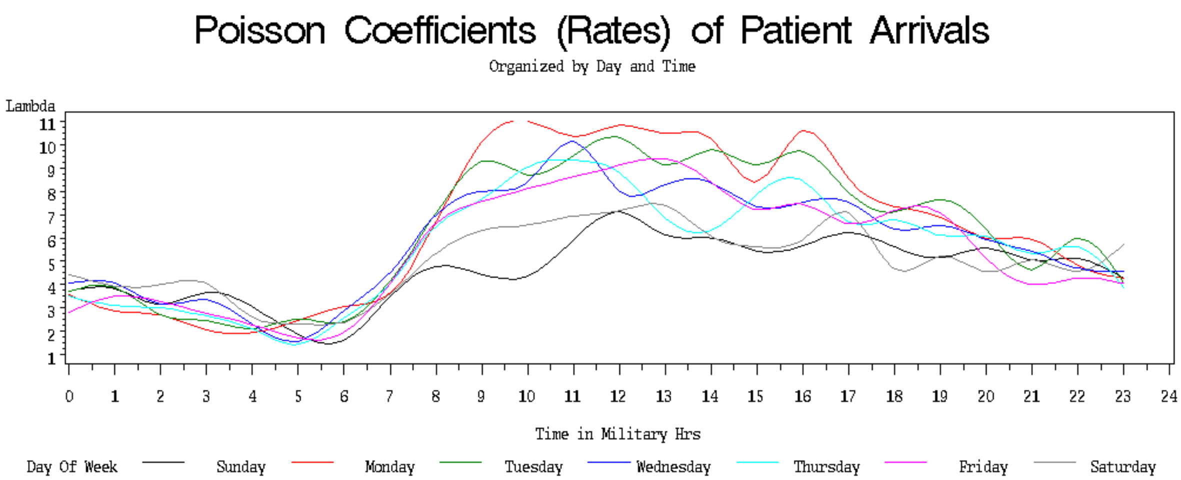

Figure 5-1: Arrival rates by day of week and time of day

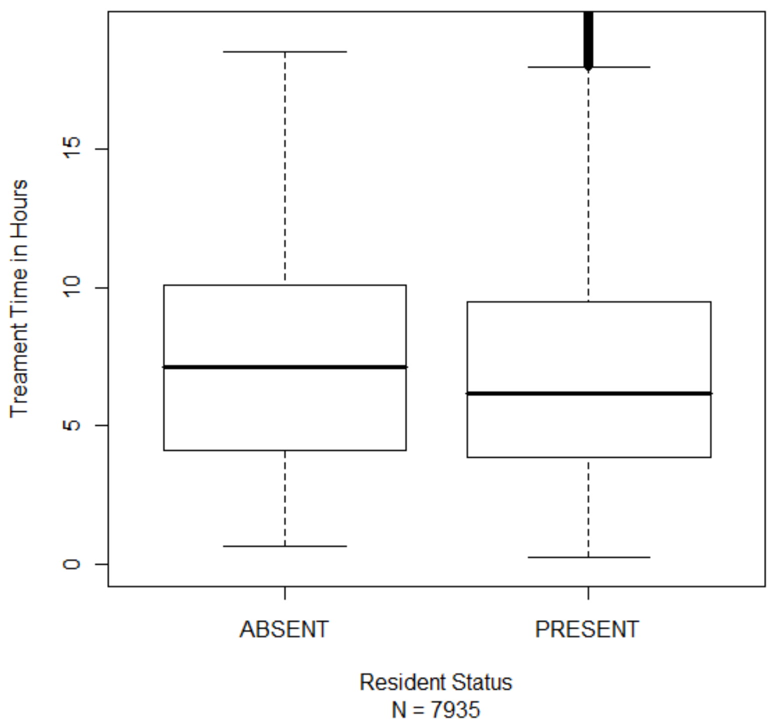

Figure 5-2: Treatment times for patients treated based on resident presence

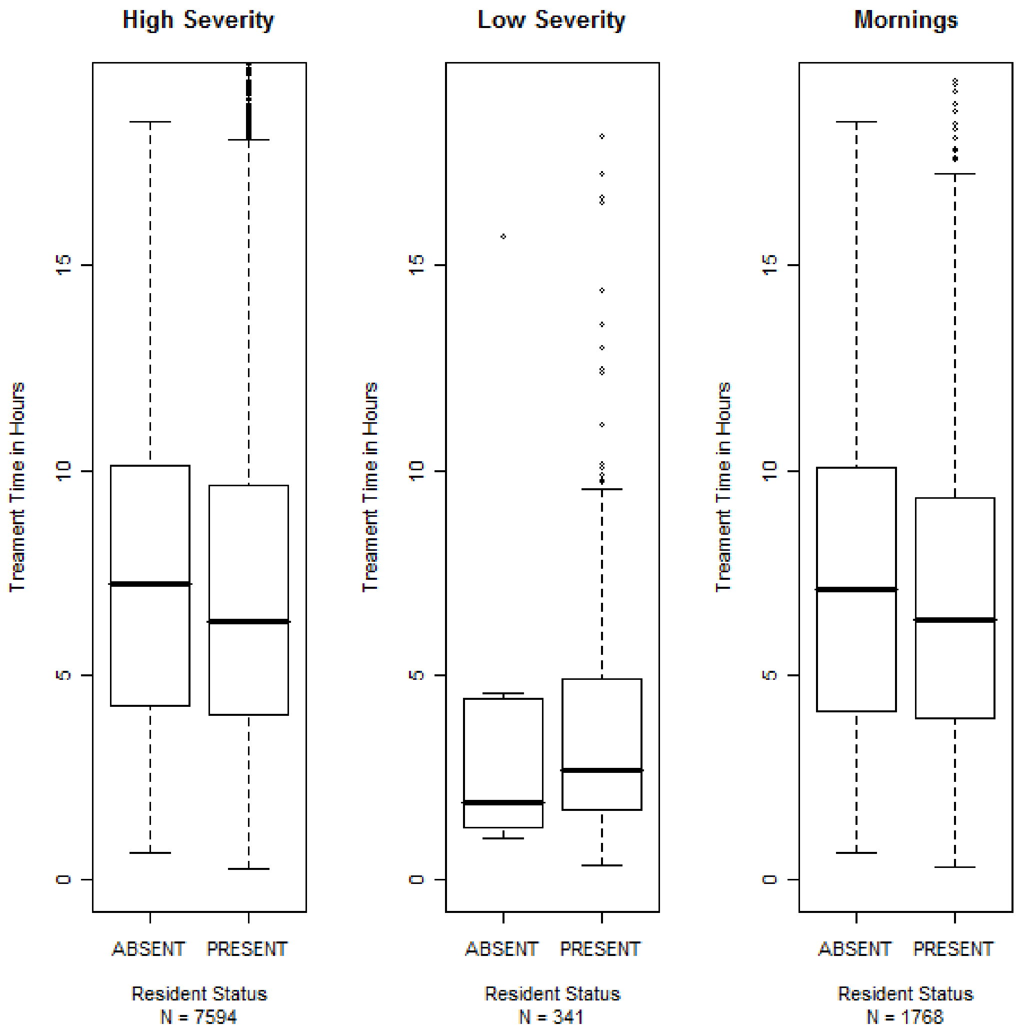

Figure 5-3: Treatment times for patients based on resident presence for high severity, low severity, and morning patients

Table 5-1: Variable descriptions

Variable	Description	Range
NoRes	Dummy variable that is 1 for all patients first treated on Wednesday mornings (when residents are absent)	[0,1]
Line	The number of patients in the waiting room, used as a measure of congestion	[0,28]
Admit	Dummy variable that is 1 if the patient was admitted as an inpatient upon being discharged from the ED and 0 if he/she was sent home.	[0,1]
Numlab	The number of lab tests the patient had	[0,97]
Labs	Dummy variable that is 1 if the patient had any labs at all	[0,1]
Numrad	Number of radiology tests the patient had	[0,19]
Rad	Dummy variable that is 1 if the patient had any radiology tests at all	[0,1]
Weekend	Dummy variable that is 1 if the patient arrived on Saturday or Sunday	[0,1]
Night	Dummy variable that is 1 if the patient arrived during the night shift (11 p.m. to 7 a.m.)	[0,1]
Severity	The severity score given to the patient by the triage nurse, with 1 being the most severe. Patients arriving by ambulance, or otherwise not receiving a score are given NA.	[1,5] or NA
Treatment Time	The time, in hours, from first being placed in a bed until the patient is either discharged or admitted to the hospital	[0.15,23]

Table 5-2: Regression results on all patients
(Adjusted R2 = .5355, N = 7935)

Variable	Coefficient	Std. Error	t-value	*p*-value
(Intercept)	5.002	0.020	247.475	<.001
NoRes	0.075	0.034	2.242	0.025
Line	0.010	0.002	5.455	<.001
Admit	0.088	0.015	5.819	<.001
NumLab	0.032	0.001	35.847	<.001
Labs	0.335	0.018	18.716	<.001
NumRad	0.057	0.004	13.509	<.001
Rad	0.148	0.016	9.376	<.001
Weekend	-0.044	0.013	-3.311	<.001
Sev1	-0.148	0.096	-1.544	0.123
Sev2	0.048	0.017	2.730	0.006
Sev3	0.031	0.015	2.080	0.038
Sev4	-0.178	0.032	-5.511	<.001
Sev5	-0.543	0.090	-6.001	<.001

Table 5-3: Regression results on high severity patients
(Adjusted R2 = .5133, N = 7549)

Variable	Coefficient	Std. Error	t-value	*p*-value
(Intercept)	5.027	0.020	245.581	<.001
NoRes	0.073	0.034	2.138	0.033
Line	0.009	0.002	4.784	<.001
Admit	0.090	0.015	5.955	<.001
Numlab	0.032	0.001	35.832	<.001
Labs	0.316	0.018	17.242	<.001
Numrad	0.056	0.004	13.331	<.001
Rad	0.143	0.016	8.881	<.001
Weekend	-0.055	0.014	-4.010	<.001
Sev1	-0.146	0.095	-1.528	0.126
Sev2	0.049	0.017	2.828	0.005
Sev3	0.029	0.015	1.987	0.047

Table 5-4: Low severity patients results
(Adjusted R2 = .5737, N = 341)

Variable	Coefficient	Std. Error	t-value	*p*-value
(Intercept)	4.234	0.104	40.558	<.001
NoRes	0.110	0.189	0.581	0.562
Line	0.041	0.011	3.711	<.001
Admit	0.010	0.127	0.081	0.935
Numlab	0.035	0.007	4.899	<.001
Labs	0.553	0.087	6.324	<.001
Numrad	0.133	0.037	3.610	<.001
Rad	0.144	0.093	1.559	0.120
Weekend	0.135	0.062	2.183	0.030
Sev4	0.281	0.099	2.834	0.005

Table 5-5: Morning only results
(Adjusted R2 = .5712, N = 1768)

Variable	Coefficient	Std. Error	t-value	*p*-value
(Intercept)	4.630	0.055	84.908	<.001
NoRes	0.068	0.034	2.008	0.045
Line	0.023	0.006	3.792	<.001
Admit	0.146	0.031	4.669	<.001
Numlab	0.030	0.002	15.628	<.001
Labs	0.328	0.038	8.750	<.001
Numrad	0.054	0.009	5.901	<.001
Rad	0.188	0.033	5.763	<.001
HighSev	0.345	0.054	6.359	<.001

Table 5-6: Survival analysis results

Variable	Coefficient	Standard Error	*z*-value	*p*-value
NoRes	-0.2505	0.0860	-2.9140	0.0036
Numlab	0.0037	0.0055	0.6680	0.5044
Numrad	-0.0358	0.0254	-1.4090	0.1587
Labs	-0.6133	0.1067	-5.7490	0.0000
Rad	-0.2198	0.0905	-2.4290	0.0152
Line	0.0327	0.0104	3.1310	0.0017
Sev1	0.6403	0.4540	1.4100	0.1585
Sev2	-0.0447	0.1023	-0.4370	0.6622
Sev3	-0.1140	0.0836	-1.3640	0.1725
Sev4	-0.0320	0.1932	-0.1660	0.8685
Sev5	0.6317	0.5864	1.0770	0.2814

CHAPTER 6: SIMULATING THE EFFECT OF RESIDENTS ON THE EMERGENCY DEPARTMENT

6.1 Introduction

In this chapter, we use a simulation approach to determine the effect of residents on emergency department (ED) efficiency. Instead of building statistical models of treatment times, we designed and implemented a model to directly simulate the ED. A simulation model allows flexibility in designing experiments and greater exploration of the mechanism through which residents impact efficiency. It also allows us to measure the effect on throughput and average waiting time, instead of simply examining treatment times. In Section 6.2, we review the relevant literature. In Section 6.3, we discuss our data and provide a detailed description of the simulation model. Validation of the model is given in Section 6.4. In Section 6.5, we discuss the results and implications. The conclusions are presented in Section 6.6.

6.2 Literature Review

In this section, we discuss studies about the effects of residents on ED efficiency. The resident education model creates a dual role for attending physicians in the ED, because a resident's role includes both treating patients and learning medicine. Thus, the resident care model can affect patient throughput because of the additional time spent on instruction.

Recent research has found that residents do decrease efficiency in hospital settings. In one study, researchers aim to review ED patient waiting times, time until an admission decision was made, and total ED length of stay during periods when residents were on strike versus times of normal resident staffing patterns (Harvey et al., 2008). They find that without residents, the ED had higher throughput and the length of stay was reduced. Lammers et al. (2008) examine the effect of adding residents to an ED at a community hospital. They conclude that there is a weak, positive correlation between ED patient length of stay and the presence of residents. Dowd et al. (2008) study the efficiency of residents as they gain experience. They find that as residents become more experienced they increase their throughput. Salazar et al. (2008) observe the effects of a resident strike on quality and throughput in an ED at a large teaching hospital. They determine that replacing residents with staff physicians leads to an increase in throughput and in quality of care.

Other studies, however, show that residents have no negative effects on throughput or treatment times. Eappen et al. (2004) look at the introduction of anesthesiology residents to surgical wards. They find no significant adverse economic or health effects. Offner et al. (2003) study the addition of residents to a trauma care center and conclude that residents improve efficiency while having no effect on the quality of care.

Methodologically, our work relies on simulation modeling and queueing theory. These methods are used extensively in the hospital operations management literature. Jun et al. (1999), Fone et al. (2003), Jacobson et al. (2006), and Brailsford et al. (2009) provide surveys of simulation models used in healthcare research. Simulation has a wide variety of applications in healthcare, such as modeling patient flow (Ceglowski et al. 2007), optimizing resource allocation (Lehaney and Hlupic 1995), and evaluating surgery scheduling strategies (Dexter et al. 2000).

Queueing theory is another technique widely used in the hospital operations management literature. Green (2006) and Fomundam and Herrmann (2007) provide surveys of applications of queueing theory to healthcare problems. For example, queueing theory has been used in the emergency department to determine appropriate staffing levels in order to reduce the proportion of patients who leave without being seen (Green et al. 2006) and to assist in bed management planning (Gorunescu et al. 2002).

6.3 Data and Simulation Model

We were motivated by the inconclusive literature to study whether residents help or hurt efficiency in the ED. At the UMMC, every Wednesday morning there was a seminar that the residents had to attend, so they were not present in the ED. Because of this, patients who were treated on Wednesday mornings were not seen by a resident, but only by attending physicians. This observation (residents present vs. not) suggests a natural experiment to determine what effect removing residents would have during other parts of the week. We designed a simulation model to exploit this natural experiment.

Because there are no changes in staffing levels in the ED on Wednesday morning other than the presence or absence of residents, the differences in treatment times for similar patients can be attributed entirely to the presence or absence of residents. Typically, there are two attending physicians on duty and four or five residents in the ED. When the residents are present, they do almost all of the “hands-on” treatment of patients, while the attending physicians play a managerial/supervisory role. When the residents are present, they are simultaneously treating patients and receiving instruction from the attending physicians. The attending physicians oversee the care and teach the residents. Therefore, our simulation model assumes that when residents are present they treat every patient who arrives. When the residents are absent, due to the seminar, the attending physician’s role shifts from supervisory to active care-providing. As a consequence, they now spend their time treating patients, rather than supervising and teaching residents. The changes in treatment times that we see when residents are not present are a result of this shift. This assumption was motivated by conversations with ED physicians at the UMMC.

To attribute treatment time changes on Wednesday morning to staffing levels, we must verify that Wednesday morning is similar to the rest of the week in terms of arrival rates and patient severity. To do this, we compare the patients who arrive on Wednesday morning (when residents are absent) to those who arrive at all other times of the week (when residents are present). Figure 6-1 shows the historical arrival rates over the course of the week. There is a wide range of arrival rates for Wednesday morning. In general, there are more arrivals than on weekend mornings and fewer than on Monday or Tuesday mornings. In addition, morning arrival rates are higher than overnight rates and lower than afternoon rates. So, Wednesday morning arrival rates are not atypical in any way. Furthermore, the patient population mirrors that of the rest of the week, in terms of severity and admission rate. We compared the two patient populations (those treated when residents were present and those treated when they were absent). We found that when residents were absent, 47%, 50%, and 3% of patients were of high, medium, and low severity, respectively, while those numbers were 45%, 51%, and 4% when residents were present. A chi-square test fails to reject (p = .81) the hypothesis that the underlying severity distribution is the same between the two patient populations. Similarly, the proportion of patients needing lab tests (72.4% vs. 75.8%) is not provably different (p = .22) between the two populations. The fact that the two patient populations are so similar gives us further confidence that the differences that we observe in treatment times between the two populations is caused by the presence or absence of residents, and not by other factors.

Therefore, we are fortunate to have a representative sample of patients not treated by residents on Wednesday morning, which enables us to quantify the effect of having residents work in the ED.

Based on historical arrival and severity data, we built a simulation model of the ED. Figure 6-2 shows a flow diagram of the ED simulation. We use this model to determine the effect of residents not just on treatment times for patients, but on the ED system as a whole. By building a simulation model, we can show how the presence of residents in the ED affects waiting times, throughput, and total time in the system. Moreover, the ED is a complex system with many interdependent parts. Because of this complexity, we felt that a simulation model would be more appropriate than other types of models (e.g., queueing models). Building a simulation model also allows us to easily experiment with the system to see how changing parameters of the system would affect performance. We implemented the simulation using SimPy, a discrete-event simulation language for Python.

The effect that residents have on treatment times is handled implicitly by the simulation. As discussed previously, we assume that every patient treated when residents are present is treated by a resident, while those treated when residents are absent are not. We do not model the specific movements of individual physicians through the ED or every doctor-patient interaction. Instead, we take a higher-level view of the ED and simply simulate patient flow.

For this study, we used historical data from the UMMC ED. The UMMC ED is divided into separate sections that treat adult medical patients, pediatric patients and psychiatric patients. There is a separate area outside the ED for patients with significant trauma. The main adult ED, the site for the prospective data collection, sees approximately 50,000 primarily adult medical and urgent care patients annually.

We build our model from the UMMC patient database data from October 1, 2009 to January 31, 2010 that contained data from the adult medical and psychiatric areas. The patient identities were masked. There were almost 17,000 patient visits during these four months and each record contained information about the patient's triage score, treatment process, and when and why they left the ED.

6.3.1 Patient Creation

Patients enter the simulation model according to a nonhomogeneous Poisson process, with the arrival rate based both on time of day and day of the week, drawn from the historical arrival data. After each patient is generated, he/she is seen by the triage nurse. At the triage station, the patient is assigned a severity score from 1 (highest) to 5 (lowest), and held for a random amount of time based on historical average triage times. A small number of patients are not given a severity score. These correspond to patients brought in via ambulance and with extremely high severity. In addition to the severity score, the simulation determines the amount of lab work the patient needs and whether or not the patient will eventually be admitted to an inpatient ward, based on the severity score.

We chose these three attributes (severity, labs, and admission) because they were the most important in determining the treatment time that a patient required and the most medically relevant. Higher severity patients take, on average, longer to treat. A high-severity patient will require more intensive care and will be held longer in the treatment bed. Similarly, a patient who is admitted to the hospital from the ED is likely to be held longer. Patients who are admitted have more severe and complex problems than those who are not. Finally, the number of labs that a patient needs directly affects the treatment time. Lab work takes time to process, which causes the patient to stay longer.

6.3.2 Patient Selection

Once a patient is discharged and a bed becomes free, the physician must select a patient from the waiting room. While we might expect the patients to be selected strictly according to severity, the historical data confirms that this is not the case. Based on the historical data, we found that the number of times that a patient was passed over lowered his future chances of being selected for treatment. This means that a severity 2 patient who has been passed over a few times might be less likely to be picked than a newly arrived severity 3 patient, even though the patient is in a higher severity class.

There is no deterministic rule for how patients are selected, so we constructed a discrete choice model, using logistic regression, to model how patients were selected. Patients were split into 4 severity categories: high (severity score of 1 or 2), medium (score of 3), low (score of 4 or 5) and N/A (no score given). Within each of the severity categories, we split the patients again 4 ways, based on how many times they had been passed over in the selection process: never, once, 2-3 times, and 4+ times, giving us 16 different patient categories. The probability that each patient would be chosen from a waiting room with one patient of each type is shown in Figure 6-3. We see that high severity patients are much more likely to be chosen than low severity patients, but also the more times a patient has been passed over the less likely he/she is to be selected.

This presented us with a discrete choice problem. Each time a bed becomes free, triage nurse must select one and only one patient from the waiting room. Each time a patient is selected, in the historical data, we note which patient type is selected, and how many of each type are still in the waiting room. Typically, this class of problems is solved using multinomial logistic regression. However, this approach requires on the order of 2^N terms in a choice set with N alternatives. In our case, this would require estimating more than 65,000 terms, which is computationally prohibitive. Instead, from the constructed dataset, we built a series of logistic regression models (see Hilbe, 2009) that measure the probability of each type of patient being selected given the distribution of patients in the waiting room. The probabilities from these regressions were used to choose which type of patient would be selected next in the simulation model. These sequential logistic regression models approximate what multinomial logistic regression does.

Because patients sometimes leave the waiting room before being treated, our simulation must take abandonment into account. From the historical data, we know the probability that a patient of a given severity will still be in the waiting room based on the number of hours he/she has been waiting. After a patient is selected from the waiting room to be treated, we determine if he/she is still in the waiting room. If the patient is absent, another patient is selected from the remaining patients in the waiting room. Once a patient has been selected and is still present, he/she is assigned to a treatment bed and held until treatment is over. The probability that a patient of each severity class is still in the waiting room is plotted in Figure 6-4. The curves are not smoothly decreasing because the sample size becomes very small as waiting times increase. Very few patients wait over six hours to be seen, and the data only record whether or not the patient was present when selected, not the exact time that they left the waiting room. We only know when patients who have left without being seen are called to be placed in a bed.

6.3.3 Treatment Time

Once in the treatment bed, the patient remains there for a length of time drawn from empirical distributions. We used empirical distributions because they were able to capture the long tails of treatment times better than kernel densities. When possible, we categorized each patient by a number of binary splits. The first split was based on whether or not the ED was congested (defined as more than 4 patients in the waiting room). Second, we split the patients based on whether or not they were eventually admitted to the inpatient ward, as admitted patients and discharged patients have different ED length of stays and different service needs. Third, we split the patients based on the amount of lab work they needed, their severity level, and whether or not they were seen by a resident. However, due to data sparseness issues, we were not able to make every split. For example, there were very few low severity patients with no lab tests who were admitted to the inpatient ward. The length of treatment time for each patient was drawn from the empirical distribution for that patient's category.

Once the treatment time had elapsed, the patient left the simulation (either via discharge or admittance to the inpatient ward), and the bed was held shortly while being prepared for a new patient. Once the bed has been cleaned, a new patient is called back, and the cycle repeats. Because one of the parameters that determine treatment time is whether or not the patient is treated by a resident, we can run experiments with our simulation by varying that parameter for groups of patients. In Section 5, we present these experiments.

By measuring treatment times based on the treatment and ED characteristics (labs, severity, congestion, admission to the hospital), we are able to control for possible confounding of the effect of residents, enabling us to isolate the effect that residents have on ED efficiency. For instance, whenever a simulated low-severity patient with no lab tests enters the ED during an uncongested time with residents present and is later discharged, we draw treatment times for that patient from an empirical distribution of all similar patients in the historical ED data who were treated when the ED was uncongested and when residents were present (all times except Wednesday mornings). If we were simulating the same patient, except without residents present, we would draw treatment times from an empirical distribution of all similar patients who were treated when the ED was uncongested and when residents were not present (Wednesday mornings).

6.4 Validation

After building the simulation model, we tested it to make sure that it was a valid replication of the system we were simulating. We did this by comparing the similarity of the outputted data from our model with the observed performance of the ED. While validating the model, we mirror the actual system, with residents present all week except for Wednesday mornings. We compared statistics from the simulation regarding patients per bed per day, the rate at which patients abandoned the waiting room before being seen, time spent until placed in a bed, and total time in system with those from the historical database. These are metrics often used to evaluate ED performance and efficiency. By demonstrating that the data generated for these metrics were statistically similar to the data from the historical database, we were able to confirm that we have a valid simulation model.

We chose the above-mentioned comparison metrics because they describe the overall performance of the ED. We simulated 20 years of data to compare to the historical values. From the simulated data, we calculated the mean and standard deviation for each of the performance metrics. Table 6-1 illustrates the similarities between the simulation model and the historical data. None of the metrics were provably different from the historical values.

We used a Kolmogorov-Smirnov (K-S) test to test the similarity of the total time in system distributions from the simulation and the historical data. The K-S statistic for two samples measures the difference between the empirical cumulative distribution functions (ECDFs) of the two samples. The ECDFs are step functions that approximate the underlying distributions from which the samples are drawn. We find the maximum vertical distance between the two ECDF curves, and compare it to the expected difference if the two samples were drawn from the same population. If they are farther apart than what would happen in 95% of cases, we can say with 95% confidence that the two samples were drawn from different distributions.

The K-S test statistic for the total time in system metric was .0075, meaning that the farthest distance between the two distributions was 0.75%. This translates to a *p*-value of .513, meaning that we cannot reject the null hypothesis that the simulation output and the historical data have the same length of stay distribution. Our time in system distribution matches the historical data almost perfectly, and the other performance metrics are similar to the historical data at the means. Noting that simulations by their very nature simplify a complex system and, therefore, cannot perfectly replicate that system's performance, we felt comfortable with the model validation results.

6.5 Experiments and Results

In our first experiment, to determine the effect that residents have on ED efficiency, we varied the proportion of patients seen by residents from 0 to 1, in increments of 0.1, and observed the changes in efficiency metrics, such as throughput and average waiting time. From one run to the next, the only change in the system is the percent of patients seen by a resident. In this experiment, residents see each patient with the same probability, regardless of patient severity. Because treatment by a resident is a parameter in the simulation, we randomly select whether a patient is treated by a resident when that patient enters the ED. We ran 20 years' worth of simulations for each level of resident presence and recorded the performance metrics from these simulations. These experiments test the hypothesis that the addition of residents to the ED slows down doctor performance and harms system efficiency.

We found strong linear trends in the relationship between the patient-based characteristics and the presence of residents. For example, we saw decreases of over 16% in total time (from 11.5 to 9.5 hours) for both high- and low-severity patients when residents were added. Additionally, we saw decreases in the time to get patients into a bed of 23% for high-severity patients and 20% for low-severity patients. Figure 6-5 shows the relationship between the total time in the ED for patients and the percent of patients treated by residents.

We also observed increases in system-wide efficiency. We measured total throughput in terms of patients treated per bed per day and found that having residents treat patients helped improve throughput. In particular, we found a 6% increase in total throughput (from 2.26 patients per bed per day to 2.38) when resident presence was increased from 0 to 100%. Figure 6-6 shows a plot of patient throughput versus resident presence.

The third performance metric we monitored was time to first bed. Again, we found that increasing the fraction of patients seen by a resident helped to improve system performance. This is especially important in an ED because patient welfare often depends on how quickly they can be seen and treated by a doctor. Figure 6-7 shows the effect of increasing the percent of patients seen by a resident on time to first bed. The addition of residents lowers average waiting times by 35% (from 92 minutes to 60).

In our second experiment, we independently varied both the percentage of high- and low-severity patients seen by residents. Because the residents' main purpose in the ED is to learn, and because the high-severity cases are the most instructive, we fixed the fraction of high severity patients seen by residents always at or above one half. We simulated 20 years with 121 different patient mixes, varying the fraction of high severity patients seen from .5 to 1 (in increments of 0.05), and of low severity patients between 0 and 1 (in increments of 0.1).

We found that the driving factor in increasing efficiency was the fraction of high-severity patients seen. This effect is illustrated in Figure 6-8, a contour plot of total time in system for all patients vs. the percent of each type of patients seen by residents. The contour lines are all nearly vertical, which shows that the driving factor is percent of high severity patients seen. The reasons are threefold: the majority (75%) of patients in the UMMC ED are high severity, residents have a much bigger effect on the service time for high-severity patients than for low-severity ones (5.3% vs. 1.9%), and high-severity patients take about twice as long to treat (8 hours vs. 4 hours), so a similar percent reduction in their service time will more heavily influence the average total time in system. We hypothesize that residents increase treatment speed for high-severity patients more than low-severity patients because more complex care is required and there are more chances for work to be done in parallel with attending physicians on high-severity patients. On the other hand, with lower severity patients, the complexity of treatment is lower, and there are fewer chances for work to be done in parallel, so the treatment times are not reduced as much.

The effect of residents on throughput is similar. The percent of high severity patients seen by a resident has a strong effect on throughput, while the percent of low severity patients seen has no detectible effect on throughput. Figure 6-9 shows a contour plot of throughput vs. resident presence. Raising the percent of high severity patients treated by a resident from 50 to 100 increases throughput by 2.7%. The contour lines are essentially vertical, meaning that changing the percent of low severity patients treated by a resident has no effect on throughput. Again, this may be because high severity patients take longer to treat, are a higher fraction of the ED patient population, and because residents have a larger effect on their service times.

In our third experiment, we tested the effect of resident presence on efficiency when treating a variety of patient populations, to see the effects of residents in medical centers that are similar to UMMC but that have different patient characteristics. We generated two additional patient populations, one with a predominantly high-severity patient population (90% high severity), and one with a predominantly low-severity population (50% high severity). All other treatment and patient attributes were held the same. We then looked at the effect of having residents present on efficiency.

We saw that, regardless of patient mix, residents still have a positive effect on system efficiency. Figures 6-10 and 6-11 show the effect residents have on total time in the ED and waiting time, respectively. In both patient populations, residents have positive effects on efficiency. The total time in system effect is about the same for both populations, reducing average time from 12.3 hours to 10.5 hours in the more severe population, and from 7.2 hours to 6.1 hours in the less severe population. The lines for the two populations are essentially parallel, meaning that the effect is the same in both patient populations. In both populations, total time in system is reduced by about 15%.

Residents also had an effect on waiting time in both populations. Figure 6-9 shows a graph of time to first bed vs. resident presence for the more and less severe populations. In this case, residents had a much more significant effect on first bed time in the more severe population than in the less severe population. This may be because more severe patients take longer to treat, and, therefore, they increase the stress on the system. This leads to longer queues and more variation in waiting time. This means that a similar reduction in processing time has a greater impact on waiting times in the high severity population than in the low severity population.

6.6 A Related Queueing Model

In addition to the simulation, we also used queueing theory to model the flow of patients through the ED. Specifically, we chose to use an *M/G/k* queue to represent the system, with each bed being treated as a server. This requires a few simplifying assumptions. First, we assume that patients arrive according to a Poisson process with a fixed arrival rate. Second, we assume that all patients who enter the queue will wait until they are served (no abandonment). Third, we assume that patients are treated in the order in which they arrive (first come, first served). We analyzed the queue with two different service time distributions. The first was the empirical distribution for patients treated on Wednesday mornings, when residents were absent. The second distribution was the empirical distribution for all patients treated during the mornings of the other weekdays (when residents were present), excluding weekends. These simplifying assumptions make the model tractable and allow us to estimate the average waiting times and queue lengths for the system with and without residents.

No closed-form solution to the *M/G/k* queue exists, so we use the approximations derived by Nozaki and Ross (1978), which take into account the first two moments of the service time distributions. An average of 2.94 patients arrived at the ED per hour and there are 27 beds in the ED. The mean treatment time of patients when residents were absent was 8.22 hours and the mean squared treatment time was 121.65 hours. When residents were present, the mean treatment time was 7.9 hours and the mean squared treatment time was 103.6 hours. The queueing model reports that the average waiting time of patients when residents are present is 55 minutes, compared to 135 minutes when residents are absent. So, when residents are present, we observe a 59% reduction in waiting time. The residents have a similar effect of the average number of patients in the waiting room (average queue length). If residents were always present, the average queue length would be 2.7, compared to 6.6 if residents were always absent, again a reduction of about 59%.

The waiting times predicted by the queueing model are lower than the historical averages, as a result of the simplifying assumptions. The queueing model has less variability than the real system, so it will have fewer occurrences of high congestion, which leads to lower average waiting times. While the waiting times are smaller, the effect that residents have on waiting times is a 59% reduction in the queueing model compared to 35% in the simulation. Although the queueing model cannot address all of the questions that the simulation can, with respect to time to first bed, the two models at least point in the same direction. This serves to enhance our confidence in the simulation model.

6.7 Conclusions

A common hypothesis in the medical community is that residents slow down treatment in EDs and have a negative impact on system efficiency, compared to just attending physicians. In this chapter, we have shown that, to the contrary, residents have a positive effect on throughput and treatment times. In particular, we found that, when treating high severity patients, residents help to decrease waiting times, decrease treatment times, and increase throughput. While efficiency might not be a main concern in deciding which patients are seen by residents, we would recommend that they see as many high severity patients as is feasible. This fits with the mission of the ED residency program. Furthermore, since residents cannot work as many hours per week as in the past, it is important for them to use their time wisely and productively. The main contribution of this chapter is to provide evidence refuting the hypothesis that residents slow down progress in the ED and that they have a negative effect on efficiency.

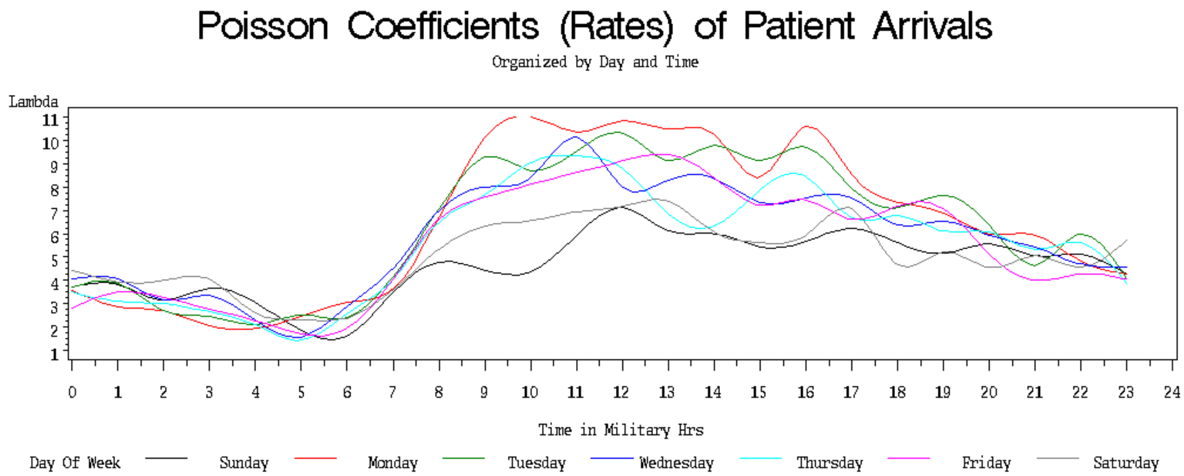

Figure 6-1: Arrival Rates by Day of Week

Figure 6-2: Flow diagram of simulation

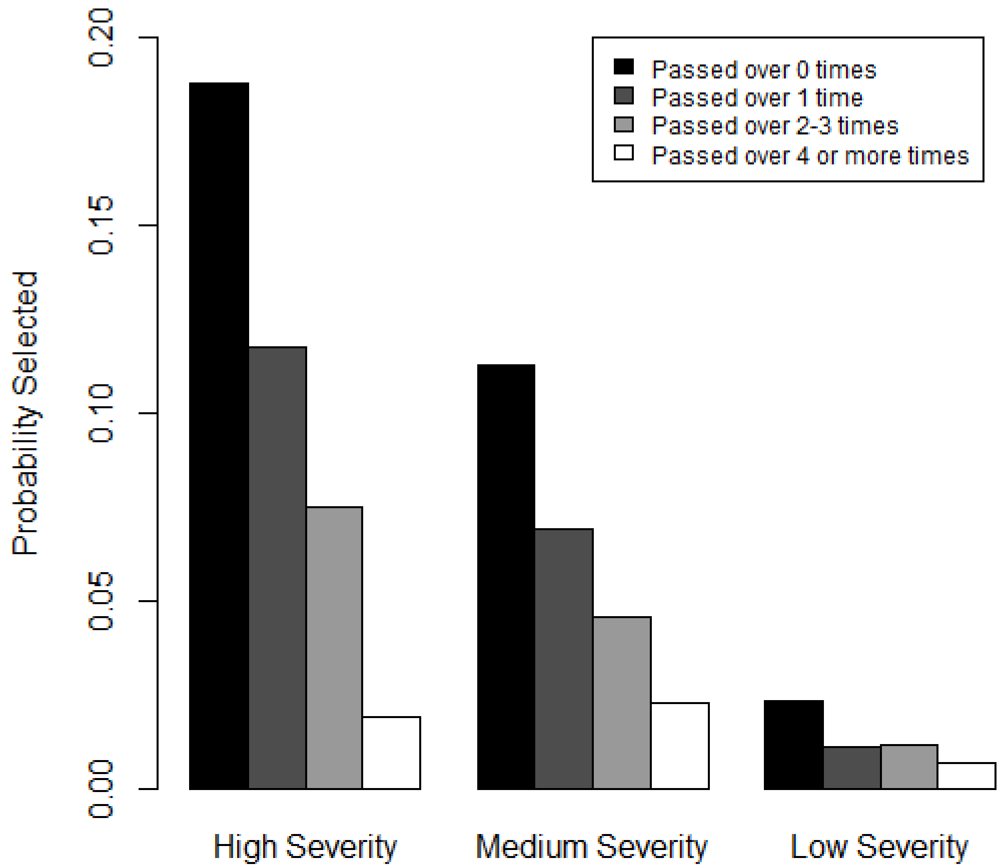

Figure 6-3: Probability each patient type is chosen from the waiting room

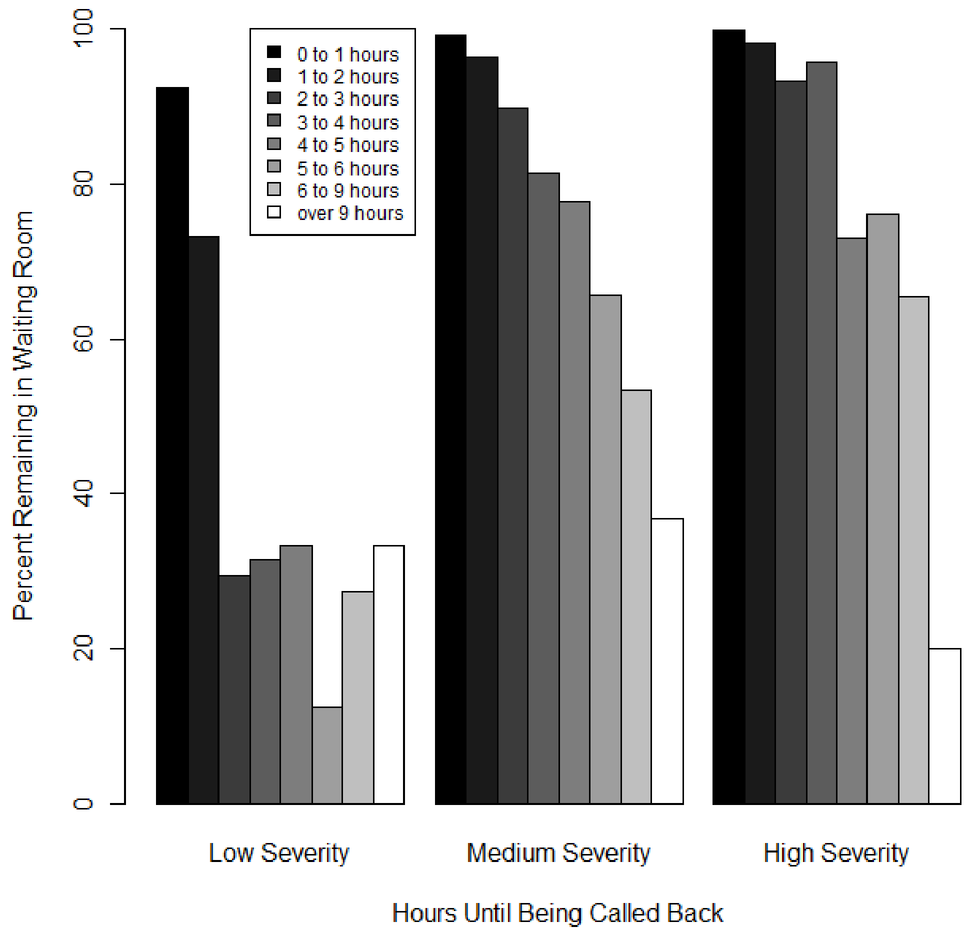

Figure 6-4: Percent remaining vs. time until called back

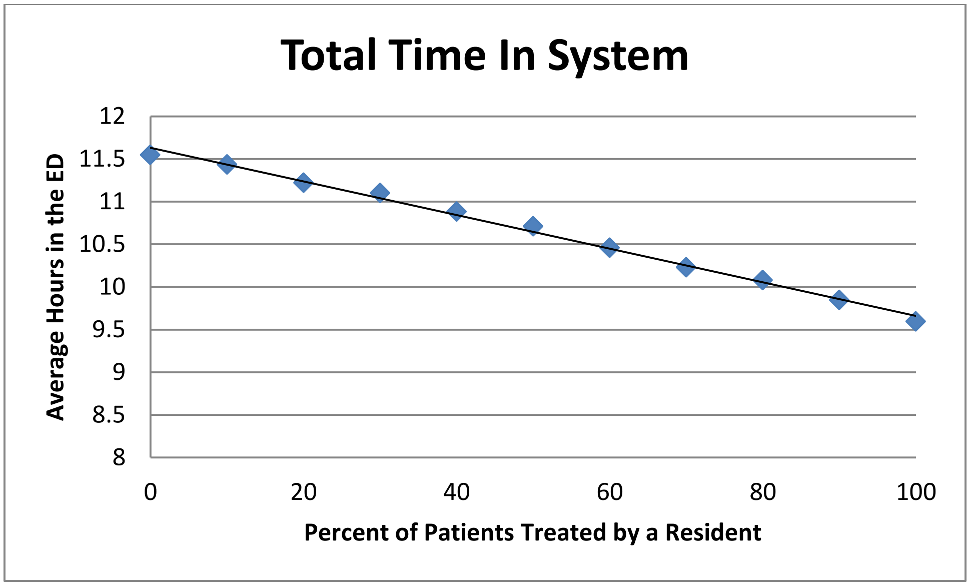

Figure 6-5: Total time in system (in hours) vs. resident presence

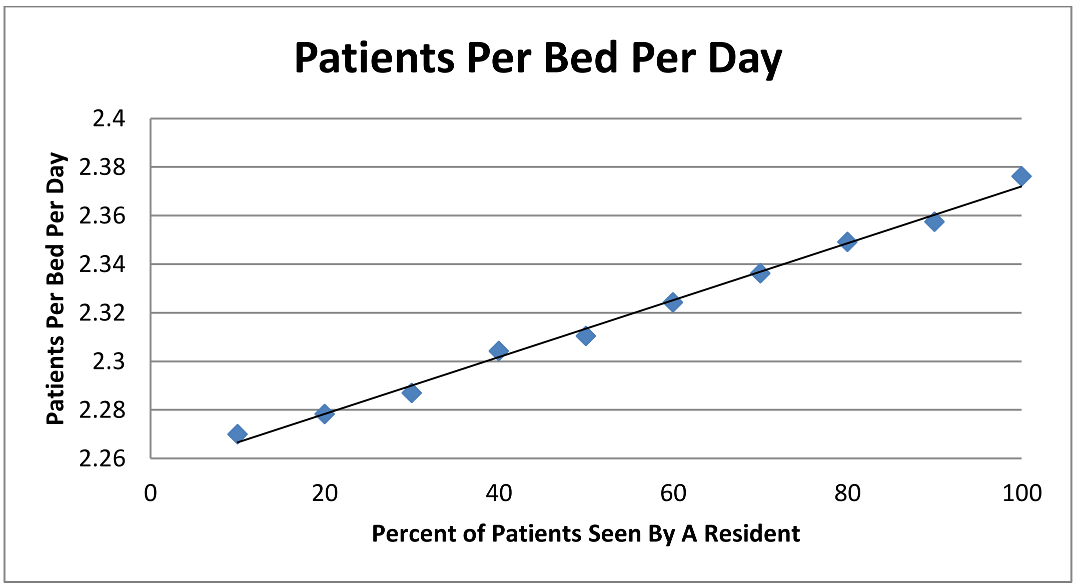

Figure 6-6: Throughput (in patients per bed per day) vs. resident presence

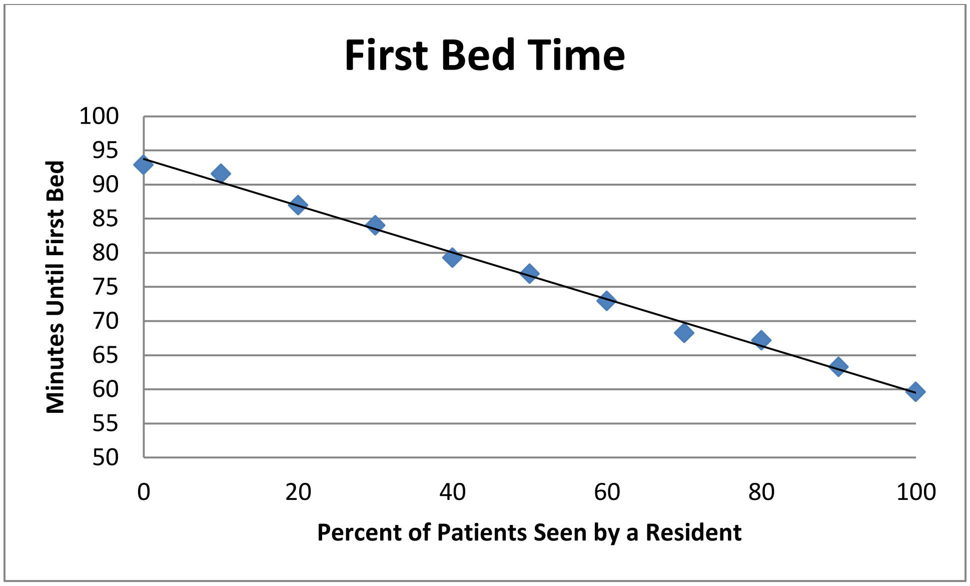

Figure 6-7: Average time to first bed (in minutes) vs. resident presence

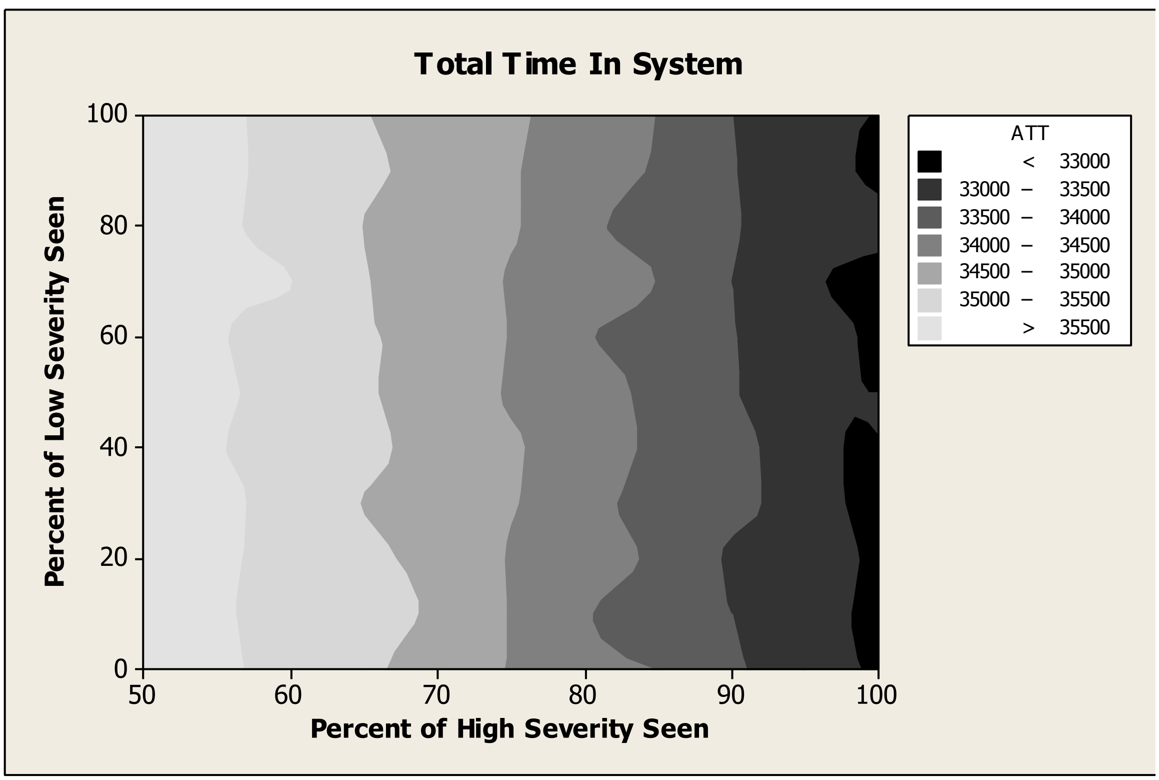

Figure 6-8: Total time in system (in seconds) vs. resident presence

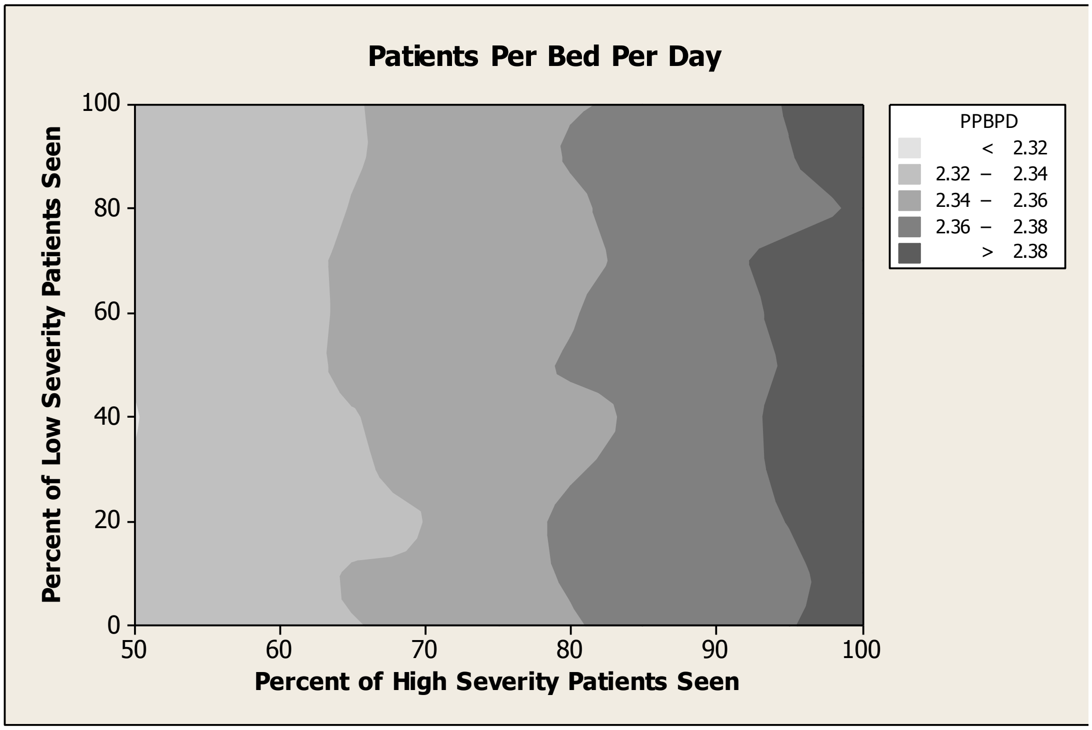

Figure 6-9: Contour plot of throughput (in patients per bed per day) vs. resident presence

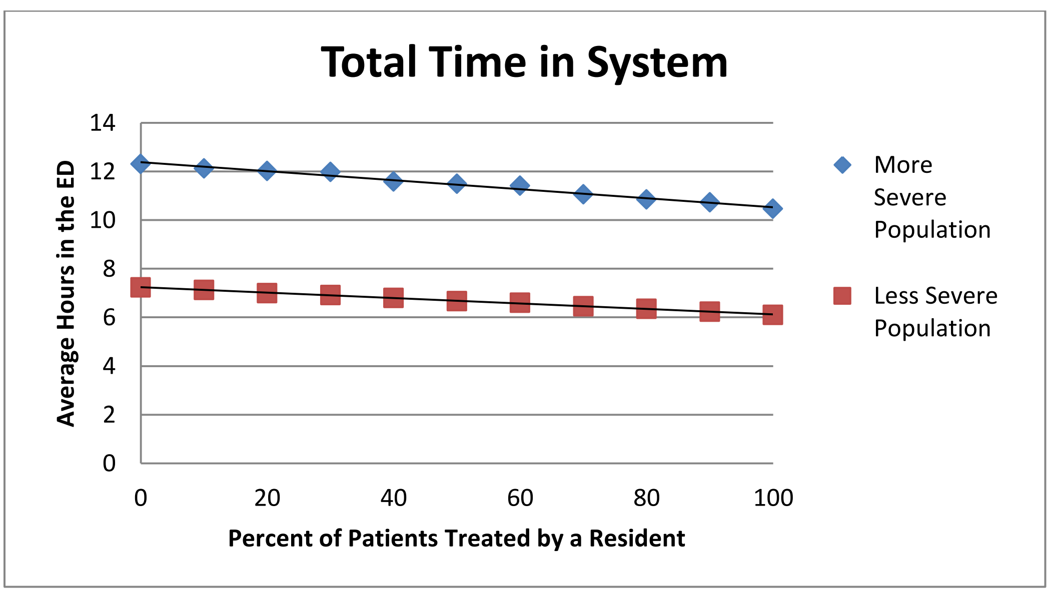

Figure 6-10: Total time in system (in hours) vs. resident presence for different patient severity mixes

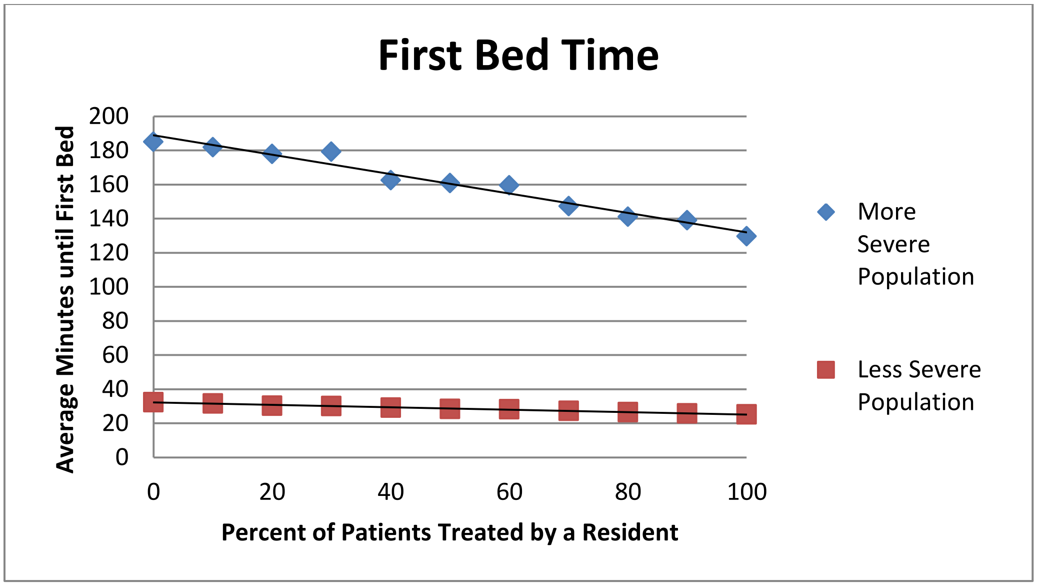

Figure 6-11: Time to first bed (in minutes) vs. resident presence for different patient severity mixes

Table 6-1: Comparison of simulated means with historical means for key ED efficiency metrics

Metric	Historical Mean	Simulation Mean	*p*-value
Patients per bed per day	2.38	2.39	.4413
Abandonment rate (in percent)	8.02	7.76	.4611
Time to first bed placement (in minutes)	80.32	81.12	.8650
Total time in system (in minutes)	550.15	549.28	.9134

CHAPTER 7: LONG DISTANCE TRIAGE

7.1 Introduction

Inter-hospital transfer (IHT) patients tend to be among the highest acuity patients in the UMMC internal medicine department. Anecdotally, doctors noticed that these patients were much more likely to suffer an adverse event (either transfer to the ICU within 48 hours of admission or death) than similar patients admitted through the emergency department (ED). While there are decision support tools to assist assigning the correct triage level to ED patients in person, there are no such tools for over-the-phone triage prior to patient transfer. It would be useful to have a clinical prediction tool to identify high-risk patients prior to arrival. We show that high-risk patients can be identified reasonably well using just four features (shown in Table 7-1). A patient with any of the four risk criteria is significantly more likely to suffer an adverse event. By dichotomizing each variable, we provide a simple tool (called HALT) that can be used by physicians to assess the severity of IHT patients. We extend this work by building classifiers that consider the continuous variables.

7.2 Preliminary Analysis

First, we looked at every subset of features from the HALT tool to measure its predictive ability. For example, the AT tool uses Anemia and Tachycardia but not Hypertension or Leukocytosis. Second, we constructed three statistical classifiers using the continuous data. We used logistic regression, naïve Bayes classification, and a combination of the two (called Max). Naïve Bayes classification is a commonly used classification technique. It calculates the likelihood that a given observation came from each category based on the distributions of the independent variables of the observed data. The combination tool (Max) classified a patient as high risk if either the logistic regression model or the naïve Bayes model found the patient to be high risk. The continuous tools have fewer observations because some data were missing. The classification results from each model are shown in Table 7-2. This table gives the number of patients of each type correctly identified, the number of Type I errors (false positive) and Type II errors (false negative), the sensitivity (percent of high-risk patients correctly identified), and specificity (percent of low-risk patients correctly identified) for each tool. Using the results in Table 7-2, we see that the naïve Bayes and logistic regression classifiers strictly dominate the HALT tool in every dimension. They have a higher percentage of patients correctly identified, and fewer Type I and Type II errors. We observe that the four individual features (H,A,L,T) do not perform well individually, but do much better when combined. This means that each feature uses different information about a patient, and complements the others well.

7.3 Error Costs

For UMMC, committing a Type II error is worse than a Type I error. It is worse to incorrectly classify a very sick patient as low risk than to incorrectly classify a relatively healthy patient as high risk. To illustrate this, we assign different costs to each error, ranging from equal costs to a Type II error being 100 times worse than a Type I error (shown in Table 7-3 in ascending cost order at a

5:1 ratio). The higher the relative cost of Type II error, the more the system rewards methods that aggressively classify patients as high risk. We see that, regardless of the cost structure, the naïve Bayes and logistic regression classifiers have lower costs than the HALT tool.

7.4 Varying the Threshold

We construct the receiver operating curves (ROC) for the naïve Bayes and logistic regression models. ROC plots show the tradeoff between detecting high risk cases and committing a Type I error. Changing the threshold for high risk is equivalent to moving along the *x*-axis. A lower threshold means that more patients are classified as high risk, which implies a move to the right along the *x*-axis. The vertical distance above the 45-degree line gives the improvement of the model over random guessing. In Figure 7-1, both models correctly identify about 50% of patients who eventually have an adverse event, while only classifying 10% patients without an adverse event as high risk. However, after that point, the curve flattens out. In order to correctly identify 80% of patients with adverse events as high risk, we must incorrectly label 50% of patients without a negative outcome as high risk as well.

7.5 Implementation

The logistic regression results dominate the HALT tool in every dimension. We constructed a combination tool, similar to Max, that classifies a patient as high risk if either HALT or the logistic regression model classifies that patient as high risk. This tool has the best sensitivity of any tool we tested (correctly detecting 67% of all high-risk patients). However, it gives more false positives (sensitivity of 78%). The two-by-two tables for each tool are given in Table 7-4. The rows show if the model classifies the patient as high risk or not, while the columns show whether or not the patient had an adverse event. A zero denotes no adverse event occurred the model classified the patient as low risk. A one denotes an adverse event or the model classified the patient as high risk. For example, with the HALT tool, 128 patients were classified as high risk but did not suffer an adverse event. The number of observations differs between the tools because there were missing data. When data are missing, the tools using continuous data (logit and the combination tool) cannot estimate the risk of the patient, while the HALT tool just assumes that any missing value is in the low-risk range.

To generate the logistic regression tool, we fit the following model:

$$\text{Adverse Event} = \beta_1\ MAP + \beta_2\ HGB + \beta_3\ WBC + \beta_4\ Pulse + \varepsilon$$

where *MAP* is the patient's mean arterial pressure, *HGB* is the patient's hemoglobin count, *WBC* is the patient's white blood count, and *Pulse* is the patient's pulse.

The logistic regression model is easy to construct with a calculator or a spreadsheet program. The equation for the probability that a patient is going to experience an adverse event is given by:

$$\frac{1}{1+e^{-(-1.5623-.0222\,MAP-.1767\,HGB+.1119\ WBC+.0214\ Pulse)}}\,.$$

After fitting the model, we generated the probability of an adverse event for each patient. We wanted the logistic regression model to classify patients as high risk at the same rate as the HALT tool. To do this, we sorted the patients according to their predicted probabilities of an adverse event. The HALT tool classifies 14.5% of patients as high risk. The highest 14.5% of patients had a predicted probability of greater than 0.13 for an adverse event. This means that a patient with a 13% chance or higher of an adverse event, given by the logistic regression tool, would be classified as high risk.

Consider a patient with MAP = 82, Pulse = 94, WBC =18, and HGB = 10.4. This patient meets none of the HALT criteria, thereby producing a HALT score of 0. Using our logistic regression equation, the probability an adverse event for this patient is

$$\frac{1}{1+e^{-(-1.5623+.1119\ (18)-.0222\ (82)+.0214\ (94)-.1767\ (10.4))}} = .232\ .$$

This probability exceeds the threshold of 0.13. We would classify this patient as high risk using either the logistic regression model or the combination model.

7.6 Extending With Age

We also included that patient's age as a predictor of mortality or admission to the ICU. As a result, we were able to achieve marginal improvements in predictive accuracy. Including age as a continuous variable had no effect on the predictive power of the tool, but creating a dichotomous variable for elderly patients (defined as at least 80 years old), we were able to achieve small improvements. We included the dichotomous age variable in the logistic regression, and estimated the probabilities of an adverse event for each patient based on the extended regression model. We then classified each patient as either high or low risk, using the same procedure described in Section 5. Table 7-5 shows the results of this classification.

When compared with the original logistic regression model, the extended model performs slightly better. The sensitivity of the model increases from 52.2% to 53.7%, while the specificity increases from 87.7% to 88.0%. This translates to a decrease in "cost" of between 3.3% (1:1 ratio) and 6.7% (100:1 ratio). Including patient age in the model yields a small, but positive increase on the predictive power of the HALT tool.

7.7 Conclusions

In this chapter, we show that basic medical information can be used to predict which IHT patients will require higher levels of care. This information can be used when making admission decisions, as well as when planning staffing levels. Having accurate information about the severity of incoming patients before they arrive at the hospital will help decision makers better match supply with demand.

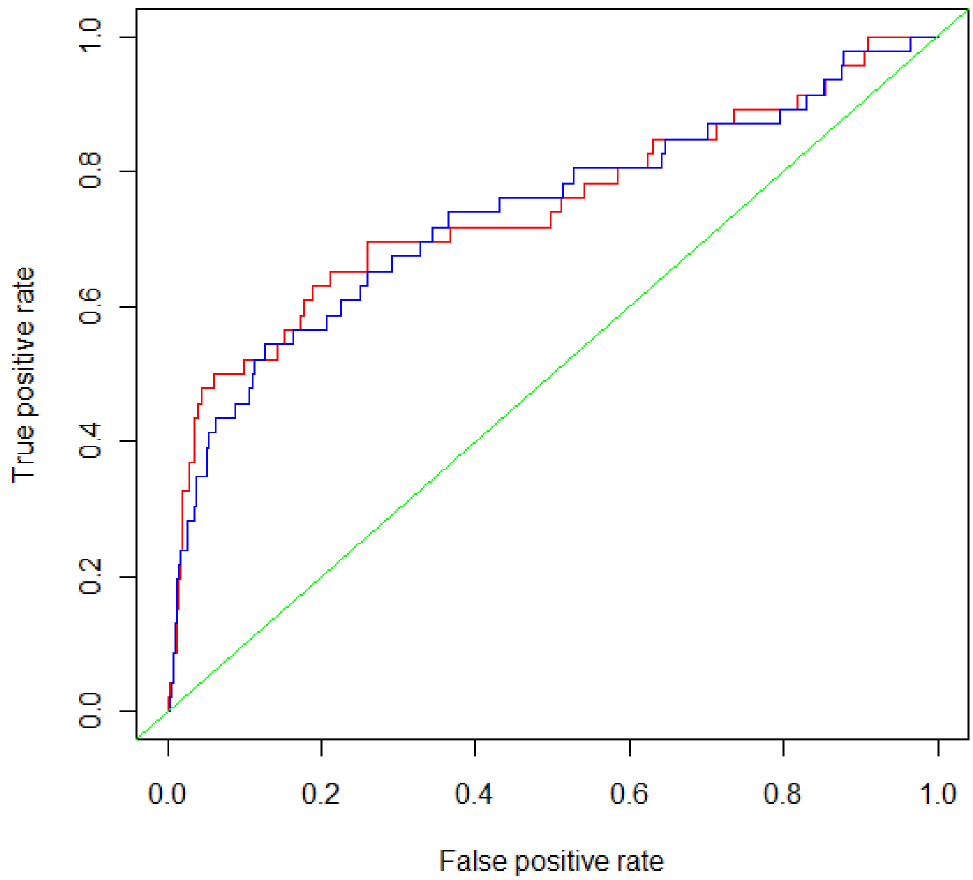

Figure 7-1: ROC plots for Naïve Bayes (Red) and Logistic Regression (Blue)

Table 7-1: Independent variables

Feature	Description
Hypertension	Measured by mean arterial pressure. Considered high if > 65 mmHg.
Anemia	Measured by hemoglobin. Considered low if < 7 g/dL.
Leukocytosis	Measured by white blood cell count. Considered high if > 20,000 cells/mcL.
Tachycardia	Measured by pulse. Considered high if > 100 beats per minute.

Table 7-2: Classification results

Tool	True +	True -	False +	False -	Sensitivity	Specificity
Logistic Regression	24	465	65	22	0.5217	0.8774
Naïve Bayes	24	468	62	22	0.5217	0.8830
Max	27	450	80	19	0.5870	0.8491
H	7	1051	12	88	0.0737	0.9887
A	5	1051	12	90	0.0526	0.9887
HA	12	1039	24	83	0.1263	0.9774
L	19	1034	29	76	0.2000	0.9727
HL	24	1022	41	71	0.2526	0.9614
AL	24	1022	41	71	0.2526	0.9614
HAL	29	1010	53	66	0.3053	0.9501
T	17	985	78	78	0.1789	0.9266
AT	19	974	89	76	0.2000	0.9163
HT	23	973	90	72	0.2421	0.9153
HAT	25	962	101	70	0.2632	0.9050
LT	33	958	105	62	0.3474	0.9012
ALT	35	947	116	60	0.3684	0.8909
HLT	37	946	117	58	0.3895	0.8899
HALT	39	935	128	56	0.4105	0.8796

Table 7-3: Error costs for each tool

Tool	Cost 1:1	Cost 3:1	Cost 5:1	Cost 10:1	Cost 20:1	Cost 50:1	Cost 100:1
Naïve Bayes	0.1458	0.2222	0.2986	0.4896	0.8715	2.0174	3.9271
Max	0.1719	0.2378	0.3038	0.4688	0.7986	1.7882	3.4375
Logistic Regression	0.1510	0.2274	0.3038	0.4948	0.8767	2.0226	3.9323
HAL	0.1028	0.2168	0.3307	0.6157	1.1857	2.8955	5.7453
HL	0.0967	0.2193	0.3420	0.6485	1.2617	3.1010	6.1667
AL	0.0967	0.2193	0.3420	0.6485	1.2617	3.1010	6.1667
HLT	0.1511	0.2513	0.3515	0.6019	1.1028	2.6054	5.1097
HALT	0.1589	0.2556	0.3523	0.5941	1.0777	2.5285	4.9465
L	0.0907	0.2219	0.3532	0.6813	1.3377	3.3066	6.5881
LT	0.1442	0.2513	0.3584	0.6261	1.1615	2.7677	5.4447
ALT	0.1520	0.2556	0.3592	0.6183	1.1364	2.6908	5.2815
HA	0.0924	0.2358	0.3791	0.7375	1.4542	3.6045	7.1883
HT	0.1399	0.2642	0.3886	0.6995	1.3212	3.1865	6.2953
HAT	0.1477	0.2686	0.3895	0.6917	1.2962	3.1097	6.1321
H	0.0864	0.2383	0.3903	0.7703	1.5302	3.8100	7.6097
A	0.0881	0.2435	0.3990	0.7876	1.5648	3.8964	7.7824
T	0.1347	0.2694	0.4041	0.7409	1.4145	3.4352	6.8031
AT	0.1425	0.2737	0.4050	0.7332	1.3895	3.3584	6.6399

Table 7-4: Two-by-two tables for three prediction tools (HALT, logistic regression, and the combination tool).

	Actual 0	Actual 1			Actual 0	Actual 1			Actual 0	Actual 1
HALT 0	935	56		Logit 0	465	22		(Logit + HALT) 0	414	15
HALT 1	128	39		Logit 1	65	24		(Logit + HALT) 1	116	31

Table 7-5: HALT Confusion Matrix With Age

	Actual 0	Actual 1
Extended Logit 0	434	19
Extended Logit 1	59	22

CHAPTER 8: DETECTING PROSTATE CANCER USING MRI DATA

8.1 Introduction

Prostate cancer is widely prevalent and hard to diagnose. The National Cancer Institute estimates that 16% of men born today will be diagnosed with prostate cancer in their lifetime (Howlader et al. 2012). Currently, the two main methods for diagnosing prostate cancer are a prostate specific antigen (PSA) test and a biopsy. Both of these methods have serious drawbacks. Although it has some predictive power, a PSA test can be unreliable with a high error rate (Hoffman et al. 2002). While a biopsy is more accurate, it is expensive and highly invasive with negative side effects (Cooper et al. 2004). Since a biopsy is conducted randomly within the prostate gland, it can result in a significant number of misses in cancer diagnosis, as well. We propose a new method for classifying patient risk using Magnetic Resonance Imaging (MRI) data. An MRI is being used more frequently to evaluate the prostate, because of its effectiveness in assessing both the anatomy and the physiology of the prostate tissue.

The widespread use of biopsies is an expensive and possibly inefficient use of resources. Biopsies are roughly three times as expensive as MRIs, costing an average of $2,100, compared to $700 for an MRI. Better pre-biopsy information about which patients have the highest risk for cancer will allow hospitals to direct diagnostic resources to those patients for whom they will do the most good. If giving each patient an MRI first can reduce the number of biopsies by a third, using MRIs as a screening method would reduce overall costs while simultaneously reducing the number of men who suffer side-effects from biopsies. Having better information about patient risk will allow hospitals to more efficiently allocate diagnostic resources, and will allow patients to make better informed decisions about whether or not to undergo a biopsy.

Our data was collected from multi-parametric images generated from Dynamic Contrast Enhanced (DCE)-MRI, which provides vascular permeability, Diffusion Weighted (DW)-MRI, which provides microstructural cell density, and Magnetic Resonance Spectroscopic Imaging (MRSI), which provides metabolic signatures of malignancy. Data from patients who had radical prostatectomy were analyzed. After radical prostatectomy, the prostate specimen was fixed in formalin. Axial sections (3mm) from the specimen were made using a prostate slicer. Digital images of both the slice specimens and the pathologic slides were obtained. Each prostate was subdivided into octants. This resulted in 223 octants (one was missing) that were examined. Histology information for these octants was given by an experienced radiologist using Gleason scores. Sample images are shown in Figure 8-1.

A Gleason score (Gleason 1977) is used as a measure of the severity of prostate cancer. In our data set, scores range from 0 to 8, with 0 indicating no cancer cell identified, 1 to 3 indicating indolent (slow developing) disease, and 4 to 8 indicating a tumor. In Figure 8-2, we show the distribution of Gleason scores in our data set. There are 223 Gleason scores corresponding to the severity of the cancer in each prostate octant in the data set. Since a portion of the population with elevated PSA but indolent disease receives unnecessary over-treatment, and a portion of the population may benefit from early and accurate detection of extremely aggressive prostate cancer, it would be beneficial to distinguish between aggressive cancer and indolent disease. Therefore, a Gleason score of 5 or higher designates aggressive cancer, and a score below 5 indicates indolent disease or no cancer.

8.2 Literature Review

While prostate cancer diagnostic methods are improving, no true “gold standard” test exists yet. The current standard diagnostic test, measuring prostate-specific antigen (PSA) levels in the blood has a high false positive rate, and can be quite inaccurate. Welch and Albertsen (2009) find that the introduction of the PSA test as a standard diagnostic tool has led to hundreds of thousands of false diagnoses and excess treatments. Because of the high false-positive rate of PSA tests, medical guidelines are conflicted as to whether or not PSA tests should be used regularly to screen for cancer (Cooper et al. 2004).

Another commonly used method for diagnosing prostate cancer is the digital prostate exam. Digital exams may add value to the standard PSA test, but still suffer from high false positive and false negative rates (Akdas et al. 2008). Furthermore, the results of digital exams can vary from physician to physician (Smith and Catalona 1995).

Currently, biopsies are the most accurate method for diagnosing prostate cancer. However they also are the most expensive, invasive, and have the highest likelihood of causing moderate or severe side effects (Catalona et al. 1994). Because of the cost, pain, and side effects of biopsies, other methods are sought to diagnose prostate cancer, or to identify high risk patients who should undergo biopsy to confirm the diagnosis. Recently, MRIs have been used as a method for identifying patients with a high risk of prostate cancer. Preliminary studies have found that MRIs can be used to identify high risk patients (Amsellem-Ouazana et al. 2005, and Padhani et al. 2000).

8.3 Analysis

First, we use logistic regression to predict whether or not the prostate slice has cancer. We use four features: apparent diffusion coefficient, which measures the magnitude of diffusion (the magnitude of diffusion of the prostate tumors is lower than the normal gland), volume transfer constant (K^{trans}), which reflects blood flow and vessel permeability, conventional average of T2 values, and spectroscopy scores. These features are taken from the multi-parametric MRI images and are used as input variables to predict the probability that each slice has cancer. Because our data set is relatively small, we use leave-one-out cross-validation to separate the data set into a training set and a test set. This method yields good results, with 64.6% accuracy, and an area under the ROC curve (AUC) of 0.66. AUC is one common measure of the predictive power of a model. It is equivalent to the probability that a randomly chosen positive observation will be ranked higher than a randomly chosen negative observation (Fawcett 1977). The confusion matrix and ROC curve for the logistic regression method are given in Table 8-1 and Figure 8-3, respectively.

Second, we use K-Nearest-Neighbor (KNN) to classify the data. A distance matrix is generated using Euclidean distances on the four input features.

For each observation in our data set, the outcomes for the five most similar observations are then recorded. If three or more of the five nearest neighbors have a Gleason score of five or higher, the observation is classified as high risk. If three or more have a Gleason score below five, the observation is recorded as low risk. This method has a predictive accuracy of 74%. The confusion matrix for KNN is given in Table 8-2. Table 8-3 gives the breakdown by number of neighbors with cancer. Furthermore, among the 62 observations classified as being the highest risk (four or five neighbors being cancerous), 55 observations are cancerous. In addition to performing well overall, KNN does a good job of identifying very high risk slices.

Third, we ran the initial logistic regression model discussed earlier, again using leave-one-out cross-validation, but augmenting the data with the number of cancerous neighbors from KNN. This augmented logistic regression model outperforms both the original logistic regression model and KNN, in terms of predictive accuracy and in terms of AUC. Predictive accuracy is 77%, and the area under the ROC curve is 0.85. The ROC curve for each model is given in Figure 8-4, and the confusion matrix for this method is given in Table 8-4. This model performs particularly well on the highest risk patients. Looking at the ROC curve, we see that it requires a false positive rate of 6% to identify 46% of true positives. This is important because this method could be used in conjunction with a PSA test to determine which patients might require a biopsy to definitively diagnose and locate the cancer.

Finally, we examined how our method performed on classifying just the highest severity cases of cancer. Because of the low mortality rates of prostate cancer and the side effects of treatment, many people choose to treat only the most aggressive cancers. For these reasons, we retrained the previous models with the objective of identifying observations with a Gleason score of 7 or 8. The augmented logistic regression model is very good at identifying very high severity cancer, especially when compared to the basic logistic regression model. The augmented logistic regression model has a predictive accuracy of 82%, and an AUC of 0.86, compared to 58% accuracy, and an AUC of 0.65 for the basic logistic regression model. The ROC curves are given in Figure 8-5, and the confusion matrices are given in Tables 8-5 and 8-6. We see that the augmented method has a very low false negative rate, misclassifying only five patients who have cancer with a Gleason score of 7 or 8.

8.4 Conclusions

The results from the three models show that MRIs could be used as predictive tools to assess patient risk. Using logistic regression and nearest-neighbor classification, we can accurately assess the risk that a patient has prostate cancer. This information can then be used to determine whether or not a patient should undergo further diagnostic tests, such as a biopsy. Because an MRI is not invasive, has few side effects, and is relatively inexpensive, it can be a useful tool in preventing an unnecessary biopsy from being performed. An MRI can lower the overall testing cost and reduce the number of side effects caused by a biopsy, while helping to identify and diagnose a patient who has cancer. In addition, our method can be used to direct where in the prostate a biopsy should be done. By targeting only suspicious areas for biopsy, our method could improve biopsy accuracy and protect healthy areas of the prostate from the damage associated with a biopsy. Furthermore, Hoffman et al. (2002) found that PSA scores have an area under the ROC curve of 0.67, significantly worse than that of the proposed augmented logistic regression method. This means that our method has the potential to improve diagnostic ability, while minimizing side effects.

This work shows, as a proof of concept, that MRIs can be used to detect prostate cancer with reasonably good accuracy. This work employs simple data mining techniques on a small dataset, but still yields good accuracy. Essentially, we show that computer-generated features pulled from the MRI images contain significant information that can be used to detect prostate cancer.

In future work, we hope to incorporate MRI images from healthy patients into the data set. This would give us a sample that is more representative of the overall population. We also hope to incorporate the result from a PSA test into our model, in order to further increase the power of our test. A model that uses both MRI and PSA data should be able to accurately diagnose prostate cancer, reduce the number of unnecessary biopsies done, and increase the number of high severity cancers that are treated.

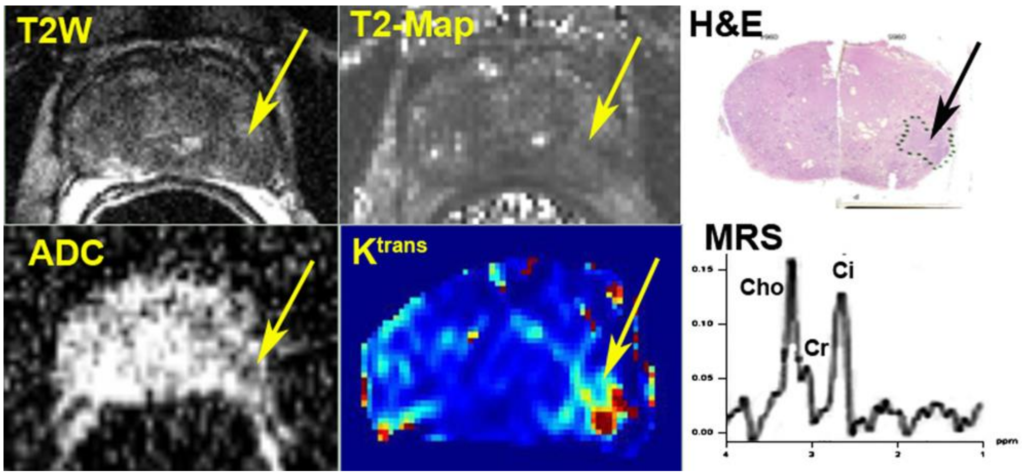

Figure 8-1: Multi-parametric imaging of prostate cancer:
Examples of five types of MRIs are shown, and corresponding spectra from the tumor showing low citrate and high choline. Histology and images showing cancer as marked by the arrow.

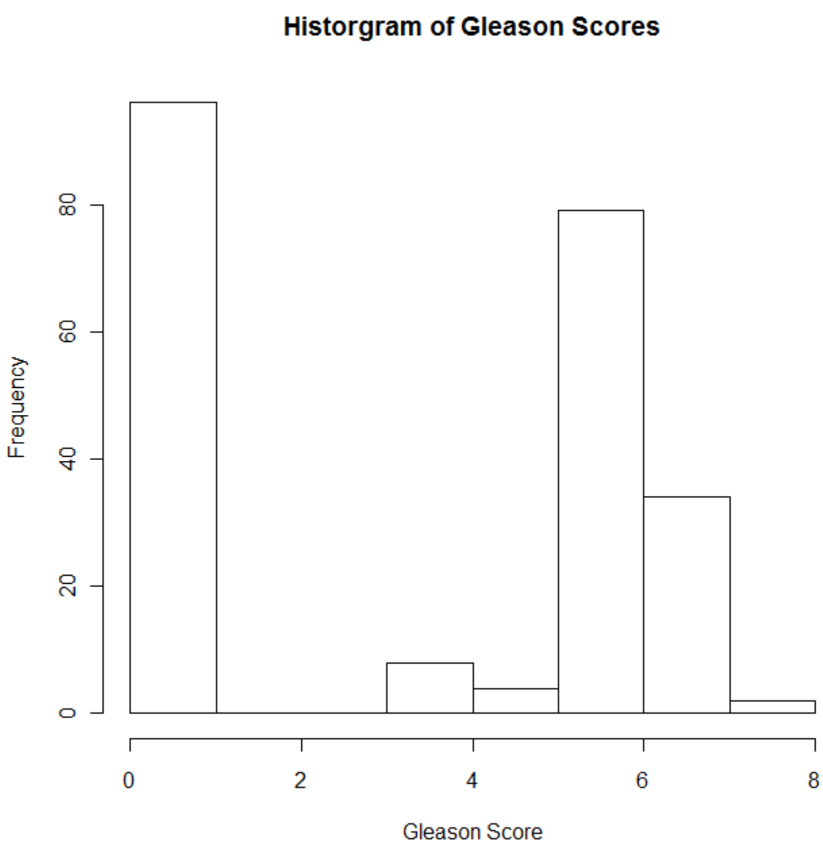

Figure 8-2: Histogram of Gleason Scores

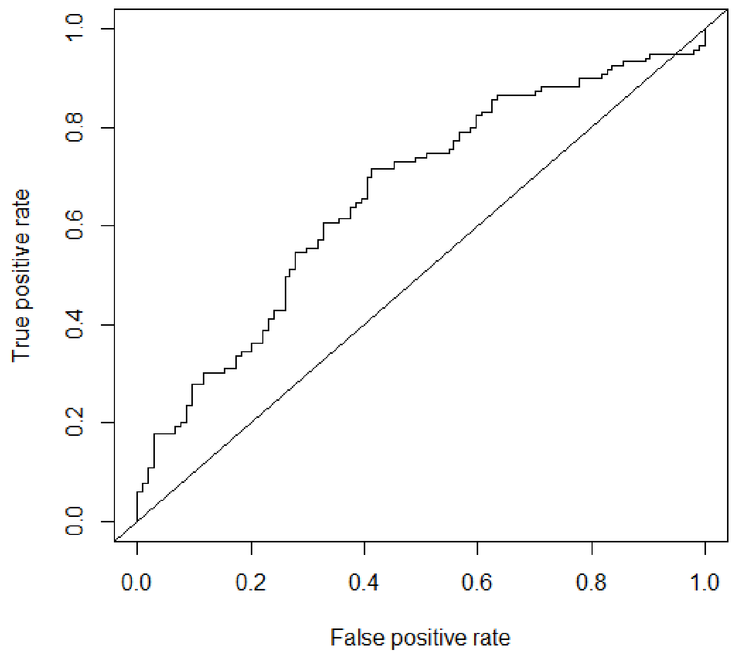

Figure 8-3: ROC Curve for Logistic Regression Model

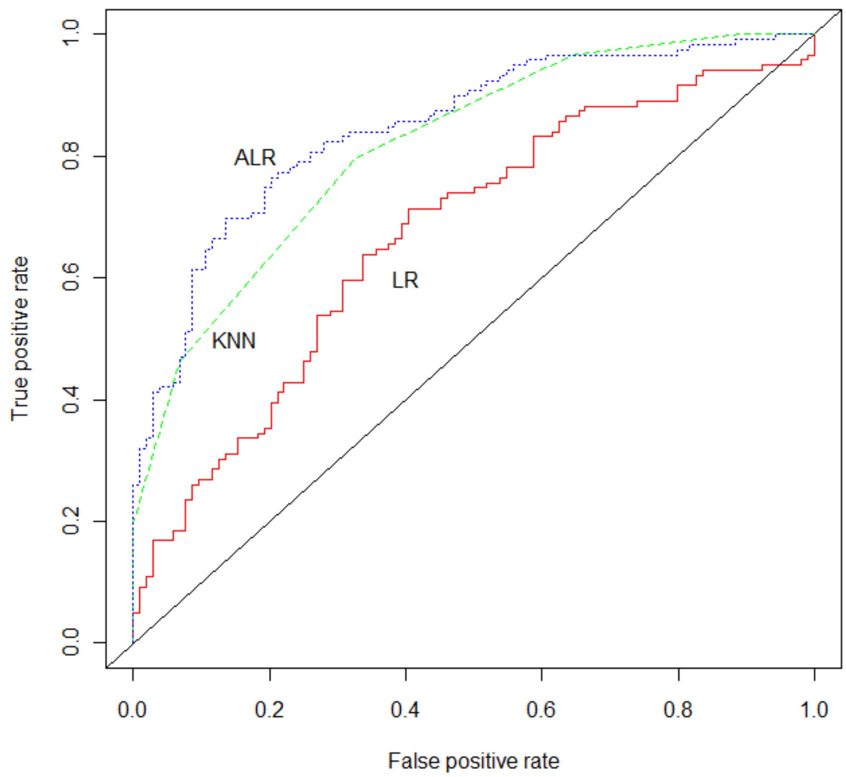

Figure 8-4: ROC curves for the logistic regression model (LR), the KNN model (KNN), and the augmented logistic regression model (ALR)

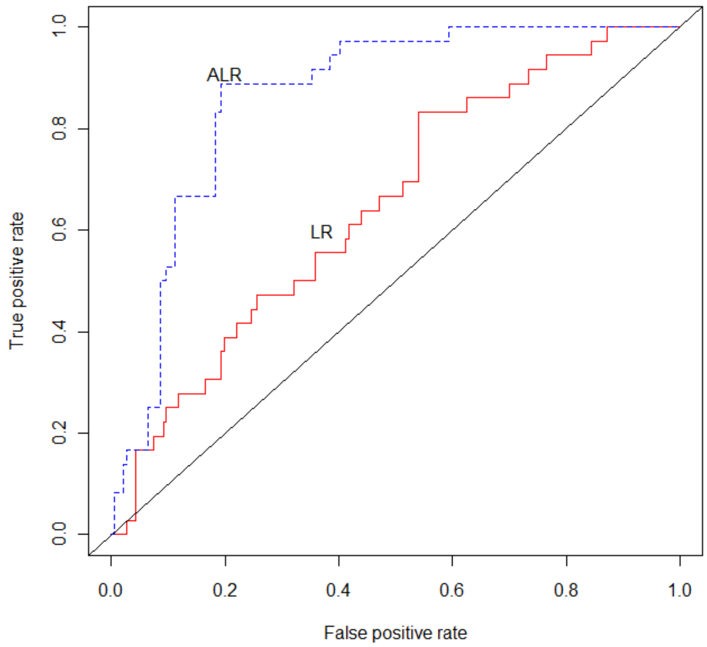

Figure 8-5: ROC curve for high severity cancer for the logistic regression model (LR) and the augmented logistic regression model (ALR)

Table 8-1: Confusion matrix for the logistic regression model

	Gleason Score	
	0 – 4	5 – 8
Predicted Healthy	59	34
Predicted Cancer	45	85

Table 8-2: Confusion matrix for the KNN model

	Gleason Score	
	0 – 4	5 – 8
Predicted Healthy	78	25
Predicted Cancer	26	94

Table 8-3: Breakdown in outcomes by number of cancerous neighbors

	Gleason Score	
Neighbors with Cancer	0 – 4	5 – 8
0	11	0
1	26	4
2	33	20
3	27	40
4	7	32
5	0	23

Table 8-4: Confusion matrix for the augmented logistic regression model

	Gleason Score	
	0 – 4	5 – 8
Predicted Healthy	79	22
Predicted Cancer	25	97

Table 8-5: Confusion matrix for high severity cancer for the augmented logistic regression model

	Gleason Score	
	0 – 4	5 – 8
Predicted Healthy	151	5
Predicted Cancer	36	31

Table 8-6: Confusion matrix for high severity cancer for the basic logistic regression model

	Gleason Score	
	0 – 4	5 – 8
Predicted Healthy	108	14
Predicted Cancer	79	22

CHAPTER 9: CONCLUSIONS AND FUTURE WORK

The research presented in this dissertation contributes to the literature on healthcare operations management in two key ways. First, we show that operational constraints and resource availability, such as staffing levels and bed utilization, impact medical decision making and the quality of care received by patients. This is important managerially, because it shows that operations management decisions cannot be made separately from medical decisions, but that the system must be viewed as a whole.

The decisions made, for example, about how to schedule surgeries have system-wide effects that must be taken into account. When surgeries with long lengths of stay are scheduled early in the week, recovery beds fill up, which lead to early discharges and a higher readmission rate. If the whole system were taken into account when surgeries were scheduled, we would be able to improve quality of care and throughput.

Similarly, we see how operations constraints impact patient outcomes when we examine how the quality of care hospitals provide changes over the course of a day. A lack of specialized resources available overnight leads to worse outcomes for patients. We show that operations management decisions, i.e., staffing levels and resource allocation, affect hospital efficiency and the quality of care delivered. Operations management as a discipline studies how processes can be made more efficient through intelligent allocation of resources. In this dissertation, we extended this to show that in addition in improving efficiency, operations management can also help hospitals improve the quality of care that they deliver as well.

The second main contribution is showing novel ways for medical data to be used in predicting patient risk. We built predictive models to address two problems faced by clinicians. The first problem we address is predicting which intra-hospital transfer patients will require an ICU bed, or die, within the first 48 hours of transfer. This work will allow our partner hospital prepare for incoming patients and deliver the appropriate quantity of care. The second model we built uses MRI data to diagnose prostate cancer. Our work shows that MRIs can be competitive with traditionally used diagnostic measures, but minimizes side-effects and costs.

These predictive models are important, because they demonstrate two ways that data can be used to predict medical outcomes. As healthcare information technology systems improve, there will be an explosion in the amount of medical data available to researchers. The ability to extract medically relevant information from these new data sources will help to improve medical practice. Having better advance information on patients should help doctors make better decisions while treating patients.

The growth of newly available data sources makes predictive modeling an attractive arena for future work. Electronic medical records, health insurance claims data, and large national databases all offer data that can be mined for medically relevant information. New algorithms to analyze new emerging medical data sources need to be developed an implemented. For example, models using either insurance claims data or electronic medical records could be built to predict which patients are at risk of negative outcomes like diabetes or heart disease that are potentially preventable. These data are available, and should have the richness of information needed to predict these negative outcomes. If the predictive models are good enough, this information could be used by clinicians to intervene.

Bibliography

Akaike H. A new look at the statistical model identification. *IEEE Trans. Automatic Control.* 1974; 19: 716-723.

Anderson D, Price C, Golden B, Jank W. Wasil, E. Examining the Discharge Practices of Surgeons at a Large Medical Center. *Health Care Management Science.* 2011; 14: 338-34.

Akdas A, Tarcan T, Turkeri L, Cevik I, Biren T, Gurmen N. The diagnostic accuracy of digital rectal examination, transrectal ultrasonography, prostate-specific antigen (PSA) and PSA density in prostate carcinoma. *British Journal of Urology.* 1995:76(s1), 54–56.

Amsellem-Ouazana D, Younes P, Conquy S, Peyromaure M, Flam T, Debré B, Zerbib M. Negative prostatic biopsies in patients with a high risk of prostate cancer. Is the combination of endorectal MRI and magnetic resonance spectroscopy imaging (MRSI) a useful tool? A preliminary study. *European Urology*. 2005:47(5), 582-586.

Anderson D. Golden B. Jank W, Wasil, E. The Impact of Hospital Utilization on Patient Readmission Rate. *Health Care Management Science.* 2012; 15: 29-36.

Anderson, G. F., Lave, J. R. Financing Graduate Medical Education Using Multiple Regression to Set Payment Rates. *Inquiry*, 1986; 23(2), 191-9.

Belien J, Demeulemeester E. Building Cyclic Master Surgery Schedules With Leveled Resulting Bed Occupancy. *European Journal of Operational Research*. 1986; 176(2). 1185-1204.

Bell CM, Redelmeier DA. Mortality Among Patients Admitted to Hospitals on Weekends as Compared with Weekdays. *New England Journal of Medicine*. 2001; 345:663–668.

Blake, J., Carter, M., Richardson, S. An Analysis of Emergency Room Wait Time Issues via Computer Simulation. *INFOR*. 1996; 34(4), 263-273.

Blake JT, Dexter F, Donald J. Operating Room Managers' Use of Integer Programming for Assigning Block Time to Surgical Groups: A Case Study. *Anesthesia & Analgesia*. 2002; 94:143-148.

Bohmer RMJ, Newell J, Torchiana DF. The Effect of Decreasing Length of Stay on Hospital Discharge Destination and Readmission After Coronary Bypass Operation. *Surgery* 2002;132:10–1.

Brailsford SC, Harper PR, Patel B, Pitt M. An Analysis of the Academic Literature on Simulation and Modeling in Health Care. *Journal of Simulation.* 2009;3:130-140.

Catalona WJ, Richie JP, Ahmann FR, Hudson MA, Scardino PT, Flanigan RC, deKernion JB, Ratliff TL, Kavoussi LR, Dalkin BL. Comparison of digital rectal examination and serum prostate specific antigen in the early detection of prostate cancer: results of a multicenter clinical trial of 6,630 men. *The Journal of Urology*. 1994:151(5), 1283-1290.

Campbell AJ, Cook JA, Adey G, Cuthbertson BH. Predicting Death and Readmission After Intensive Care Discharge. *British Journal of Anaesthesiology*. 2008; 100: 656–662.

Cardoen B, Demeulemeester E, Beliën J. Operating Room Planning and Scheduling: A Literature Review. *European Journal of Operational Research*. 2010: 201(3): 921-932.

Casaletto JA, Gatt R. Post-Operative Mortality Related to Waiting Time for Hip Fracture Surgery. *Injury*. 2004; 35(2): 114-120.

Ceglowski R, Churilov L, Wasserthiel J. Combining Data Mining and Discrete Event Simulation for a Value-Added View of a Hospital Emergency Department. *Journal of the Operational Research Society*. 2007;58:246-254.

Chan CW, Yom-Tov G, Escobar G. Does Speedup Reduce Congestion? An Examination of Intensive Care Units with Readmissions. Working Paper, Columbia University Business School. (2012).

Chan, L., & Kass, L. E. Impact of medical student preceptorship on ED patient throughput time. *The American Journal of Emergency Medicine*. 1999; 17(1), 41-43.

Chevalier P, Tabordon N. Overflow Analysis and Cross Trained Servers. *International Journal of Production Economics*. 2003; 85(1): 47– 60.

Chow G. Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica*. 1960; 28(3): 591-605.

Centers for Medicaid and Medicare Services, "Medicare Hospital Compare Quality of Care." 2013.

Coleman EA, Parry C, Chalmers S, Min SJ. The Care Transitions Intervention: Results of a Randomized Controlled Trial. *Archives of Internal Medicine*. 2006;166:1822-1828.

Cooper CP, Merritt TL, Ross LP, John LV, Jorgensen CM. To screen or not to screen, when clinical guidelines disagree: primary care physicians' use of the PSA test. *Preventive Medicine*. 2004:38(2), 182-191.

Cowper PA, DeLong ER, Hannan EL, Muhlbaier LH, Lytle BL, Jones RH, Holman WL, Pokorny JJ, Stafford JA, Mark DB, Peterson ED. Is Early Too Early? Effect of Shorter Stays After Bypass Surgery. *Annals of Thoracic Surgery.* 2007; 83: 100–107.

Cox DR. Regression Models and Life Tables. *Journal of the Royal Statistical Society*. 1972; 34(2). 187-220.

Custer, W., & Wilke, R. Teaching Hospital Costs: the Effects of Medical Staff Characteristics. *Health Services Research*. 1991; 25(6), 831-857.

DeBehnke, D. J. Clinical Supervision in the Emergency Department: a Costly Inefficiency for Academic Medical Centers. *Academic Emergency Medicine*. 2001; 8(8), 827-8.

Delaney CP, Fazio VW, Senagore AJ, Robinson B, Halverson AL, Remzi FH. 'Fast Track' Postoperative Management Protocol for Patients With High Co-Morbidity Undergoing Complex Abdominal and Pelvic Colorectal Surgery. *Br J Surg.* 2001; 88: 1533–1538.

Dexter F, Macario A, O'Neill L. Scheduling Surgical Cases into Overflow Block Time – Computer Simulation of Scheduling Strategies on Operating Room Labor Costs. *Anesthesia & Analgesia.* 2000;90(4):980-988.

Diette GB, et al. Comparison of Quality of Care by Specialist and Generalist Physicians as Usual Source of Asthma Care for Children. *Pediatrics*. 2001; 108(2): 432-437.

Dimick B. et al. Surgical Volume and Quality of Care for Esophageal Resection: Do High-Volume Hospitals Have Fewer Complications? *Annals of Thoracic Surgery*. 2003; 75(2): 337-341.

Dobson G, Lee H-H, Pinker E. 2011. A Model of ICU Bumping. Working Paper. 2011.

Dobson G, Hasija, Pinker EJ. Reserving Capacity for Urgent Patients in Primary Care. *Production and Operations Management*. 2011; 20: 456-473.

Dowd, M. D., Tarantino, C., Barnett, T. M., Fitzmaurice, L., & Knapp, J. F. Resident Efficiency in a Pediatric Emergency Department. *Academic Emergency Medicine*. 2005; 12(12), 1240-4.

Downs SH, Black N. The Feasibility of Creating a Checklist for the Assessment of the Methodological Quality Both of Randomised and Non-Randomised Studies of Health Care Interventions. *Journal of Epidemiology and Community Health*. 1998; 52:377–84.

Eappen, S., Flanagan, H., & Bhattacharyya, N. Introduction of Anesthesia Resident Trainees to the Operating Room Does Not Lead to Changes in Anesthesia-Controlled Times for Efficiency Measures. *Anesthesiology*. 2004; 101(5), 1210-4.

Egol KA et al. Mortality Rates Following Trauma: The Difference is Night and Day. Journal of Emergencies. *Trauma and Shock.* 2011; 4(2): 178-183.

Fawcett, Tom. An introduction to ROC analysis. *Pattern Recognition Letters*. 1977: 27, 861–874.

Fineberg, D. A., & Stewart, M. M. Analysis of Patient Flow in the Emergency Room. *The Mount Sinai Journal of Medicine*. 1977; 44(4), 551-9.

Fone D, Hollinghurst S, Temple M, Round A, Lester N, Weightman A, Roberts K, Coyle E, Bevan G, Palmer S. Systematic Review of the Use and Value of Computer Simulation Modeling in Population Health and Health Care Delivery. *Journal of Public Health Medicine.* 2003; 25(4), 325-335.

Fomundam, S, Herrmann J. A Survey of Queuing Theory Applications in Healthcare. Working Paper. Digital Repository at the Univeristy of Maryland. 2007.

Fritze, J. Medical Expenses Have 'Very Steep Rate of Growth'. *USA Today*. 2010; 2/4/2010.

Gawande A. The Checklist Manifesto: How to Get Things Right. New York, NY: Metropolitan Books. 2009.

Gleason DF. The Veteran's Administration Cooperative Urologic Research Group: histologic grading and clinical staging of prostatic carcinoma. *Urologic Pathology: The Prostate*. 1977: 171–198.

Gorunescu F, McClean I, Millard PH. A Queueing Model for Bed-Occupancy Management and Planning of Hospitals. *Journal of the Operational Research Society.* 2002;53(1):19-24.

Green L. Queueing Analysis in Healthcare. In *Patient Flow: Reducing Delay in Healthcare Delivery* (RW Hall, ed.). Springer, 2006; 281-307.

Green LV, Soares J, Giglio JF, Green RA. Using Queueing Theory to Increase the Effectiveness of Emergency Department Provider Staffing. *Academic Emergency Medicine : Official Journal of the Society for Academic Emergency Medicine*. 2006;*13*(1), 61-68.

Hasan M. Readmission of Patients to Hospital: Still Ill-Defined and Poorly Understood. *International Journal for Quality in Health Care*;. 2001; 13: 177–179.

Harvey, M., Al Shaar, M., Cave, G., Wallace, M., & Brydon, P. Correlation of Physician Seniority With Increased Emergency Department Efficiency During a Resident Doctors' Strike. *New Zealand Medical Journal*. 2008; 121(1272).

Haynes AB, Weiser TG, Berry WR, et al. A Surgical Safety Checklist to Reduce Morbidity and Mortality in a Global Population. *New England Journal of Medicine*. 2009; 360:491–499.

Heron M, Hoyert DL, Murphy SL, Xu J, Kochanek KD, Tejada-Vera B. National Vital Statistics Reports. Atlanta (GA): Centers for Disease Control and Prevention. 2009.

Hilbe J. *Logistic Regression Models*. CRC Press; 2009.

Hoffman RM, Gilliland FD, Adams-Cameron M, Hunt WC, Key CR. Prostate-specific antigen testing accuracy in community practice. *BMC Family Practice.* 2002: 3:19.

Hollingsworth, B. Non-Parametric and Parametric Applications Measuring Efficiency in Health Care. *Health Care Management Science*. 2003; 6(4), 203-218.

Howlader N, Noone AM, Krapcho M, Neyman N, Aminou R, Altekruse SF, Kosary CL, Ruhl J, Tatalovich Z, Cho H, Mariotto A, Eisner MP, Lewis DR, Chen HS, Feuer EJ, Cronin KA. *SEER Cancer Statistics Review, 1975-2009 (Vintage 2009 Populations)*. 2012: National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2009_pops09/

Huckman, R., & Barro, J. Cohort Turnover and Productivity: the July Phenomenon in Teaching Hospitals. NBER. 2005.

Hughes RG, Hunt SS, Luft HS. Effects of Surgeon Volume and Hospital Volume on Quality of Care in Hospitals. *Medical Care*. 1987; 25(6): 489-503.

Hwang SW, Li J, Gupta R, Chien V, Martin RE. What Happens to Patients Who Leave Hospital Against Medical Advice? *Canadian Medical Association Journal*. 2003; 168:417–420.

Jacobson SH, Hall SN, Swisher JR. Discrete-Event Simulation of Health Care Systems. In *Patient Flow: Reducing Delay in Healthcare Delivery* (RW Hall, ed.). Springer, 2006; 211-252.

Jun JB, Jasobson SH, Swisher JR. Application of Discrete-Event Simulation in Health Care Clinics: A survey. *Journal of the Operational Research Society*. 1999:50(2):109-123.

Kc, D., Terwiesch C. Impact of Workload on Service Time and Patient Safety: An Econometric Analysis of Hospital Operations. *Management Science*. 2009; 55(9): 1486-1498.

Kc, D., Terwiesch C. An Econometric Analysis of Patient Flows in the Cardiac ICU. *Manufacturing and Service Operations Management*. (2011). Forthcoming.

Keehan S, Sisko A, Truffer C, Smith S, Cowan C, Poisal J, Clemens KM. Health Spending Projections Through 2017: The Baby-Boom Generation is Coming to Medicare. *Health Affairs*. 2008; 27:W145-55

Kurhanewicz, J, Swanson MG, Nelson SJ, Vigneron DB. Combined magnetic resonance imaging and spectroscopic imaging approach to molecular imaging of prostate cancer. *Journal of Magnetic Resonance Imaging*. 2002:16, 451–463.

Lam SK. SERVQUAL: A Tool for Measuring Patients' Opinions of Hospital Service Quality in Hong Kong. Total Quality Management. 1997; 8(4): 145-152.

Lammers, R. L. The Effect of a New Emergency Medicine Residency Program on Patient Length of Stay in a Community Hospital Emergency Department. *Academic Emergency Medicine*. 2003; 10(7), 725-730.

Lehaney B, Hlupic V. Simulation Modeling for Resource Allocation and Planning in the Health Sector. *The Journal of the Royal Society for the Promotion of Health*. 1995:115(6), 382-385.

Locker TE, Mason SM. Analysis of the Distribution of Time That Patients Spend in Emergency Departments. *British Medical Journal*. 2005; 330:1188-9.

Magid DJ, Wang Y, Herrin J, McNamara RL, Bradley EH, Curtis JP, et al. Relationship Between Time of Day, Day of Week, Timeliness of Reperfusion, and in-Hospital Mortality for Patients with Acute ST-segment Elevation Myocardial Infarction. *JAMA*. 2005; 294:803–12.

McClellan M, Staiger D. Comparing Hospital Quality at For-Profit and not For-Profit Hospitals.: National Bureau of Economic Research. 1999. NBER Working Paper no. 7324.

McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, Kerr EA. The Quality of Health Care Delivered to Adults in the United States. *New England Journal of Medicine*. 2003; 348: 2635–2645.

McManus ML, Long MC, Cooper A, et al. Variability in Surgical Caseload and Access to Intensive Care Services. *Anesthesiology*. 2003; 98:1491–1496.

McManus ML, Long MC, Cooper A, Mandell J, Berwick D, Pagano M, and Litvak E Variability in Surgical Case Load and Access to Intensive Care Services. *Anesthesiology.* 2003; 98(6):1491-1496.

Millard PH, Christodoulou G, Jagger C, Harrison GW, McClean SI. Modeling Hospital and Social Care Bed Occupancy and Use by Elderly People in an English Health District. *Health Care Management Science.* 2001; 4: 57-62.

Miró O et al. Decreased Health Care Quality Associated with Emergency Department Overcrowding. European Journal of Emergency Medicine: *Official Journal of the European Society for Emergency Medicine*. 1999; 6(2): 105-107.

Niehaus D, Koen L, Galal U, Dhansay K, Oosthuizen P, Emsley R, Jordan E. Crisis Discharges and Readmission Risk in Acute Psychiatric Male Inpatients. *BMC Psychiatry.* 2008; 1:8-44.

Nozaki, S. Ross, S. Approximations in Finite-Capacity Multi-Server Queues with Poisson Arrivals. *Journal of Applied Probability*. 1978; *15*(4): 826-834.

Offner, P. J., Hawkes, A., Madayag, R., Seale, F., & Maines, C. General Surgery Residents Improve Efficiency but not Outcome of Trauma Care. *The Journal of Trauma*. 2003; 55(1), 14-9.

O'Neill, L., & Dexter, F. Evaluating the Efficiency of Hospitals' Perioperative Services Using DEA. *Operations Research and Health Care.* 2005; 70(2), 147-168.

Padhani AR, Gapinski CJ, Macvicar DA, Parker GJ, Suckling J, Revell PB, Leach MO, Dearnaley DP, Husband JE. Dynamic contrast enhanced MRI of prostate cancer: correlation with morphology and tumour stage, histological grade and PSA. *Clinical Radiology* 2000:55(2), 99-109.

Philibert, I. New Requirements for Resident Duty Hours. *JAMA: The Journal of the American Medical Association*. 2002; 288(9), 1112-1114.

Pinker EJ, Shumsky RA. The Efficiency-Quality Trade-off of Cross-Trained Workers. *Manufacturing and Service Operations Management*. 2000; 2(1): 32– 48.

Price C, Babineau T, Golden B, Harrington M, Wasil E. Capacity Management in a Cardiac Surgery Line. presented at INFORMS 2007 Meeting, Seattle, WA.

Price C Applications of Operations Research Models to Problems in Health Care. Ph.D. Thesis. 2009. University of Maryland. College Park, MD.

Price C, et al. Reducing Boarding in a Post-Anesthesia Care Unit. *Production and Operations Management.* 2011; 20(3): 431-441.

Reeves MJ, Smith E, Fonarow G, Hernandez A, Pan W, Schwamm LH. Off-hour Admission and In-Hospital Stroke Case Fatality in the Get With the Guidelines-Stroke Program. *Stroke*. 2009; 40: 569–76.

Rogowski, J. A., & Newhouse, J. P.. Estimating the Indirect Costs of Teaching. *Journal of Health Economics*. 1992; 11(2), 153-171.

Roth AV, Dierdonck R. Hospital Resource Planning: Concepts, Feasibility, and Framework. *Production and Operations Management*. 1995; 4(1): 2-29.

Rosko, M. D. Understanding Variations in Hospital Costs: An Economics Perspective. *Annals of Operations Research*. 1996; 67(1), 1-21.

Salazar, A., Corbella, X., Onaga, H., Ramon, R., Pallares, R., & Escarrabill, J. Impact of a Resident Strike on Emergency Department Quality Indicators at an Urban Teaching Hospital. *Academic Emergency Medicine. 2001;* 8(8), 804-808.

Saposnik G, Baibergenova A, Bayer N, Hachinski V. Weekends: A Dangerous Time for Having a Stroke? *Stroke*. 2007; 38:1211–5.

Sarkis, J., & Talluri, S. Efficiency Measurement of Hospitals: Issues and Extensions. *International Journal of Operations & Production Management*. 2002; 22(3), 306-313.

Singer DE, Carr P L, Mulley A G, Thibault GE. Rationing intensive care–physician responses to a resource shortage. *New England Journal of Medicine*. 1983; 309(19):1155-1160.

Singer JD, Willet JB. It's About Time – Using Discrete-Time Survival Analysis to Study Duration and the Timing of Events. *Journal of Educational Statistics.* 1993; 18:155-95.

Smith DS, Catalona WJ. Interexaminer variability of digital rectal examination in detecting prostate cancer. *Urology*. 1995:45(1), 70-74.

Soteriou AC, Hadjinicola GC. Resource Allocation to Improve Service Quality Perceptions in Multistage Service Systems. *Production and Operations Management*. 1999; 8(3): 221-239.

Strauss M J, LoGerfo JP, Yeltatzie JA, Temkin N, Hudson LD. Rationing of Intensive Care Unit Services: An Everyday Occurrence. *Journal of the American Medical Association. 1986;* 255(9):1143-6.

Swisher, J., & Jacobson, S. Evaluating the Design of a Family Practice Healthcare Clinic Using Discrete-Event Simulation. *Health Care Management Science*. 2002; 5(2), 75-88.

Theokary C, Zhong JR. An Empirical Study of the Relations Between Hospital Volume, Teaching Status, and Service Quality. *Production and Operations Management*. 2011; 20(3): 303-318.

Thomas JW, Guire KE, Horvat GG. Is Patient Length of Stay Related to Quality of Care? *Hospital & Health Services Administration*. 1997; 42(4): 489-507.

Thomas JW, Holloway JJ, Guire KE. Validating Risk-Adjusted Mortality as an Indicator for Quality of Care. *Inquiry: A Journal of Medical Care Organization, Provision and Financing*. 1993; 30(1): 6-22.

Trzeciak S. Emergency Department Overcrowding in the United States: An Emerging Threat to Patient Safety and Public Health. *Emergency Medicine Journal*. 2003; 20(5): 402-405.

U.S. News & World Report. America's best hospitals. June 22, 2012.

Vandamme R, Leunis J. Development of a Multiple-item Scale for Measuring Hospital Service Quality, *International Journal of Service Industry Management*. 1993; 4(3): 30 – 49.

Welch HG, Albertsen PC. Prostate cancer diagnosis and treatment after the introduction of prostate-specific antigen screening: 1986-2005. *Journal of the National Cancer Institute.* 2009:101(19), 1325-9.

Welch, W. P. Do All Teaching Hospitals Deserve an Add-On Payment Under the Prospective Payment System? *Inquiry*, 2987; 24(3), 221-32.