An Econometric Analysis of Patient Flows in the Cardiac ICU

Diwas KC, Christian Terwiesch
Operations and Information Management Department
The Wharton School, University of Pennsylvania
November 28, 2007

Abstract

This paper explores the rationing of bed capacity in a cardiac intensive care unit (ICU). We find that the length of stay for patients admitted to the ICU is influenced by the occupancy level of the ICU. In particular, a patient is likely to be discharged early when the occupancy in the ICU is high. This in turn leads to an increased likelihood of the patient having to be readmitted to the ICU at a later time. Such “bounce-backs” have implications for the overall ICU effective capacity – an early discharge immediately frees up capacity, but at the risk of a (potentially much higher) capacity requirement when the patient needs to be readmitted. We analyze these capacity implications, shedding light on the question if an ICU should apply an aggressive discharge strategy or if it should follow the old quality slogan and “do it right the first time.” By comparing the total capacity usage for patients who were discharged early versus those who were not, we show that an aggressive discharge policy increases the net peak capacity in the ICU, despite the associated increase in readmissions.

The authors are grateful to Chris Lee, Jiri Horak, Jack Iwashinya, Devin Pope and the participants in the Empirical Operations Management Workshop held at the Wharton School. They can be reached at diwas@wharton.upenn.edu and terwiesch@wharton.upenn.edu.
1 Introduction

According to the Center for Disease Control and Prevention (2005), cardiovascular diseases affect one out of every three adults and are the leading cause of death of both men and women in the United States. Cardiac surgery also is one of the key contributors to hospital profits, making the topic interesting not just from the perspective of society but also the one of hospital management.

Numerous studies (Hall 2006, IHI 2007) have found that resource constraints often plague patient flows in hospitals in general and in particular cardiac care. A resource constraint at one stage can lead to delays, congestion and overall reduction in patient throughput for the hospital. For example, in the cardiac care process, the surgical Intensive Care Unit (ICU) has often been identified as the process bottleneck. The ICU is an expensive resource with the cost of patient care being multiple times higher than in a regular ward (see e.g. Henning et al 1987). Consequently, many ICU’s operate at high levels of occupancy, leading to increased waiting times upstream of the ICU and an overall reduction in patient throughput (see also McConnell et al 2005).

Given the scarce ICU capacity, hospitals are often forced to ration the available ICU capacity. This means that when the ICU reaches its full occupancy, the healthiest (in relative terms) patient gets discharged, more or less independent of their absolute health condition. While such early discharges clearly increase patient throughput in the short-term, they have the potential to lead to medical complications and to increase the likelihood that a patient has to revisit the ICU in the future.

It is this interplay between the medical variables that determine the ICU length of stay of a patient and the operational variables such as ICU occupancy and capacity rationing that are at the heart of this paper. Based on a study of 1365 cardiothoracic patients in a large US teaching hospital, including their medical records as well as detailed operational data of patient flow, we develop an econometric model of patient recovery, discharge from the ICU and potential readmission to the ICU. This allows
us to make the following three contributions.

**First**, we estimate the impact of ICU occupancy on the ICU length of stay of a patient. This allows us to study the discharge pattern of the ICU. We show that a patient who arrives at a busy ICU has an expected length of stay that is 11% shorter compared to a patient (with similar medical conditions) that arrives at a lower level of occupancy.

**Second**, we show that this capacity rationing behavior has serious medical implications. Specifically, we show that a patient who is discharged early due to a high occupancy level is 18% more likely to be readmitted to the ICU at some later time (creating a so called “bounce-back”). Moreover, we show that patients have a dramatically longer length of stay in the ICU when they are admitted to the ICU for a second stay.

**Third**, we analyze the capacity implications of the hospital’s discharge pattern. An early discharge immediately frees up ICU capacity, but at the risk of a (much higher) capacity requirement upon readmission. By comparing the total capacity usage for patients who were discharged early versus those who were not, we show that an aggressive discharge policy increases the net peak capacity in the ICU, despite the associated increase in readmissions.

The remainder of this paper is organized as follows. We first discuss the recovery process of cardiac patients with a focus on the ICU followed by a review of the relevant literature. We then develop our models and present econometric specification to test them. In section 6, we provide a description of our data collection. Finally we present our results and conclude with a discussion of the implications of our findings.

## 2 Process Description

After a cardiac patient is admitted to the hospital, a number of pre-surgery diagnostic tests are conducted and the patient is prepared for surgery. These activities are
referred to as the pre-operative stage. Immediately following surgery, the patient is taken to the intensive care unit (ICU). At this point, the patient is typically unconscious and is on breathing assistance via a ventilator. In the immediate post-operative stage, various medications are administered to sedate and stabilize the patient. The patient is under constant monitoring, often requiring a one-to-one patient-to-nurse ratio during the first twelve hours following surgery and a physician available to attend if any complications arise.

Following discharge from the ICU, the patient is taken to a step-down unit (often referred to as “the floor”). The step-down unit has reduced intensity of treatment and monitoring. For example, the patient may no longer be on heavy medication and is typically no longer on ventilator support. Also, an attending physician may not be available immediately day and night. If serious complications arise while the patient is in the step-down unit requiring increased level of care and monitoring, the patient is readmitted to the ICU. For the majority of patients, however, no significant complications arise, and after a period of stay in the step down unit, the patient is discharged from the hospital.

This process is common across hospitals, including the hospital underlying this research study, a large teaching hospital in the US. The hospital is ranked as one of the top 20 hospitals in cardiac surgery in the US and performs over 1200 cardiothoracic surgeries a year. Given this size, the hospital has an 18-bed ICU dedicated to cardiothoracic care. A typical cardiothoracic surgery patient spends a median of 1.1 days in the first visit to the ICU and 11 days in the hospital. About 14% of the patients require readmission to the ICU.
3 Literature Review

Capacity planning in healthcare delivery has been an active and fruitful area for research in Management Science and Operations Research. Previous studies (e.g. Kwak and Lee 1997, Green and Meissner 2002, Huang 1995, Green 2003) have looked at the sizing of capacity including beds, equipment and number of staff. This stream of research is quite extensive and we refer the readers to Smith-Daniels et al (1988) and Green (2004) for more comprehensive overviews. Queueing theory is one of the most commonly used analytical methods to describe care processes because of the the stochastic nature of demand and service rates in healthcare as well as the importance of performance measures such as waiting time, queue length, or turn-away probability.

A common assumption in the previous body of literature is that the service rate is drawn from a probability distribution (usually exponential) that is exogenous and independent of the current state of the system, including the number of people waiting in line. However, there exist several papers, some analytical and some empirical, that challenge this independence assumption and postulate that resources should increase their service rate when the load on the system is high. This literature of dynamic queueing control started with a set of analytical models, including work by Bertsekas (2000), George and Harrison (2001), Stidham and Weber (1989) and Crabill (1972), that derive the optimal service rate that balances the costs of acceleration with the costs of waiting times. Collectively, these papers show that it is optimal for resources to accelerate as the length of the queue increases.

From an empirical perspective, using data from two distinct healthcare services, patient transport and cardiothoracic surgery, Kc and Terwiesch (2007) validate that workers adapt to increasing levels of load in the system by increasing their service rate. The authors also show that such temporary service rate increases are not sustainable (i.e., workers become fatigued) and can have potentially serious quality implications. The results obtained by Kc and Terwiesch complement a set of prior lab experiments conducted by Schultz et al (1998, 1999) that establish that workers in an assembly
line adjust their service rates in response to the amount of work in process inventory between the workers.

Just like the literature reviewed above, the theory underlying our work also is in the tradition of optimal queueing control and we empirically investigate the relationship between system load and service rate. However, what sets the present paper apart from the prior literature is its focus on rework. We consider a service setting in which the server has the option to “rush” customers currently waiting for service. While such rushing immediately increases service capacity, it comes at the risk that the customer has to be reworked at a later point in time. This, potentially much longer, rework has a negative impact on service capacity, leaving the server with a decision of “rushing now and reworking later” or “doing it right the first time”.

The quality management literature has taken a firm point on this decision. Rework is seen as one of the seven sources of waste initially observed by the Japanese production movement (see e.g. Ohno 1988). For example, in their work of benchmarking automotive production plants across the world, Womack et al (1990) found that GM’s Framingham plant spent over 40 hours on the average vehicle, including the rework of 1.3 defects per vehicle, while Toyota in its Takaoka plant only needed 18 hours, largely reflecting substantially less time wasted on rework. In his description of Toyota’s production system, Liker (2004) emphasizes the importance of getting quality right immediately as opposed to relying on rework downstream in the assembly line. This quality paradigm as also been analyzed in healthcare operations. Tucker (2004) finds in her study of nursing work that nurses waste a large part of their time reworking what either they themselves or other members of the care process got wrong earlier on.

Given that effective capacity allocation is critical to improving hospital revenues, Roth and van Dierdonck (1995) have developed a paradigm they define as Hospital Resource Planning (HRP). They find that anticipating and allocating resource needs more accurately is an effective way for hospitals to match capacity to patient-specific
needs. In our paper, in addition to anticipating ICU bed usage based on individual patient characteristics, we empirically examine whether such adaptive discharge decisions in the ICU can in fact lead to a performance improvement in the hospital.

While our paper is written to contribute to the literature in Operations Management, we also draw on the medical literature to appropriately capture various patient level risk factors and their impact on clinical outcomes as well as ICU capacity consumption. In particular, we apply a widely used risk stratification method called EuroSCORE (Nashef et al. 2002, Kurki et al. 2002, Tournouilis et al. 2005) to assess the impact of an early discharge on the likelihood of a bounceback to the ICU. We also build on previous work in the medical literature to develop models of patient recovery (Peake et al. 2006).

4 Model Development

Cardiac surgery can be seen as a substantial “shock” to the patient’s body. Thus, the time in the ICU following surgery is primarily one of stabilization and recovery. Major milestones in this recovery process include the removal of ventilator assistance and the weaning off from heavy medication.

First and foremost, healing and recovery require time. Consequently, everything else equal, the longer a patient has spent in the ICU, the more likely she is to be ready for discharge. In other words, the hazard \( h_i(t) \) of patient \( i \) being discharged at any given time \( t \) increases with the time spent in the ICU.

\[
\frac{\partial h_i(t)}{\partial t} > 0
\]

This formulation is in line with an extensive body of research in biostatistics, modeling the effect of time on patient recovery. Let \( S_i(t) \) denote the probability that patient \( i \) remains in the ICU at a given point \( t \):

\[
S_i(t) = \Pr\{LOS_i > t\}
\]
where $LOS_i$ is patient $i$’s length of stay in the ICU.

Patients admitted to the ICU have are heterogeneous in their medical conditions, in other words, the risk of complications and the case severity vary from patient to patient. A single bypass procedure performed on a 50 year old simply has a lower risk associated than a triple bypass surgery of an 80 year old. A higher case severity typically requires a longer time for recovery, i.e. a longer stay in the ICU.

We use the accelerated failure time (AFT) analysis to model the impact of severity on the relative recovery times and lengths of stay of patients in the ICU. The AFT model assumes that patient $j$ is likely to stay $\phi_{ij}$ times longer in the ICU than patient $i$, where $\phi_{ij}$ is a constant specific to patients $i$ and $j$ describing the relative risk levels of the two patients. Thus, we can write the probabilities of any two patients $i$ and $j$ remaining in the ICU as:

$$S_i(t) = S_j(\phi_{ij}t) \quad \forall \ t$$  \hspace{1cm} (3)

As the occupancy level in the ICU increases, fewer beds are available to accommodate the inflow of new patients from the operating room. If the ICU is full (i.e. all ICU beds are occupied), the hospital has to ration the ICU capacity. Since the patients coming from the operating room are typically the ones that are in the biggest need of ICU care (capacity), they will always be guaranteed a bed. Hence, as a result of the overall ICU capacity constraint, some other patient has to leave the ICU. This is the patient that is the healthiest from among the current ICU population. Note though, that in absence of the high ICU occupancy, this same patient would have spent a longer time in the ICU, making his discharge a result of operational variables as opposed to medical variables alone.

To formalize this logic, we hypothesize that a patient in a busy ICU will have a shorter length of stay ($LOS_i$) than a patient in a less busy ICU. In short,

$$\frac{\partial LOS_i}{\partial OCCUPANCY_i} < 0$$  \hspace{1cm} (4)

where $OCCUPANCY_i$ is the occupancy level associated with the stay of patient $i$. The alternative hypothesis is that the hospital, when facing a busy ICU, cancels
or delays new surgeries and thereby allows the current ICU patients to stay in the ICU as long as medically necessary.

From a medical perspective, a longer length of stay increases the likelihood of a more complete patient recovery in the ICU. A patient who is discharged early for operational reasons related to ICU occupancy, i.e. who would have spent a longer time in the ICU if it were for medical considerations alone, is at an increased risk of experiencing complications outside of the ICU. Adjusting for medical risk factors, we therefore postulate that a patient who is discharged early has a higher likelihood of a revisit to the ICU. In other words, the greater the length of stay, the lower the likelihood of a patient bouncing back to the ICU:

\[ \frac{\partial \Pr_{BB,i}}{\partial LOS_i} < 0 \] (5)

where \( \Pr_{BB,i} \) is the probability that patient \( i \) bounces back. The alternative competing hypothesis is that a patient is only discharged after it is absolutely safe for the patient. The alternative hypothesis is that a shortened length of stay has little or no impact on the likelihood of a bounce-back.

Finally, consider the overall capacity implications of an early discharge. If a patient is more likely to bounce back to the ICU when discharged early, the discharge decision has implications for the total ICU capacity consumption of that patient. Define the total expected length of stay (\( TOTAL\_LOS_i \)) for patient \( i \), including the initial length of stay as well as the future length of stay associated with a potential readmission, as:

\[ TOTAL\_LOS_i = LOS_i + \Pr_{BB,i}(LOS_i) * LOS\_REVISIT_i \] (6)

where \( LOS\_REVISIT_i \) is the revisit length of stay and \( \Pr_{BB,i}(LOS_i) \) is the likelihood of a patient \( i \)'s revisit when the initial length of stay is \( LOS_i \).

Earlier, we postulated that a shorter initial LOS increases the likelihood of a bounceback. If this is true, there exists an optimal LOS that minimizes \( TOTAL\_LOS \).
In other words, the ICU faces a trade-off between discharging a patient early ("rushing a patient"), in which case the initial length of stay is short, but the bounce-back probability is high, and following a more conservative discharge policy ("doing it right the first time"), in which case the initial length of stay is long, but the bounce-back probability is low.

The total length of stay in the ICU is not the only performance measure the hospital cares about. Under the diagnosis related group (DRG) payment system, a hospital is reimbursed a fixed payment amount depending on the diagnosis for the patient, irrespective of the actual cost incurred by the hospital. A hospital thus has little financial incentive to keep a patient longer in the ICU than necessary. The operational performance measure that maximizes hospital revenues is thus the overall patient throughput.

By definition, the early discharge of a patient as a result of capacity rationing happens at a time when the ICU is capacity constrained. Saving a patient day in the ICU at that time and having the patient come back at some point in the future may or may not increase the patient throughput in the ICU. In particular, if the ICU is rarely capacity constrained, it is likely that a future readmission of a patient discharged early will occur at a time when there exists excess ICU capacity. In this case, an early discharge helps increase the overall patient throughput. If however, the readmission happens at a time when the ICU is (still or again) capacity constrained, the overall patient throughput would decrease because of the early discharge.

We assess the throughput implications of the early discharge decision (long initial length of stay vs short initial length of stay) by estimating its impact on the peak ICU capacity. Unlike our previous analysis of patient length of stay, this peak capacity calculation explicitly considers if the ICU is currently capacity constrained or not. Let \( 1_{\{ICU=BUSY\}} \) be an indicator function indicating if the ICU is at full occupancy (and thus presently operating at peak capacity). \( 1_{\{ICU=BUSY\}} = 1 \) if the ICU is at capacity and \( 1_{\{ICU=BUSY\}} = 0 \) otherwise. We can estimate the peak bed capacity
consumption of a patient by computing:

\[ \text{LOS}_i \cdot 1_{\{ICU=BUSY\}} + \text{Pr}_{BB,i} (\text{LOS}_i) \cdot \text{LOS}_i \text{REVISIT}_i \cdot 1_{\{ICU=BUSY\}} \]  

(7)

If an aggressive discharge decision decreases the peak bed capacity consumption, it helps to increase the overall patient throughput in the ICU. Put to the extreme, a patient staying in a half empty ICU for one week has a lower peak bed capacity consumption than a patient who just spends one day in an ICU that is full.

5 Econometric Specifications

We model the length of stay (LOS) of a patient in the ICU using the Weibull distribution. We test the validity of this assumption by fitting our model against the log-normal and the exponential distributions. We find that the Weibull and the exponential have similar levels of fit, while the log-normal fits our data quite poorly. Thus, we can safely reject the model \((p < 0.001)\) in which the ICU length of stay is log-normally distributed in favor of either the Weibull or the exponential. Since the Weibull approximates a wide range of commonly used distributions, including the normal and the exponential, we use this distribution for the remainder of our analysis. It is also the distribution that is typically used for AFT models.

With the Weibull distribution, the survival function, \(S_t(\text{LOS}_i)\) is given by:

\[ S_t(\text{LOS}_i) = \exp[-\text{LOS}_i \cdot e^{(-X_i \beta)^{1/\sigma}}] \]  

(8)

where the variables in \(X_i\) capture the various patient-level and system-level factors that affect the patient’s length of stay. For example, patient-level variables are age or procedure type while system-level variables are the ICU occupancy or the day of the week. The flexibility offered by the Weibull distribution allows us to model increasing, decreasing and constant hazard rates. In particular, \(\sigma > 1\) implies an increasing hazard rate.\(^1\)

\(^1\)An alternative model formulation is in terms of the hazard rate. \(h(\text{LOS}_i) = \alpha \cdot \text{LOS}_i + X_i \zeta\)
Given that the \( LOS \) has a Weibull distribution, we obtain the following econometric specification:

\[
\log(LOS_i) = X_i \beta + \sigma * e_i
\]  

(9)

5.1 Effect of Load

The above specification states that the \( LOS \) in the ICU in the absence of capacity constraints can be explained by the parameters in \( X \). To assess the effect of ICU occupancy on the length of stay, we append the binary variable \( BUSY \) in the following expanded specification. A value of \( BUSY = 1 \) denotes that the ICU is at high occupancy and a value of 0 indicates otherwise:

\[
\log(LOS_i) = X_i \beta + \gamma * BUSY_i + \sigma * \epsilon_i
\]  

(10)

The coefficient \( \gamma \) provides us the estimate of the effect of occupancy on the length of stay of the patient. We assume that \( \epsilon_i \) has a constant mean and variance. We also assume that the only source of autocorrelation between consecutive \( LOS \) observations is through the occupancy in the ICU. After we include \( BUSY \) explicitly in our model, we can assume that error terms \( \epsilon_i \) are independent across observations.

5.2 Effect of Early Discharge on Bounce Back

We next estimate the likelihood of a patient revisiting the ICU upon being discharged early. We define the variable \( LOS\_EARLY \) to be an estimate of the number of days by which a patient was discharged early. Estimating \( LOS\_EARLY \) first requires us to know what the length of stay would have been for a given patient in the absence of a capacity constraint. In other words, we would like to estimate the the length of stay in the ICU determined by medical variables alone, free from any polluting impacts of time pressure created by capacity rationing. We estimate this variable,
which we define as $LOS_{MED}$, by using the non-parametric method of Propensity Scores Matching.

To do this, we first divide up the patient population into two separate groups – a group that is admitted to the ICU when it is busy and a group that is admitted when the ICU is not busy (Figure 2). We assume that the time of admission and hence the assignment to these two groups is random and independent of the severity of the case. Note that emergency admissions are, by definition, always random. About 23% of all admissions fall into that category. Moreover, we found in our discussion with the doctors working in cardiac care that the state of the ICU was not considered when scheduling new surgeries in pre-operative planning. To validate this assumption of independence between case severity and occupancy, we computed the correlation between the level of pre-operative severity, as measured by the New York Heart Association Severity index, and the occupancy in the ICU. The resulting correlation coefficient is not statistically different from zero ($p = 0.32$). Given this independence between case severity and occupancy, we can think of the assignment of patients to either of the two groups as a natural experiment.

Each of these two groups of patients is then further segmented into strata of comparable risk levels determined using patient-level preoperative risk factors. After this process, the two groups have a comparable stratum of patients with similar levels of risk. For each stratum in the low-occupancy group, we estimate the mean length of stay, which we take to be the length of stay determined by medical variables alone ($LOS_{MED}$). The difference between $LOS_{MED_i}$ and the observed $LOS_i$ is our estimate for the measure $LOS_{EARLY_i}$. That is,

$$LOS_{EARLY_i} = LOS_{MED_i} - LOS_i$$ (11)

The EuroSCORE model is a widely used risk stratification model to estimate medical outcomes of cardiac surgery, including bounce-backs. Typically, the EuroSCORE model only includes medical variables. Our approach is to augment the EuroSCORE model by adding the operational variable $LOS_{EARLY_i}$. This allows us to estimate
the effect of an early discharge on the likelihood of a bounceback:

$$logit(p_i) = \eta \ast LOS_{EARLY_i} + \mu \ast Y_i$$

(12)

where \(p_i\) = probability that patient \(i\) revisits the ICU, and the variables included in \(Y_i\) capture the patient level risk factors that are associated with a bounceback. The coefficient \(\eta\) allows us to estimate of the likelihood of a bounceback as a result of an early discharge by \(LOS_{EARLY_i}\).

5.3 Capacity Implications of Discharge Decisions

We address the capacity implications of discharge decisions in two ways. First, we consider the population of patients who bounced back to the ICU. For this group of patients, we estimate the net capacity savings had we kept them longer in the ICU by \(LOS_{EARLY_i}\) amount of time. Recall that from a pure medical perspective, the appropriate length of stay (\(LOS_{MED_i}\)) is given by:

$$LOS_{MED_i} = LOS_i + LOS_{EARLY_i}$$

For the medically recommended length of stay \(LOS_{MED_i}\), the associated bounce-back probability is \(Pr_{BB,i}(LOS_{MED_i})\). This can be seen as the base-line probability of a bounceback as a result of medical reasons alone. For the actual length of stay, \(LOS_i\) the associated bounce-back probability is \(Pr_{BB,i}(LOS_i)\). The increased bounce-back probability (\(Pr_{BB,i}(LOS_i) - Pr_{BB,i}(LOS_{MED_i})\)) can be attributed to the early discharge.

Define the length of stay following a readmission as \(LOS_{REVISIT_i}\). Then, the hospital faces an expected increase in length of stay of discharging patient \(i\) early of \([Pr_{BB,i}(LOS_i) - Pr_{BB,i}(LOS_{MED_i})] \ast LOS_{REVISIT_i}\). This increase needs to be seen relative to the capacity saving that is obtained from the early discharge. This capacity saving is obtained with certainty and corresponds to \(LOS_{EARLY_i}\) amount of time.
We define $LOS_{SAVED\_ED_i}$ as the net expected bed-days saved from an early discharge of patient $i$. The total length of stay saved across all $i$ patients who bounced back is thus:

$$TOTAL\_LOS\_SAVED = \sum_i LOS\_SAVED\_ED_i$$

$$= \sum_i [LOS\_EARLY_i - \{Pr_{BB,i}(LOS_i) - Pr_{BB,i}(LOS\_MED_i)\} * LOS\_REVISIT_i]$$

Next, we look at the impact of the discharge decision on the peak capacity savings. When the ICU is busy, the early discharge of a patient allows the ICU to admit a “fresh” patient from the OR. Thus, when evaluating the impact of the early discharge on the peak capacity savings, this early discharge counts to our advantage, just as it did in equation (13). However, unlike (13), we now only count the expected length of stay after readmission, if it occurs when the ICU is full. A bounce back that happens to a half empty ICU has no peak capacity implications. We can thus write the peak length of stay saved by an early discharge as for patient $i$ as $PEAK\_LOS\_SAVED\_ED_i$, and across all $i$ early-discharged patients as the summation:

$$TOTAL\_PEAK\_LOS\_SAVED = \sum_i PEAK\_LOS\_SAVED\_ED_i$$

$$\begin{align*}
&= \sum_i [LOS\_EARLY_i * 1_{ICU=BUSY} - \{Pr_{BB,i}(LOS_i) - Pr_{BB,i}(LOS\_MED_i)\} * LOS\_REVISIT_i * 1_{ICU=BUSY}] \\
\end{align*}$$
6 Data Collection

We collected our data at the cardiothoracic intensive care unit at a major east-coast teaching hospital. For each of the 1365 patients in our sample, we compiled data from three different sources, a medical database, a patient tracking system, and the patient billing records.

The medical data that we use comes from the cardiac surgery clinical database from the Society for Thoracic Surgeons (STS). This clinical database provides a set of medical variables that enable us to capture the medical differences across patients. For example, the type of procedure, pre-existing conditions, age, gender and risk factors affect both the recovery time (and hence length of stay in the ICU) as well as the likelihood of developing complications that could lead to bouncebacks. Table 1 provides a list of the control variables related to the system-level covariates, $X_i$, defined previously. Most prominently, this includes the day of the week and the number of procedures performed that day. Table 2 provides a comprehensive listing of the patient level covariates, $Y_i$. These variables are used in the EuroSCORE model and include, for example, age and gender, as well as a set of indicators for potential sources of complications such as previous cardiac surgery, an unstable angina, or a neurological dysfunction. Table 3 provides summary statistics for the control variables $X_i$ and $Y_i$.

Our second source of data comes from the hospital’s patient tracking system. This information system, NaviCare©, is tracks patients and resources such as hospital beds and patient transporters in real time and supports the hospital in its patient flow management. Our research site was one of the first implementation sites of NaviCare in the country, providing us with access to patient flow data beyond what had previously been feasible. NaviCare generates timestamps for a set of events associated with the patient moving through the hospital, including the exact time the patient entered in and departed from the ICU. This information allows us to impute the length of stay for each individual patient. More importantly, we can also
use this data to impute an accurate, hour-by-hour level of occupancy in the ICU. Prior to NaviCare, such micro-level data had not been available, and researchers typically had to rely on less accurate census data to estimate occupancy levels.

Finally, we used the patient billing records to determine the payer type and the insurance status of the patient. In addition to medical variables, one might argue, hospitals discriminate the level of service they offer depending on the insurance status of the patient. 54 of the patients did not have the payer type specified, and we created a dummy payer type called "None" to which we assigned these patients.

We merge these three data sets based on a unique patient identifier creating a comprehensive and consolidated data set consisting of both the operational and medical variables. This yields a total of 1365 unique patient records from June 2006 through June 2007.

This initial set of patients also includes a small set of patients who were admitted with primarily pulmonary conditions, and acutely severe (e.g. heart transplant) outlier patients. Although these patients could not be risk stratified using EuroSCORE, we nevertheless include them in our calculations of the ICU occupancy. However, we only use the remaining 1237 patients in our other hypotheses testing. Table 4 displays detailed summary statistics of the operational variables.

7 Results

We find that the occupancy level in the ICU has a significant impact on an admitted patient’s length of stay. When we estimate equation (10), we find that the coefficient estimate (γ) for the explanatory variable indicating that the ICU is busy (BUSY = 1) is 0.11 (p = 0.07) (Table 5). In other words a patient who enters a busy ICU has a length of stay that is 11% shorter than that of a comparable patient admitted to a low occupancy ICU. The effect of the occupancy on the length of stay is apparent from non-parametric Kaplan-Meier estimates of the aggregate survival functions generated
for busy and non-busy estimates shown in Figure 4. This finding is in concord with the observations of physicians and nursing staff, who indicated to us that when the ICU gets busy, the least severe patients are discharged faster. In other words, the competing alternative hypothesis that surgeries are delayed or cancelled because the ICU is not able to discharge patients faster, is not supported. Instead we find that existing ICU patients are discharged faster in order to accommodate new surgery patients.

We next study the impact of the early discharges on the likelihood that a patient has to revisit the ICU. From our evaluation of the logistic regression equation (12), we estimate the coefficient ($\eta$) value for the explanatory variable measuring early discharge ($LOS\_EARLY$) to be 0.18 ($p = 0.05$) (Table 6). That is, a patient who is discharged early by one day has increased odds of bouncing back by 18%. In other words, patients could have been kept longer in the ICU to reduce the likelihood of a revisit. In the medical literature, we find that the bounceback rate is often taken to be a measure of the quality of care. If one takes this perspective on the quality of care, our finding suggests that an early discharge has a negative impact on the quality of care.\textsuperscript{2} Thus, the alternative hypothesis that patients are discharged only when they are deemed suitable for discharge, is not supported.

We next examine the impact of the increased rate of bounceback on ICU capacity. Our summary statistics (Table 4) show that on average, revisits have a longer length of stay than first-time ICU visits. We find that an average revisit lasts four days while a first-time stay in the ICU lasts a day and a half.

Using the method of propensity scores matching, we estimate the additional bed days allocated to a patient with a given set of risk factors in the absence of a capacity constraint. Figure 5 graphically displays the estimated measure $LOS\_EARLY$ for the patients who bounced back.

\textsuperscript{2}The bounceback rate is considered to be a lower bound on the incidence of complications outside the ICU. For example, there may potentially have been complications developed outside the ICU that were not severe enough to warrant a revisit.
We then evaluate the net capacity the hospital would have saved from keeping the bounceback patients longer the first time. In estimating TOTAL_LOS_SAVED using (13) for all the patients who bounced back, we find that revisits are costly as far as the total bed days are concerned. In particular, had the ICU kept the revisit patients longer the first time, on average 175 fewer bed-days would have been used.

However, when we look at the ICU’s peak capacity as opposed to the total bed days, a different picture emerges. When we estimate (14) for the TOTAL_PEAK_LOS_SAVED, we find a net savings of 75 peak bed days through the practice of aggressive discharges. Thus the early discharge behavior is superior as far as peak capacity is concerned. This is due to the fact that many patients who are discharged early revisit when the ICU is not busy. The net peak-capacity savings of 75 bed-days allows the ICU to accommodate on average 50 more patients who have an average length of stay of 1.52 days. Thus, the practice of rushing and revisiting allows the ICU to increase the availability of beds and treat a larger number of people.

8 Model Validations and Robustness

We first estimated the value of BUSY using exploratory analysis by plotting the risk-adjusted length of stay of patients against the occupancy in the ICU at the time of admission. In Figure 3 we show the the average risk-adjusted length of stay against the occupancy in the ICU. Note that although the ICU has a total of 18 beds, there are patients who have a length of stay of less than a day. As a result, the daily ICU occupancy can exceed 18. From Figure 3, we find that the risk adjusted length of stay declines steadily until the occupancy reaches 16, after which it remains unchanged for higher levels of occupancy. We thus denote occupancy levels exceeding 16 as BUSY = 1. We have tried other values of BUSY and obtained the most statistically significant results with this measure.

Models (10) and (12) were estimated using the method of maximum likelihood.
In estimating (10), we find that the estimate for $\sigma$ is 1.001 (s.e. 0.02), suggesting that the hazard rate of a patient being discharged is time invariant (Table 5).

In our analysis of the effect of an early discharge on the likelihood of a bounce-back, we use all the key explanatory variables identified by physicians and medical researchers in the popularly used EuroSCORE model. Admittedly, there are unobserved underlying patient risks that could potentially bias the estimate of $\eta$. However, such risk factors would not only require a patient to stay longer but also increase the likelihood of an adverse outcome requiring a bounceback to the ICU. Thus, if the estimate for gamma is biased, it is most likely an underestimate of the effect of an early discharge on the likelihood of a bounceback.

9 Conclusion and Future Research

In this paper we looked at the management of bed capacity in a cardiac intensive care unit. We find that the ICU adapts to new demand for beds by aggressively discharging some patients earlier. An adaptive ICU is a healthy ICU because it is able to rapidly respond and reduce the likelihood of becoming a bottleneck in the patient flow process.

However, we also find that an early discharge leads to an increased likelihood of a patient revisit. That is, the practice of aggressively discharging some patients to the step-down unit in order to free up capacity, does lead to some patient revisiting the ICU for a potentially more costly length of stay. However, not all patients end up revisiting as a result of an early discharge. Furthermore, those who end up bouncing back often do so when the ICU is not capacity constrained, and consequently do not consume peak bed capacity. As a result the total throughput is not adversely impact.

In summary, we find that the practice of rushing and revisiting allows the ICU to increase the availability of peak capacity beds, and thus to increase the throughput of cardiac patients. This analysis could be used to study other services such as call
centers and financial services operations, where an inadequate initial resource allocation can lead to further customer revisits and future resource utilization.

References


Green, L. V., V. Nguyen. 2001. Strategies for cutting hospital beds: the impact on
patient service. *Health Services Research*, **36**(2) 421–442


Productivity Press, Portland.


### Table 1: Factors Affecting Length of Stay (X Controls)

<table>
<thead>
<tr>
<th>Measure</th>
<th>Description and Coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weekday</td>
<td>Day of week of procedure. Admissions on Sunday – Wednesday had shorter lengths of stay (and were coded 1) than admissions on other day (which were coded 0).</td>
</tr>
<tr>
<td>Num Procedures</td>
<td>Number of sub-procedures performed during surgery. E.g. single bypass plus a valve repair is coded 2.</td>
</tr>
<tr>
<td>NYHA Classification</td>
<td>New York Heart Association Risk Classification (Ranging from 1 to 4)</td>
</tr>
<tr>
<td>Payer Type</td>
<td>Categorical Variable to denote Medicare, Medicaid, Insurance, Self-Pay or None</td>
</tr>
<tr>
<td>Y</td>
<td>Variables used in EuroSCORE (Table 2 below)</td>
</tr>
<tr>
<td>Measure</td>
<td>Description and Coding</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Age</td>
<td>Following the coding convention used in the EuroSCORE system, patient ages less than 60 were coded 1, while any additional year above 60 accrued an additional point. For example, age for a 75 year old patient would be coded $1 + 75 - 60 = 16$. This coding ascribes a higher weight to much older patients, as they generally carry higher risk.</td>
</tr>
<tr>
<td>Gender</td>
<td>Females are coded 1 and males 0.</td>
</tr>
<tr>
<td>Chronic Pulmonary Disease</td>
<td>Indicates whether the patient is on medication for lung conditions or if the chronic lung disease condition is moderate or severe.</td>
</tr>
<tr>
<td>Extracardiac Arteriopathy</td>
<td>Indicates the presence of vascular disease.</td>
</tr>
<tr>
<td>Neurological Dysfunction</td>
<td>If a cerebrovascular disease exits, this explanatory variable is coded 1. The time of occurrence of the dysfunction, or the type of the cerebrovascular disease is ignored.</td>
</tr>
<tr>
<td>Previous Cardiac Surgery</td>
<td>Indicator to denote if patient has had prior cardiac surgery. The type of cardiac surgery is not considered.</td>
</tr>
<tr>
<td>Serum Creatinine</td>
<td>If the level is higher than 200mmol/l this risk factor is coded 1.</td>
</tr>
<tr>
<td>Active Endocarditis</td>
<td>Indicator to denote that endocarditis is active.</td>
</tr>
<tr>
<td>Critical Preoperative State</td>
<td>This indicator variable denotes the pre-operative state of the patient (critical state or not). The factors that determine whether the patient is in critical state or not are the presence of arrhythmia (irregular heartbeats), cardiogenic shock, need for resuscitation, the need for an intra aortic balloon pump (IABP), or the use of nitrates administered through an I.V.</td>
</tr>
<tr>
<td>Unstable Angina</td>
<td>Indicates syndrome that is intermediate between stable angina and a myocardial infarction.</td>
</tr>
<tr>
<td>Left Ventricular Dysfunction (&lt;30% Ejection fraction)</td>
<td>Indicates whether ejection fraction is less than 30%</td>
</tr>
<tr>
<td>Left Ventricular Dysfunction (30% &lt; Ejection fraction &lt; 50%)</td>
<td>Indicates whether ejection fraction is between 30% and 50%</td>
</tr>
<tr>
<td>Recent Myocardial Infarction</td>
<td>Indicates whether myocardial infarction occurred in the last 90 days.</td>
</tr>
<tr>
<td>Pulmonary Hypertension</td>
<td>Indicates that the systolic pulmonary pressure exceed 60 mmHg.</td>
</tr>
<tr>
<td>Emergency</td>
<td>Indicates status of admission</td>
</tr>
<tr>
<td>Other than isolated CABG</td>
<td>This is coded 1 if in addition to a CABG, another heart procedure was performed.</td>
</tr>
<tr>
<td>Surgery on Thoracic Aorta</td>
<td>Indicator for the presence of Aortic Aneurysm.</td>
</tr>
<tr>
<td>Post-infarction Septal Rupture</td>
<td>Indicates whether ventricular septum ruptured following a heart attack.</td>
</tr>
</tbody>
</table>

Table 2: EuroSCORE Controls (Variables in Y)
<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weekday</td>
<td>0.59</td>
</tr>
<tr>
<td>Num Procedures</td>
<td>1.87</td>
</tr>
<tr>
<td>NYHA Classification</td>
<td>2.2</td>
</tr>
<tr>
<td>Age</td>
<td>62.3</td>
</tr>
<tr>
<td>Gender / Females</td>
<td>0.34</td>
</tr>
<tr>
<td>Chronic Pulmonary Disease</td>
<td>0.144</td>
</tr>
<tr>
<td>Extracardiac Arteriopathy</td>
<td>0.147</td>
</tr>
<tr>
<td>Neurological Dysfunction Disease</td>
<td>0.17</td>
</tr>
<tr>
<td>Previous Cardiac Surgery</td>
<td>0.295</td>
</tr>
<tr>
<td>Serum Creatinine &gt; 200 mmol/L</td>
<td>0.11</td>
</tr>
<tr>
<td>Active Endocarditis</td>
<td>0.06</td>
</tr>
<tr>
<td>Critical Preoperative state</td>
<td>0.327</td>
</tr>
<tr>
<td>Unstable Angina</td>
<td>0.172</td>
</tr>
<tr>
<td>LV Dysfunction 30 (E.F &lt; 30%)</td>
<td>0.22</td>
</tr>
<tr>
<td>LV Dysfunction 50 (30% &lt; E.F. 50%)</td>
<td>0.35</td>
</tr>
<tr>
<td>Recent Myocardial Infarction</td>
<td>0.1</td>
</tr>
<tr>
<td>Pulmonary Hypertension</td>
<td>0.0037</td>
</tr>
<tr>
<td>Emergency</td>
<td>0.3</td>
</tr>
<tr>
<td>Other Than Isolated CABG</td>
<td>0.17</td>
</tr>
<tr>
<td>Surgery on Thoracic Aorta</td>
<td>0.16</td>
</tr>
<tr>
<td>Postinfarction Septal rupture</td>
<td>0.005</td>
</tr>
</tbody>
</table>

Table 3: Controls Summary Statistics (N = 1237)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOS</td>
<td>1.52</td>
<td>1.2</td>
<td>1.1</td>
</tr>
<tr>
<td>LOS_REVISIT</td>
<td>4.1</td>
<td>4.6</td>
<td>2.8</td>
</tr>
<tr>
<td>OCCUPANCY</td>
<td>16.5</td>
<td>3.5</td>
<td>17</td>
</tr>
<tr>
<td>BUSY</td>
<td>0.54</td>
<td>0.49</td>
<td>1</td>
</tr>
<tr>
<td>BB</td>
<td>0.12</td>
<td>0.33</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4: Model Variable Summary Statistics (N=1365)
### Table 5: Effect of Occupancy on Length of Stay

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Estimate (SE)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.28 (0.15)</td>
<td>0.03</td>
</tr>
<tr>
<td>BUSY</td>
<td>-0.11 (0.06)</td>
<td>0.07</td>
</tr>
<tr>
<td>σ</td>
<td>1.001 (0.02)</td>
<td>0.001</td>
</tr>
</tbody>
</table>

*N=1237. Log-likelihood of Model Fit = -1709.49

Dummy Variable Estimates for $X_i$ (Appendix) not Displayed

### Table 6: Effect of Early Discharge on Likelihood of Bounceback

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Estimate (SE)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.28 (0.15)</td>
<td>0.03</td>
</tr>
<tr>
<td>LOS_EARLY</td>
<td>-0.11 (0.06)</td>
<td>0.07</td>
</tr>
</tbody>
</table>

*N=1237. Model Likelihood Ratio Goodness of Fit p < 0.001

Dummy Variable Estimates for $Y_i$ (Appendix) not Displayed

### Figures

Figure 2: Propensity Score Matching

Control = Low Occupancy

Treatment = High Occupancy
Figure 3: Estimating Occupancy Level for $BUSY = 1$

Figure 4: Effect of Occupancy on Length of Stay
Figure 5: Early Discharge in Hours