



An Empirical Study of the Spillover Effects of Workload on Patient Length of Stay

**Jillian Berry Jaeker
Anita Tucker**

Working Paper

13-052

July 17, 2013

Copyright © 2012, 2013 by Jillian Berry Jaeker and Anita Tucker

Working papers are in draft form. This working paper is distributed for purposes of comment and discussion only. It may not be reproduced without permission of the copyright holder. Copies of working papers are available from the author.

An Empirical Study of the Spillover Effects of Workload on Patient Length of Stay

Jillian Berry Jaeker, Harvard Business School

Anita Tucker, Harvard Business School

Abstract

We use two years of inpatient data from 243 California hospitals to quantify the relationship between hospital-level workload and patient length of stay (LOS), and its “spillover” effects across patient types. Patients are categorized as medical or surgical, and the effects of same type patient workload (occupancy) on LOS are analyzed. The analysis is repeated with workload replaced by other type patient occupancy, providing a “spillover” effect. We find that the effects of inpatient workload on LOS spillover across patient types, which we theorize results from most inpatients, regardless of type, utilizing the same shared resources (e.g. pharmacy and laboratory). These spillover effects remain even while we find that the effects of workload vary at different time points during a patient’s stay: LOS increases as inpatient workload on the day of admission increases, while inpatient workload at the end of the stay has a U-shaped effect on LOS.

1. Introduction and Patient Care Process

Healthcare spending in the US has increased from \$256 billion (9.2% of GDP) in 1980 to \$2.7 trillion in 2011 (17.9%), with hospital care accounting for nearly a third of these expenditures (Centers for Medicare & Medicaid Services Office of the Actuary National Health Statistics Group 2012). As a result, there is increasing pressure to identify mechanisms for reducing costs while maintaining quality of care, such as through the Affordable Care Act (Sebelius 2013). One important area of focus is determining drivers of productivity in acute care hospitals. Productivity can be thought of as the number of patients successfully cared for in a time period divided by the resources used during that time. Given the high fixed costs of hospitals, attempts to increase productivity have primarily focused on increasing bed occupancy levels (Roberts et al. 1999). These occupancy levels are also referred to as “patient load”, or more generally, workload.

Operations management research has shown, however, that the relationship between productivity and occupancy levels *within an inpatient unit* is ambiguous. High occupancy can lead to either increased or decreased length of stay (LOS). First, queuing theory states that under conditions of variation in service and arrival times, both of which occur in hospitals (Eddy 1984), queuing will occur. As utilization approaches 100%, queue length and overall throughput time increase dramatically. In the healthcare setting, studies have shown that as patient load become higher, LOS increases (Green and Nguyen 2001) and quality of care decreases due to the resulting delays in treatment (Chalfin et al. 2007) and mental

strain (Kuntz et al. 2013). We refer to delays in treatment due to a combination of queuing for resources and slowing down due to mental strain as “congestion effects.”

Second, employee behaviors, such as those that emerge when employees have high levels of discretion over how they complete their tasks (Hopp et al., 2007), alter predicted relationships among capacity utilization and throughput times (Debo et al. 2008, Hopp et al. 2009, Hopp et al. 2007, Jouini et al. 2008). In discretionary task settings, employees have the ability to decide what tasks to perform for their customers, in what order, and what level of quality is acceptable (Hopp et al. 2007). Thus workers have the levers to increase or decrease processing times for customers, which in turn impacts the link between workload and overall throughput time. For example, in a study of restaurant employees, Tan and Netessine (2012) found that when a restaurant was busy, employees modified the level of service they provided to customers such that workers spent less time per customer, and customers spent less time in the restaurant. We refer to this behavior as “workload smoothing.”

Operations management scholars have highlighted the need for empirical research at the intersection between human behavior and operations management principles (Boudreau et al. 2003, Hopp et al. 2009, Jouini et al. 2008). In particular, more research is needed to shed insight on the complex relationship between employee response to workload and its impact on productivity, such as throughput times (Boudreau et al. 2003, Hopp et al. 2009, Hopp et al. 2007, Schultz et al. 2003), particularly in the healthcare setting (KC and Terwiesch 2009, KC and Terwiesch 2012). Failing to properly account for the impact of patient load on productivity can result in the misalignment of labor (Green et al. 2011) and physical resources (Green and Nguyen 2001, Shapiro 1996), and result in poorer performance (Debo et al. 2008, Kuntz et al. 2013, Tan and Netessine 2012). Scholars have begun to analyze the effect of patient load on LOS, and have proposed target occupancy levels that maximize productivity, but they have only examined the effect of patient load on patients of the same service type (Chan et al. 2010). However, since hospitals are systems of interconnected, semi-autonomous inpatient units (Gittell et al. 2000), the optimal allocation of resources to maximize productivity for the hospital, accounting for workload related factors, may be different from the optimal target occupancy and resource allocation when only considering the impact of workload in one inpatient unit.

This paper addresses this call for research by empirically examining the impact of patient load on LOS, with a special focus on the “spillover” effect of patient load across the two main types of patients within a hospital: medical and surgical. Based on his or her diagnosis, a patient is assigned to a specific hospital service, which is the specialty responsible for caring for the patient. Each service has its own physicians, and while there may be some overlap in beds and nursing, the care is generally segregated by service (Gittell et al. 2000, Meyer 2011). Ancillary services, such as radiology, pharmacy, laboratory, and transport, are shared among all patients in the hospital. We refer to the occupancy of medical (surgical)

patients in the hospital as medical (surgical) inpatient workload, while we refer to the occupancy of medical (surgical) patients awaiting admission as incoming medical (surgical) patient workload (See Figure 1).

FIGURE 1 ABOUT HERE

We use survival analysis with two years of patient-level data from 243 California acute care hospitals to empirically test the effects of load on LOS. We analyze 239,400 patients across eight different diagnostic groups (DRGs), including both medical and surgical diagnoses. Upon admission to the hospital each patient is assigned one of 746 diagnostic related group (DRG) based on the clinical diagnosis and primary needs of the patient, including the patient's severity level (Office of Inspector General 2001). While there is some variation among patients of a given DRG, CMS treats it as minimal, and these DRGs are used to calculate a payment level from CMS, where CMS reimburses for the treatment of the condition, irrespective of the costs incurred (Office of Inspector General 2001). In addition, we supplement our quantitative analysis with interviews of nurses from a California hospital.

We use three measures for patient load, all of which are calculated at the level of an individual patient: inpatient workload on the focal patient's day of admission; inpatient workload on the day before expected discharge for the focal patient; and incoming patient workload on the day before expected discharge for the focal patient. By analyzing inpatient load at two different time points during the stay, we can test whether different effects dominate at different times. For newly admitted patients, we find that as the hospital occupancy level of patients of the same type increases, LOS increases, which is consistent with congestion effects. We suspect that this occurs because key tasks are not being performed for the newly admitted patients in the "golden" first hours when they are in the hospital (Blow et al. 1999, Buist et al. 2002), which results in the patient's stay being longer overall. More interesting is that these effects also spillover from one type of patient to another (e.g., a high volume of medical patients increases LOS for newly admitted surgical patients) such that as the inpatient load of the other type increases, the LOS of the newly admitted patient increases. These results are consistent with increased queuing due to patient load, which delays the start of care not just for own type patients, but also for other type patients.

Meanwhile, on the day before a patient's expected discharge, we find that an increase in the inpatient workload of a patient's own type is associated with a U-shaped effect on LOS. Specifically, as same type inpatient workload increases towards around 85% occupancy, the probability of discharge increases by about 7%, so that expected LOS becomes shorter. This is indicative of a workload smoothing effect where employees prioritize inpatients who are less far along in their hospital stay by sending home patients who are approaching the end of their hospital stay. However, once workload exceeds approximately 85% occupancy, the trend reverses, and the probability of discharge decreases, thus increasing LOS. This suggests that there is a tipping point where congestion effects dominate the effects of workload

smoothing. We find the same U-shaped relationship between LOS and inpatient workload from the other type of patients. We observe a *decrease* in LOS as workload from the different patient type increases at lower occupancy levels, followed by an increase in LOS as the occupancy of the different patient type reaches a high level. We find it surprising that—at the lower levels of occupancy of dissimilar patients—increasing occupancy from dissimilar patients is associated with decreasing LOSs, as this suggests that the behavioral response of workload smoothing spills over across patient types. Specifically, we theorize that employees perceive the strain on the hospital’s shared resources, such as pharmacy and radiology, from the other type of patients. This strain can cause delays in the care of their own patients and therefore similar to their response to workload from the same type of patients, they prioritize the hospital’s sickest patients by discharging those closest to leaving to reduce the hospital system level load. As with the same type of patients, workload smoothing requires spare capacity to undertake discharge procedures, so that as the hospital occupancy of dissimilar patients reaches high levels, congestion effects dominate and LOS increases, which we believe is due to delays obtaining shared resources.

Finally, we find that as incoming patient load of the same type increases, the probability of discharge on the day before expected discharge also increases, consistent with workload smoothing. However, the incoming patient load from the different type patients does not impact the probability of discharge on the day before expected discharge. We hypothesize that the reason that there is no spillover effect from incoming patients is because they have not yet begun to use resources, but employees caring for same type patients know that they face a heavy future workload, and are incentivized to smooth their load by discharging patients near the end of their stays.

Our work makes several contributions to the growing operations management literature on the impact of employee behaviors and workload on organizational productivity. First, we identify and quantify spillover effects of workload across different types of patients on LOS, which suggests that studies should include possible spillover effects if they are to fully account for workload’s impact on productivity. Second, we analyze the effects of inpatient workload at the beginning and end of patients’ hospital stays, and find that different effects dominate at different times, a concept that we believe has not been fully explored in previous work. Third, our data enable us to separate the effect of inpatient workload from that of incoming workload, a distinction that, to our knowledge, has been understudied in prior research. Finally, by distinguishing between incoming and inpatient workloads, as well as the timing of the loads, we show that there is a congestion spillover effect, which would be expected in situations that involve resource sharing, but also a workload smoothing spillover effect, which is a surprising behavioral response not theorized to occur with shared resources. Our study has important implications for how managers should think about allocating resources across the hospital to achieve optimal hospital-level performance.

2. Related Literature

Traditional operations models have generally made the assumptions that the time to complete the work on a particular “unit of work” is a random variable with a constant mean, and that each unit of work is independent from each other unit (Dallery and Gershwin 1992). These assumptions have been extended to research in patient care settings. However, recent research suggests that the time required to complete a unit of work is not constant, but is instead influenced by the overall workload, though the direction of the change may not be constant. One reason for this is because workers can speed up or slow down if they have discretion over two key behaviors: how many tasks they perform for customers (Batt et al. 2012, Hopp et al. 2007, Oliva and Sterman 2001) and—even if they have no discretion over which tasks to perform—how long they take to perform a standard set of tasks (KC and Terwiesch 2009, Schultz et al. 1998). For example, in service settings, a worker can respond to her own workload—as implied by the number of people waiting for service from her—by either increasing or decreasing the amount of work she performs per customer, which impacts quality and the average processing time. Hopp, Iravani et al. (2007) analytically showed that by increasing the capacity in a service setting, effectively reducing the workload per worker, servers may provide more service to each customer. Under certain conditions, this can paradoxically result in longer queues for customers, despite the increased capacity, reinforcing the importance of understanding how the state of the system affects resource use and thus, allocation decisions.

Other analytical studies have found that when customers are queued up waiting for service, the average utility of all of the customers can be increased by decreasing the time spent per customer (George and Harrison 2001, Ha 1998, Stidham and Weber 1989). This is because the cost of increased effort and/or reduced quality is offset by the reduction in waiting times (George and Harrison 2001, Stidham and Weber 1989). Empirical studies of employees’ response to workload have found a general tendency for employees to reduce the time they spend per “unit of work” when they perceive an increasing inflow of work. Research has found that as workload increases employees will first work faster (Kuntz et al. 2013), and if queuing continues, will then “cut corners” by omitting certain tasks to reduce waiting time for queuing customers (Oliva and Sterman 2001). In a healthcare setting, high patient load has been shown to result in premature discharge from the ICU (KC and Terwiesch 2012), as well as incomplete physician documentation (Powell et al. 2012). The corner cutting can negatively impact the quality of work, as evidenced by higher mortality rates (Kuntz et al. 2013) and lower reimbursement (Powell et al. 2012). Employees whose number of tasks is fixed, and whose work is standardized may not be able to cut corners, but have been shown to speed up when faced with rising inventory levels (Schultz et al. 1999, Schultz et al. 1998).

The above research suggests that when employees face an increasing workload, they respond by reducing processing time by speeding up and cutting corners. However, although a worker's psychological response to increased workload is to try to reduce LOS, congestion effects work in the opposite direction of this workload smoothing behavior. Even with the speeding up efforts of employees, if the rate of incoming demand is greater than the rate at which the system can process that demand, the queue will lengthen and average throughput time will increase (Little 1961). Thus, even though a worker may speed up to reduce a particular patient's LOS, if the change in service rate is not as high as the change in workload, then on average, high workload will result in an increase in LOS.

In addition, heavy workload can cause mental strain and exhaustion, which can also decrease service rates, thus increasing processing times. Tan and Netessine's (2012) empirical study of restaurant servers indicates that as demand increases, servers can increase effort and serve diners more quickly, while maintaining or increasing revenue, as follows from the expected optimal behavior. However, during periods of very high demand, congestion effects may govern changes in processing times, with diners staying longer, but without an increase in revenue, which is due to the inability of servers to effectively serve so many customers. In addition, increased load in a healthcare setting has been shown to increase the number of interruptions, which often results in repeat set-up periods and re-work due to errors (Tucker and Spear 2006). Prolonged heavy load in healthcare has also been linked to worker exhaustion that results in slowdown and a propensity for errors (KC and Terwiesch 2009). As a result, slow-down and longer LOSs may occur when workload is heavy.

To reconcile the above observed tension in the direction of change in LOS due to increased workload, we examine whether workload has a different effect on LOS at different times during a patient's stay, and at different occupancy levels. We first test whether, and if so, when workload increases lead to congestion effects (increased LOS) versus workload smoothing (decreased LOS). By separating the effects of congestion and workload smoothing, we are able to show that when hospitals become busy workers prioritize some types of patients by systematically shortening the LOS of other types of patients. This highlights the importance of teasing apart the "average" effect of workload to determine its specific impact on individual patient types. We further build on this research by analyzing how workload from a different type of patient affects LOS, and if this effect matches the effect from workload from the same type of patient. In the next section we describe our hypotheses on how patients of different types and at different times during their stays are affected by workload.

3. Hypotheses

3.1 Inpatient Workload

To test the spillover effects of inpatient workload on LOS of a different type of patient, we need to understand the effects related to same type patients and then compare these with the effects due to the workload of other type patients. The greatest costs associated with hospital care are for equipment, building, and labor, which are fixed in the short term (Roberts et al. 1999). Consequently, many hospitals try to leverage these costs with high utilization of their assets, usually targeting an 85% occupancy rate (Green and Nguyen 2001), leading to times where there is high system-level inpatient workload. However, in general, the start of a patient’s stay requires the most staff and physical resources (Clarke 1996), while the remainder of the stay is more focused on recovery. During this initial period, there is a set of time consuming tasks that must be completed. Based on our discussions and observations, these include going through the patient’s history, orders, allergies; getting vital signs; and—if the patient’s condition is straightforward—beginning treatment. If the patient’s diagnosis is not definitive, further tests may need to be completed before treatment can begin. During periods of high inpatient load, it is more likely that there will be queuing delays to complete these tests. Completing these admission tasks also requires substantial nursing time. Therefore, during times of high inpatient workload staff may first try to complete important work for their other patients, delaying the start of the treatment process for newly admitted patients.

Furthermore, the benefits of prioritizing the care of newly admitted patients will not be apparent until those patients are ready for discharge days later, so the immediate costs of focusing on newly admitted patients far outweighs the immediate benefits. Meanwhile, delays in care due to rework and switching back and forth between multiple patients occur when healthcare workers care for multiple patients (Tucker and Spear 2006). Thus, it is possible that during times of high workload, the additional patient load inherent with new inpatients will result in a longer LOS for them. Delaying the start of care for critical patients has been shown to increase the overall LOS for a patient, beyond the direct delay in care (Chalfin et al. 2007). Therefore we predict that workers will prioritize patients near completion to reduce their total workloads due to those patients being discharged, and thus delaying the care for the newly admitted patients. Consequently, as the inpatient workload of same type patients increases on the day of admission of a focal patient, we would expect congestion effects to dominate over workload smoothing effects for the focal patient and therefore the focal patient will end up with a longer LOS than otherwise would be expected. Thus,

$$\frac{\partial LOS_i}{\partial Occupancy_{i,0}} > 0 \tag{1a}$$

where LOS_i is the LOS of a patient of type i , (medical or surgical), and $Occupancy_{i,0}$ is the occupancy of patients of patient type i on day 0 . Formally,

Hypothesis 1a: Increased inpatient workload of same type patients on the day of admission is associated with an increased LOS.

During the initial period of the stay, patients frequently need diagnostic tests, medications, and any treatments begun. All of these tasks require resources, most of which are shared across the hospital, such as radiology, pharmacy, and transport services. Therefore, even if only other type patients are facing high inpatient loads, there will be queuing for resources, which will delay the care of a newly admitted other type patient, and increase this patient's LOS. Thus

$$\frac{\partial LOS_i}{\partial Occupancy_{j,0}} > 0 \quad (1b)$$

where $Occupancy_{j,0}$ is the occupancy of patients, j , of type j on day 0 .

Hypothesis 1b: Increased inpatient workload of other type patients on the day of admission is associated with an increased LOS.

Each patient has an expected LOS associated with his primary condition, as well as patient specific characteristics, such as age, gender, and presence of comorbidities. The decision to discharge a specific patient requires balancing the effects of sending the patient home before he is fully recovered, which increases the chance of readmission (Chan et al. 2010, KC and Terwiesch 2012) versus keeping him longer, which enables a more thorough recovery, but also incurs greater cost for the hospital and exposes the patient to the risks of hospital acquired infections or other medical errors (Hauck and Zhao 2011). We hypothesize that when inpatient load of same type patients is high around the time of discharge, the decision to discharge will be affected by more than medical concerns alone. It should be noted that if inpatient load is going to directly affect discharge timing, we expect it to have the greatest impact on the day before expected discharge because most discharge preparations begin on the day before expected discharge. If there is a slowdown due to congestion effects on the day before discharge, discharge preparations may be postponed, resulting in that patient's discharge being delayed beyond the expected day of discharge. Conversely, if there is a speed-up due to workload smoothing, all the discharge preparations may be completed early, resulting in that patient being discharged on the day before expected discharge.

If there is high inpatient workload, workers will desire to reduce their own loads (Green et al. 2011, Tan and Netessine 2012) so that they can increase the quality of care for their remaining patients. Nurses have been shown to value short-term problem-solving skills (Tucker and Edmondson 2003), and in the

case of high inpatient load, they have to prioritize the patients who need their care the most, while still caring for patients who need more maintenance, and balancing these needs with their mental and physical limits. Thus, they choose to discharge patients close to completion to have more time to focus on the sickest patients, though in doing so they delay care to the newly admitted patients while discharges are being processed. Furthermore, in contrast to the day of admission, any additional effort to speed up will more directly affect workers as it can result in a discharge, and thus reduction in workload during a shift. Reduction of a patient also eliminates the need to “hand-off” the patient at the end of the shift, another time consuming task we observed that adds to the load of nurses and physicians. Additionally, workers have more discretion over when tasks can be completed to ensure discharge. For example, the discharge paperwork can be nearly completed hours before the final discharge occurs. Finally, some tasks may even be able to be shifted to outpatient follow-up care. Therefore as inpatient workload of same type patients increases, we would expect the costs of workload smoothing to be less than the benefits for the workers so that the LOS for a patient near discharge is reduced.

However, the ability of a worker to speed-up assumes some spare capacity to work faster and/or rearrange tasks to ensure all tasks are completed for an early discharge. At very high occupancy levels, workers may only be able to maintain patients (Aiken et al. 2002) and may be too strained to speed up in order to smooth their workloads, and instead will be more likely to slow down (KC and Terwiesch 2009, Tan and Netessine 2012). Consequently, we expect that at very high levels of occupancy, congestion effects will dominate workload smoothing effects, and the LOS for a patient near discharge will increase. For that reason, we expect a U-shaped effect of inpatient workload of same type patients on the day before expected discharge on the LOS of a current patient, such that LOS will initially decrease as workload increases from low levels of workload, but LOS eventually increases at high levels of workload.

$$\frac{\partial LOS_i}{\partial Occupancy_{i,d-1}} < 0 \quad (2a)$$

$$\frac{\partial LOS_i}{\partial (Occupancy_{i,d-1})^2} > 0$$

where $Occupancy_{i,d-1}$ is the occupancy of patients of type i on the day before expected discharge, $d-1$.

Hypothesis 2a: Increase of inpatient workload of same type patients on the day before expected discharge is associated with a U-shaped LOS

As with high inpatient workload on the day of admission, we would expect that the queuing associated with shared hospital resources will make caring for current patients more difficult for healthcare personnel. Since these workers still want to prioritize the patients that need the most care, and

smooth out their own workloads, they will discharge the patients nearest completion to alleviate some of the strain on the system. As with the predicted effect for same type patients, this workload smoothing can only occur if there is additional capacity to handle these discharges. Thus, we expect the following

$$\frac{\partial LOS_i}{\partial Occupancy_{j,d-1}} < 0 \quad (2b)$$

$$\frac{\partial LOS_i}{\partial (Occupancy_{j,d-1})^2} > 0$$

Hypothesis 2b: Increase of workload of other type patients on the day before expected discharge is associated with a U-shaped LOS

3.2 Incoming Patients

In addition to the number of inpatients in a hospital, there are also incoming patients. As described above, newly admitted patients are very time consuming, and the initial care is considered some of the most important for good outcomes. Independent of the inpatient workload within the hospital, as the number of waiting to be admitted patients increases, the expected future load of the workers greatly increases. Since workers try to minimize their own loads so as to reduce the chance of poor quality of care (Green et al. 2011), we expect that workers prioritize newer patients, where the effects of proper care are greater. Moreover, workers do not want to be perceived as bottlenecks in the system (Schultz et al. 1999, Schultz et al. 1998), so they will ensure they have space for an incoming patient. However, if current patients remain in the hospital, even if they require fewer resources, they still increase mental strain. As a result, as the number of incoming patients increases, we predict that workers will focus instead on discharging patients nearing their day-of-discharge earlier than they otherwise would. This strategy enables workers to focus more fully on incoming patients. We expect

$$\frac{\partial LOS_i}{\partial Incoming_{i,d-1}} < 0 \quad (3)$$

where $Incoming_{i,d-1}$ is the relative number of incoming patients of type i on the day before expected discharge, $d-1$. Note that on the day of expected discharge, high numbers of incoming patients may cause some workload smoothing, but since the patients are already expected to go home we do not expect to observe as strong an effect. Thus, we believe that the largest effect from heavy loads of incoming patients will be observed on the day before expected discharge. Nevertheless, for completeness in our model will include the effect of incoming patients on LOS on the day of expected discharge.

Hypothesis 3: Increased same type incoming patient load on the day before expected discharge is associated with a decreased LOS for current patients.

Unlike inpatient load, the effects of incoming patients are confined to same type patients as they only put pressure on workers in these same type inpatient units since they represent future work, but are not currently consuming resources.

Combining our hypotheses, we model the LOS of patient, i , in hospital, h , as follows

$$LOS_{i,h} = \beta_1 Controls_i + \beta_2 Controls_h + \beta_3 Occupancy_{i,0} + \beta_3 Occupancy_{i,d-1} + \beta_3 (Occupancy_{i,d-1})^2 + \beta_3 Incoming_{i,d-1} + \beta_4 Incoming_{i,d} + \varepsilon \quad (4)$$

Where $Controls_i$ are patient level controls (e.g. age, sex, date of birth, medical condition), and $Controls_h$ are hospital level controls (e.g. hospital of treatment, day of week of service), and ε is an error term representing the stochastic nature of treatment response. Within the hospital controls is the number of medical and surgical admissions on the day of admission as well as scheduled admissions on the day before and day of expected discharge to control for any cancellation of surgeries due to load, which could be a buffering mechanism (For a complete list of controls see Table 1b). The patient and hospital controls provide a baseline amount of time that a patient must be in the hospital to recover for a given medical condition with a given set of patient characteristics, and the remaining variables are the workload variables of interest (Table 1c).

4. Data

Our data consists of patient level records for all inpatient discharges in the state of California from December 2007 through December 2009 (Office of Statewide Health Planning and Development). We sum all patient admissions and subtract all patient discharges for all patients admitted in December, 2007 to determine the baseline number of patients in each hospital on January 1, 2008. We restrict our sample to only include patients admitted after December 31, 2007, and before November 30, 2009 since our data only has patients who were discharged by December 31, 2009, and does not include any patients admitted in December, but discharged on or after January 1, 2010, thus making it impossible for us to get an accurate census during that month. We limit our data to acute care hospitals with at least 7000 patients over the course of our study (23 months) to ensure enough patients per day to calculate reasonable occupancy levels, and a 24 hour ED to ensure there are emergency admissions, resulting in a sample of 243 hospitals. Each admission is its own record and includes the date admitted, date discharged, demographic information on the patient, hospital of care, diagnoses, major procedures, disposition (i.e., discharged home, to home health services, etc.), if an admission was scheduled or not, and DRG, among others. Consequently, we know whether the patient was surgical or medical, and if the visit was scheduled

or emergent. Lastly, we have dates for all major procedures performed, and the LOS of each patient in days.

The hospitals in our data vary greatly in the number of admissions and number of beds, so the absolute level of inpatient workload and incoming patient load can similarly vary. To be able to analyze the effects of patient workload across these heterogeneous hospitals, we have converted the absolute number of surgical and medical patients currently in the hospital, as well as the number of incoming surgical and medical patients, for each date, into a percentage of the maximum for each patient workload category. While we would prefer to calculate occupancy based on staffed beds available for a given day, the recorded numbers of staffed beds are at the year level, and often quite different from reality. Consequently, in a manner similar to that employed by Kuntz et al. (2013) we divide the number of visits by the maximum number of daily patient visits, and refer to this value as occupancy. Our methodology differs slightly from Kuntz et al. (2013) in that we calculate our maximum number of visits taking into account the effects on staffing of seasonality, trends, and the weekend effect, whereby hospitals reduce staffing levels on weekends. Specifically, we calculate the 99th percentile of the number of patients treated or admitted on weekdays for each quarter-year combination, and then do the same for weekends for each quarter-year combination, and these numbers represent our maximum weekday and weekend censuses, respectively. We use the 99th percentile to account for any extraordinary circumstances that do not actually reflect the number of beds generally available for occupancy. We distinguish between weekdays and weekends to account for what is commonly termed the “weekend effect” in hospitals, and we distinguish the quarters of the years to account for seasonality in demand. For example, to calculate the congestion for hospital “H” on Monday, June 9, 2008, we first find the 99th percentile of the number of patients who were inpatient on a single weekday in hospital H in quarter 2 of 2008, and label that the maximum. We then divide the number of patients actually in the hospital on June 9 by that maximum.

We focus our analysis on eight DRGs: five medical and three surgical for a total of 239,400 patient visits (see Tables 1a, 1b, and 1c for the list of DRGs used as well as summary statistics). We restrict our analysis to these DRGs since they are among the 20 most common DRGs representing 3.7% of total patient visits. They also all have average LOSs (as defined by CMS) of at least 3 days, but less than 7 days. This average LOS range ensures that the LOS is long enough to detect day level changes in LOS, but not so long that it becomes difficult to predict the expected LOS patients, reducing noise in our sample. By analyzing eight DRGs, we ensure that our results are robust across patient types. In the DRG classification, there are multiple DRG numbers for the same primary complaint, but differ in severity levels. For example, 469 is Hip or Knee Replacement with major complications and/or comorbidities, while 470 is Hip or Knee Replacement without major complications or comorbidities. Each of the eight

DRGs in our study represents a specific primary reason for visit and intensity level selected because its frequency of occurrence and average LOS met the above described criteria for inclusion.

To further reduce noise from our analysis of the impact of workload on the focal patients' LOS, we exclude patients who die during their stay since it is very difficult to predict the LOS of a patient who ends up dying in the hospital (Kuntz et al. 2013), and the death of a patient should be uncorrelated with the LOS of the remaining patients outside of the effect of occupancy. Furthermore, in our sample, only 0.50% died within the hospital, so we do not think they will affect our analysis. We also exclude patients who are transferred to another hospital, as these patients have LOSs that are impacted by factors that are not directly related to the occupancy of the hospital, as there is little evidence that transfers take place due to hospital occupancy (Kuntz et al. 2013) and represent less than 2% of the total patients within these eight DRGs. Finally, we exclude patients who left against medical advice as it is difficult to predict when they will leave, and they are only 1.10% of the sample.

Note that while we focus on eight DRGs for calculating effects of workload on LOS, the inpatient and incoming patient volumes include all DRGs, except transplants, burns, and obstetrics, because the incoming work and current occupancy levels in the hospital are comprised of all patient types. We do not include transplant, burn, and obstetric DRGs as these patients are often treated in separate areas of the hospital by very specialized caregivers.

In addition to patient level data, we also interviewed five medical/surgical nurses from a California hospital about what they did to prepare for a new admission. This was part of a larger research project that one of the authors was conducting with the hospital to understand what equipment was required to be ready for a newly admitted patient to the inpatient unit.

5. Econometric Specification

The LOS of a patient can be thought of as a survival function, with discharge being equivalent to exiting the system (KC and Terwiesch 2012). Survival analysis allows us to predict a patient's likelihood of discharge on any given day based on an underlying hazard function scaled by the patient's and hospital's characteristics. Since our LOS data is at the day level, we must use a discrete-time survival analysis. In the medical literature it is common to allow the baseline hazard to vary with time (KC and Terwiesch 2012), and since our data does not fit a known distribution very well, we include a variable for each day to account for this flexible hazard rate.

We follow Jenkins' (2004) approach and use a proportional hazard complementary log-log (cloglog) regression to model the probability of "failure" (i.e., discharge) at any given time (in our case, for each day). Cloglog is an appropriate model since it can be used with discrete and censored survival times, and it is the best estimator when observation times are discrete but the events occur continuously; in our study

we have daily census, admission, and discharge data, but patients are admitted and discharged throughout the day. This gives us the following hazard for patient i at time t :

$$h_i(t) = 1 - \exp[-\gamma_i \exp(h_0(t))] \quad (5)$$

Where $h_0(t)$ is the baseline hazard on day t and

$$\gamma_i = \exp(\beta' X_i) \quad (6)$$

in which β is a vector of coefficients, and X is a vector of variables. We restrict our analysis of LOS to patients with the eight DRGs of interest, running separate analyses for each DRG given that each will have a different baseline hazard function. We calculate this hazard rate for each patient for each day, $d=1,2,\dots,D$, where D is the 99th percentile of LOS for the DRG, providing the baseline hazard function associated with time. We consider the 99th percentile of LOS to be the maximum expected LOS since we have observed that patients who have abnormally long LOS often have unobservable medical or social conditions unrelated to patient census that is causing the long LOS (e.g., they do not have a safe home environment). To account for this cutoff, we censor patients with a LOS greater than D days, which can be interpreted as having had no event before the end of the observation period.

In addition to the controls listed in Table 1b, we include our variables of interest: medical and surgical occupancy, as well as their square terms, on the day of admission and the day before expected discharge; medical and surgical admissions on the day before and of expected discharge (See Table 1c). Given that we only have midnight occupancy levels, we use the occupancy level at the start of the day of admission (the midnight immediately prior to being admitted) for the day of admission occupancy. For the day before expected discharge, we use the occupancy at the end of the day before expected discharge. The admission variables are the total number of admissions over the course of the day of interest.

Since patients have different expected LOSs due to individual patient characteristics, even within a single DRG, to calculate the occupancy and admission rates on the day before expected discharge, we first need to predict the expected day of discharge. To do this, we regress, by DRG, patient characteristics (see Appendix 1 for a list of controls and model results for the DRG related to hip and knee replacement) on log LOS, with robust standard errors clustered by hospital. We then predict each patient's LOS using the coefficients from our model, and round to the nearest integer.

We restrict our sample to patients with expected LOSs of at least three days so that we can detect day-level changes in LOS. Note, while we observe the workload measures on the day of admission and day before expected discharge, we do not control for the level of busyness on the other days of the stay since we do not expect them to have a significant effect on LOS. Furthermore, we are concerned about over-specification due to the high correlation in occupancy of the same type of patients between days, as

shown in Table 2; the correlation between day “D” and day “D+1” is 0.86 for medical patients, and 0.85 for surgical patients.

As it is difficult to interpret hazard rate changes (the β 's in eq. 6), we will present the (1) hazard ratio, and (2) hazard rate margins (Hoetker 2007, Kuntz et al. 2013). A hazard ratio is the ratio of hazard rates between the treatment and control group. In our setting, where the variables of interest are continuous, the hazard ratio is the proportional effect of a 1% increase in occupancy on the hazard rate, where a hazard ratio of less than 1 means that an event is less likely to happen (i.e., a patient is less likely to be discharged, so the LOS increases) as the occupancy increases, compared to the baseline condition. A hazard ratio greater than one means the event is more likely to happen (i.e., a patient is more likely to be discharged, so the LOS decreases), and a hazard ratio of 1 means there is no difference. For example, a hazard ratio of 0.95 is equivalent to saying that a 1% increase in occupancy is associated with a 5% decrease in the hazard rate, and thus a patient is more likely to stay in the hospital longer. The actual change in LOS will be dependent on the underlying hazard function for the patient given his characteristics.

In addition to the hazard ratio, we will also present the hazard rate margins at a variety of variable values. The hazard rate margins produce the average absolute change in the hazard rate for a 1% increase in the occupancy, given that the occupancy is at a set value. This value can be considered the average partial effect. For example, we can find the average hazard rate change that results from an increase in the medical occupancy on day of admission by 1%, given that the occupancy was 60%. The reason we choose a set of points to evaluate the hazard rate, as opposed to just averaging the partial effects of a 1% increase over the given occupancy rates, is our non-linear effects. Since we predict that increased medical and surgical occupancy will have U-shaped effects, we are able to get a better sense of what happens by selecting evaluation points. For marginal effects, we evaluate at occupancy levels of 0%, 25%, 50%, 65%, 85%, 92%, and 100%. We chose these values to represent a range on the spectrum, and specifically chose 85% as that is often cited as a target occupancy, and 92%, as Kuntz et al. (2013) found that to be a tipping point in terms of quality. We ran other values as robustness checks and found similar results.

Given the set of hazard functions, we can also solve for the patient’s survival function $S_i(t)$, which gives the chances of surviving past time t , and equals

$$S_i(t) = \exp \left\{ \sum_{d=1}^t \ln[1 - h_i(d)] \right\}, \text{ where} \quad (7)$$

$$S_i(0) = 1 \quad \forall i$$

As with the hazard function, we can calculate these survival functions for each patient for each day, $d=1,2,\dots,D$, to yield a survival curve, and find the expected survival time of the patient, or LOS, of a patient with a specific set of characteristics, and the relative effects of each of these characteristics, as

$$E(LOS) = \sum_{t=1}^K t * [S_i(t) - S_i(t - 1)] \quad (8)$$

where K is the expected maximum LOS in days. However it should be noted that the effect of occupancy on the survival function will vary based on all patient and occupancy level variables. As explained by Hoetker (2007), the treatment directly affects the hazard rate, but that rate is also determined by the patient's other characteristics as well as the baseline level of the variable of interest. For example, a 1% increase in occupancy from 0% occupancy can have a different change in LOS than a 1% increase at 50% occupancy, and the same 1% increase will have a different effect if a patient is 65 years old or 85 years old. When the treatment is binary, survival functions can be determined for each patient both with and without the treatment of interest, and the average LOSs can be compared. However, since we have multiple variables of interest, and none of our variables of interest are binary, but instead continuous and non-linear, and thus the effects of additional occupancy differ at different base occupancy levels, we cannot run survival analyses for each possible patient characteristic combination. Instead, we will present the effects of multiple changes in occupancy for a given patient simply to provide an example of the magnitude and directions of changes of LOS associated with each variable.

We use Stata 12 for all of our analysis. The hazard functions are solved using maximum likelihood estimation on our inpatient hospital dataset, yielding us a model of defined coefficients for each hazard function. We use the *hr* command for hazard ratios, and *margins* for our margins/average partial effects.

6. Results

6.1 Inpatient Workload

On the day of admission, the occupancy of same type patients has a statistically significant effect on the LOS of patients (Hypothesis 1a). As seen in Model 1 of Table 3, for the five medical DRGs, the hazard ratios for Cellulitis and Heart Failure are statistically significantly less than 1 at 0.9919 ($p < 0.05$) and 0.9928 ($p < 0.10$), respectively. Similarly, for surgical patients, the hazard ratio for Hip and Knee Replacement/Revision patients is statistically significant at 0.9948 ($p < 0.05$). The hazard ratios for the other DRGs, with the exception of spinal fusion patients, are directionally consistent, with values less than 1. This indicates that patients who are admitted to the hospital, on a day when there is a high occupancy of the same type of inpatient, will have a longer LOS than otherwise expected. Furthermore, when we look at the marginal effects (Table 4) of same patient type occupancy on day of admission, we see that at low to medium occupancy levels (under 85% occupancy), each additional patient reduces the hazard rate, thus increasing the LOS, with a flattening out (and slight reduction for some surgical DRGs)

as occupancy approaches 100%. Again, this is consistent with Hypothesis 1a that increased inpatient workload of same type patients on the day of admission increases LOS due to congestion effects.

TABLES 3 AND 4 ABOUT HERE

We believe that this is because admitting a new patient is a time consuming process that is difficult for nurses to accomplish if they have a heavy load of current patients. As a result, when occupancy levels are high, the newly admitted patients' care might be postponed until the nurse can complete important work for his or her other patients. All five nurses whom we interviewed stated that admissions were time consuming, and as a result, they tried to first make sure that their current patients' needs were met before beginning the process of admitting a new patient. For example one nurse described the challenge a co-worker had trying to balance the work demands of a new admission with her existing patients. "She had a little difficulty. She was struggling to juggle her patients, 'Should I go see this new patient first? But if I see this new patient, I'm missing the opportunity to see my other three patients who are on the floor.' When she did attend to the newly admitted patient, she was in that room for at least a good hour."

We also find evidence that the occupancy level of the other type of patients affects LOS, such that medical (surgical) patient occupancy affects the LOS of surgical (medical) patients on the day of admission (Hypothesis 1b). Looking at Model 2 of Table 3, we see that for the 3 surgical DRGs, the hazard rates associated with medical occupancy levels are all statistically significantly less than 1 (0.9704, $p < 0.01$ Cholecystectomy; 0.9806, $p < 0.05$ Spinal Fusion; 0.9674, $p < 0.01$ Hip and Knee Replacement/Revision). For medical patients, the effects of surgical occupancy are not as strong statistically, but again, are in the same direction. The effects may not be as strong due to the fact that for the typical hospital, there are two times as many medical patients as surgical, so the absolute effects of surgical patients on shared resources are smaller. We find similar results as above in the marginal effects model (See Table 4). We also observe some decreasing of LOS in the square term when the occupancy gets very high. From these collective results we conclude that there is a leveling off of LOS increases due to congestion effects. This may be because at high levels of occupancy, staff might be too busy to prioritize discharges, and as a result, the care of incoming patients is not delayed by prioritizing discharge patients, as it is for lower levels of occupancy.

To give a sense of the magnitude of the effects of same and other type congestion on the day of admission on LOS, we first generated the expected LOS for a "typical" hip and knee replacement patient (See Appendix 2 for values used). We then varied the medical and surgical occupancies on the day of admission (see Figure 2a). We find that the maximum change in LOS due to own type (surgical) occupancy occurs as occupancy increases from 0% to 51%, with a change of 0.11 days (2.7 hours), or 2.8% over the 0% occupancy. It should be noted, that this is a conservative estimate as the actual maximum occurs somewhere between 51 and 85%. The maximum change in LOS due to other type

(medical) occupancy occurs as occupancy increases from 0% to 86%, with a change of 0.76 days (18.2 hours), or 22.9% over the 0% occupancy LOS. As we mentioned above, the effect of medical occupancy is expected to be stronger due to the absolute number of medical patients being greater.

FIGURE 2a ABOUT HERE

As same type occupancy increases on the day before expected discharge, we find an increase in the hazard rate at lower occupancy levels, and a decrease in the hazard rate at high occupancy levels, consistent with a U-shaped effect on LOS (Hypothesis 2a). Model 1 of Table 5 shows that the linear occupancy terms for medical and surgical patients result in hazard ratios statistically greater than one for own type patients. For each DRG, the results are significant at the $p < 0.01$ level for all conditions, with a range in hazard ratios of 1.0539 to 1.0695 for medical patients, and 1.0201 to 1.0278 for surgical patients. Thus, as occupancy of one's own patient type increases, the probability of being discharged early increases. This is indicative of a workload smoothing effect where employees discharge current patients near the end of their stay to make room for incoming patients. However, this effect is moderated by the occupancy level. The square occupancy terms for both medical and surgical patients are statistically less than 1 ($p < 0.01$), suggesting that at high occupancy levels, an additional patient reduces the probability of discharge and increases the LOS, which is consistent with congestion effects where work stalls due to queuing and mental overload. Table 6 shows the marginal effect of increased occupancy at different base occupancy levels. We find that for occupancy levels below 85%, an increase in occupancy of 1% increases the hazard rate, thus reducing LOS, but over 85% (around the target occupancy level of most hospitals), the same 1% increase in occupancy reduces the hazard rate, resulting in an increased LOS. These results support Hypothesis 2a that increased occupancy of same type patients on the day before expected discharge has a U-shaped effect on LOS. Thus, workload smoothing defines the effect of increased load, until congestion effects take over at loads above 85%.

TABLES 5 AND 6 ABOUT HERE

Consistent with Hypothesis 2b, we find the same U-shaped effect on LOS when other type patient occupancy increases on the day before expected discharge (Model 2 of Table 5). The linear hazard ratios range from 1.0129 to 1.0697 ($p < 0.01$ for all), with the square hazard ratios ranging from 0.9996 to 0.9999 ($p < 0.01$ for all). Margin analysis shows that between 65 and 85% occupancy results in a switch from decreasing LOS to increasing LOS as occupancy increases (Table 6). These results are consistent with the fact that while beds and certain medical personnel, such as physicians, are specialized, resources throughout the hospital are shared. As a result, occupancy of other types of patients causes queuing and slow down for many resources, such as radiology tests, pharmacy, and patient transport.

Again, we use survival analysis to measure the magnitude of the change in expected LOS associated with changes in occupancy on the day before expected discharge (see Figure 2b). As same type (surgical)

occupancy increases from 0% to 51%, the LOS decreases by 0.68 days (16.4 hours), or 14.4% from the maximum. As same type occupancy continues to increase, the expected LOS increases by 0.19 days (4.5 hours) or 4.6% from the minimum. Similarly, when other type (medical) occupancy increases from 0% to 86%, the expected LOS decreases by 2.62 days (62.8 hours), or 39.1%. Once the occupancy increases to 100%, the expected LOS increases by 0.16 days (3.9 hours), or 4% from the minimum.

FIGURE 2b ABOUT HERE

6.2 Incoming Patients

Our results show that on the day before expected discharge, a higher number of same type admissions results in a hazard ratio greater than one. Four of the medical DRGs and two of the surgical DRGs are statistically greater than one, while the others are directionally, though not statistically, greater than one (See Model 1 of Table 7). Thus, a 1% increase in the relative number of same type admissions results in a 0.1%-0.2% increase in the hazard rate, resulting in an earlier discharge and lower LOS, as predicted in Hypothesis 3. On the day of expected discharge, only pneumonia has a statistically significant hazard ratio greater than one, suggesting that the greatest effect is when patients are nearing discharge, and an early discharge is possible (Model 3 of Table 7). Using survival analysis, we find that the effect of incoming surgical patients causes the LOS of a “typical” hip and knee replacement patient to decrease by 0.07 days (1.6 hours) or 1.6%. Unlike with inpatient workload, the number of the other type of patient admissions (i.e. medical) on the day before expected discharge does not affect the LOS of incoming patients (i.e. surgical) (Model 2 of Table 7). These results are consistent with our hypothesis that incoming patients cause employees to prioritize discharging current patients as a means to smooth their workloads and avoid the negative consequences of overwork.

TABLE 7 ABOUT HERE

6.3 Robustness

As a robustness check, in addition to analyzing the eight different DRGs, we also restricted our sample for analysis to only Medicare patients. Given that during the time of our study, Medicare paid a fixed fee for each DRG, there was an increased incentive for hospitals to discharge patients as soon as possible. The presence of this incentive makes any observed increase in LOS a conservative estimate of the overall effect. Although recent legislation prohibits Medicare payment for expenses related to unnecessary readmissions or medical errors made by the hospitals—which might encourage hospitals to keep patients longer than they otherwise would to help ensure that they do not have unreimbursed readmissions or medical treatment due to errors—this legislation was not enacted during our study period. Furthermore, Medicare patients are by definition over 65 and therefore are a more homogenous group than the regular

patient population, which will reduce noise in the sample. We find that these results are consistent with our findings, lending additional credibility to our results (See Appendix 3).

7. Discussion and Conclusion

We have contributed to the operations management literatures on healthcare operations and employee productivity by refining and characterizing the definition of workload and its effect on patient LOS. We distinguish between inpatient and incoming patient workloads, as well as their effects at differing points during a patient's stay. Our work shows that the type and timing, as well as the level, of patient workload matters in terms of LOS, with congestion effects dominating during the initial stage of a patient's stay, in addition to periods of very high load (>85% occupancy) towards the end of a patient's stay. On the other hand, workload smoothing characterizes the effects of increased load towards the end of the stay. Additionally, we show that when resources are shared across separate streams of patients the congestion and workload smoothing effects "spillover" across patient types. These results are significant since in most hospitals, each inpatient unit within a hospital generally optimizes its own staffing, capacity levels, and early discharging decisions. Our work shows that by doing so they may be producing less than optimal hospital-level performance, though each inpatient unit is optimized. Moreover, there has been little research into how shared resource capacity, such as laboratory staffing, affects LOS, but our work suggests that in addition to bed capacity, the availability of these shared resources has a significant impact on patient LOS. More work needs to address how the load of the hospital can be managed to improve productivity at the hospital level. Some hospitals may be better at optimizing at the hospital level, and we leave it for future work to determine which are, and what characteristics define them.

Most of the previous research on these topics has explored the effect of ICU workload on ICU LOS. We extend this body of research by analyzing workload at the hospital level, taking into account the entire LOS in the hospital. In contrast to prior research which finds that increasing workload is associated with *shorter* LOS (Chan et al. 2010, KC and Terwiesch 2012), we find that at very high workloads (>85%) or during the initial portion of the stay, LOS *increases* with increasing load. While at first glance our finding seems to contradict the prior literature, it is not necessarily inconsistent because prior studies have focused on settings where the ICU is at or very near 100% occupancy near the time of a patient's discharge, so those employees have no option but to discharge a patient from the ICU to make room for a severely ill incoming patient. In addition, this patient will likely be discharged to another inpatient unit within the hospital, which in our dataset would still be counted in the patient's LOS (making it longer), whereas in the prior studies, the unit-level discharge would be considered a shorter LOS. Furthermore, at low and medium occupancy levels, we also find that increased inpatient workload at the end of the stay increases the probability of an early discharge, but in these cases it seems that patients are truly leaving

the hospital rather than just moving to another inpatient unit. Finally, unlike previous studies, our results also included the effect of increased inpatient workload at the beginning of the stay, which results in increased LOS.

We find it interesting that our data suggest that 85% occupancy on the day before expected discharge is where additional inpatient workload increases the LOS because 85% is generally espoused as the target occupancy for optimal performance (Green and Nguyen 2001). Our work suggests that running at a slightly lower occupancy rate might reduce the hospital-level LOS of patients, thus allowing the same number of patients to be treated, but using fewer days per patient. One policy implication of our findings is that aggregate hospital capacity in a geographic region should allow for hospitals to be at less than 85% capacity. Another benefit of this policy could be a possible reduction in readmissions. While we do not analyze the effects of occupancy on clinical outcomes or readmissions, Kuntz et al. (2013) have shown that there is a tipping point, around 92% occupancy, at which mortality dramatically increases with increased workload. Even more concerning is that it has been suggested that the corner cutting that causes reductions in quality can persist even after the workload is reduced (Oliva and Sterman 2001). Others have shown that higher occupancy is also associated with higher readmissions, which effectively reduce the capacity of the hospital (KC and Terwiesch 2012), and as we show, have implications which are often underestimated as these readmitted patients affect patients beyond just the inpatient unit to which the patient is readmitted.

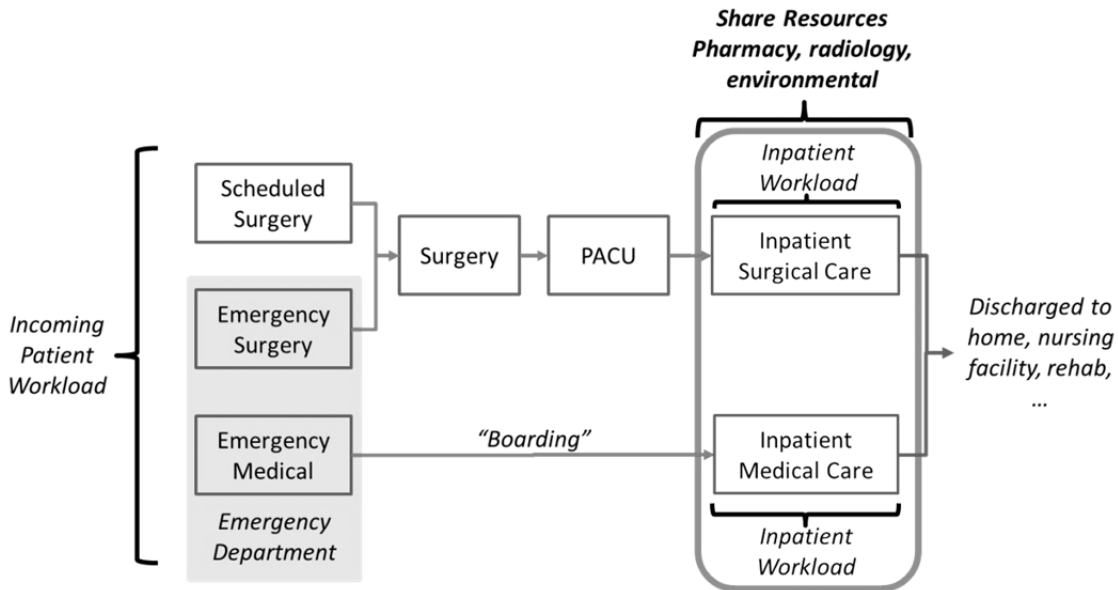
One possible mechanism to control discharge time as well as improve productivity in general so as to be able to maintain a lower occupancy is to employ an admission and discharge nurse. The time consuming nature of discharges and admissions suggests that it could be beneficial to have dedicated admission/ discharge nurses to ensure that this important work is not compromised when staff nurses become inundated with high workloads from existing patients. Our data suggest that nurses engage in workload smoothing by discharging patients earlier than they otherwise would when facing medium to high workloads and conversely become unable to discharge patients in a timely manner when they have heavy workloads. High loads from incoming patients of the same patient type also confirm these findings. A discharge nurse can eliminate the need for this prioritization and ensure the patient is discharged at the appropriate time based on the probability of readmission. When not working on discharges, the nurse can focus on admissions. Since admissions are time consuming, our work suggests that increased inpatient load causes delays in the start of care, which propagates the problem since it increases the chance of high workload throughout the stay. While some of the delay is likely due to queuing for resources, such as laboratory services, our discussions with nurses indicate that at least some is due to interruptions in the initial assessment due to the need to care for other patients. Adding a nurse who can focus solely on admissions and discharge will allow this nurse to perform the initial patient

assessment with less probability of interruption due to routine care, thus starting care sooner, reducing the chance of negative outcomes, and reducing the LOS. By removing the parts of care which generally require the largest blocks of time, other nurses may be able to attend to more patients, reducing the negative effects of high load. Further improving productivity, we would expect the behavioral workload smoothing and congestion effects that spillover across patients type would be likely to be reduced as a consequence of a discharge/admission nurse who controls the timing of these activities. Thus specialization, which is usually considered harmful to productivity in operations since it removes some benefits of flexibility, may actually be valuable in increasing in the throughput of patients.

There remains much to be explored in this area. As we mentioned, due to data limitations, we do not specifically test for changes in outcome quality associated with these changes in LOS. Furthermore, it is not yet known if some hospitals are better able to handle load, or if some shared resources have a greater effect on changes in LOS than others. Finally, we have only studied these effects in the hospital setting. While we think it is highly probable that these results apply to other settings with high worker discretion, such as law firms and call centers, more research is needed to confirm. However, we believe that this is an important contribution to better understanding how workload affects productivity at the hospital level.

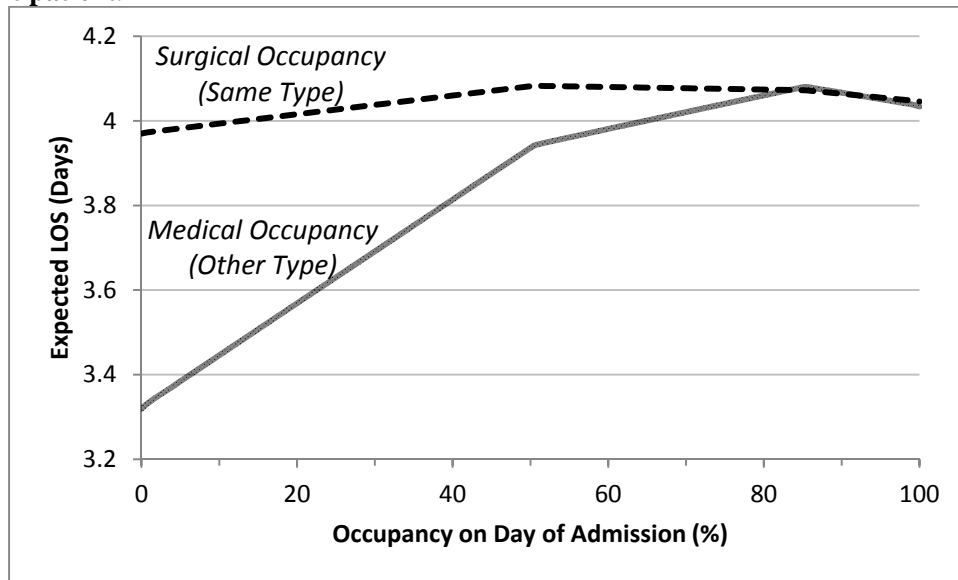
FIGURES

Figure 1. Flow of Admitted Patients Through the Hospital



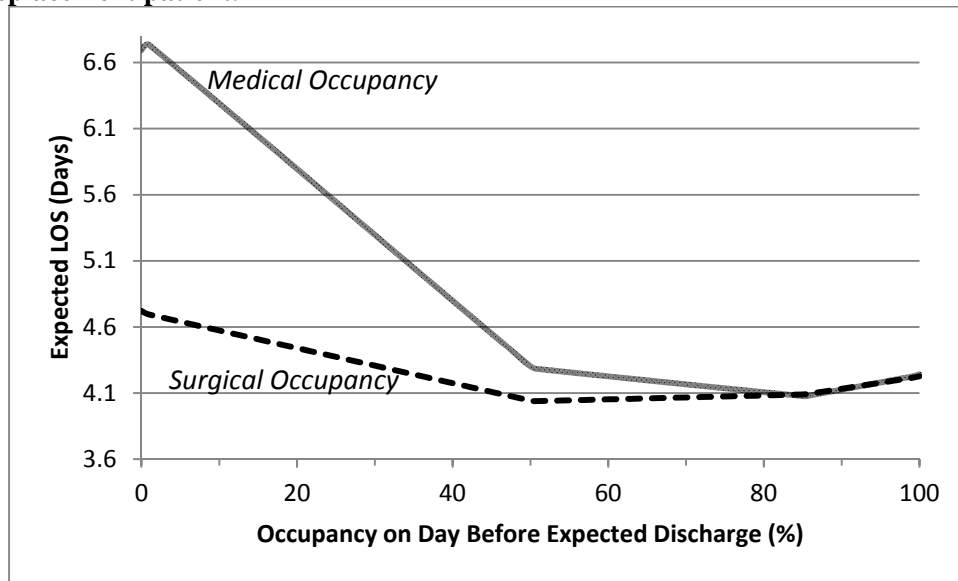
Notes: Patients are categorized as either medical or surgical, and enter either through the ED or via a scheduled surgery, and are considered incoming patients at this time. They then move to inpatient care, where they are treated separately, but share hospital resources. At this stage, they represent inpatient workload. Finally, patients are discharged to home or other nursing facility.

Figure 2a. Effect of Occupancy on Day of Admission, by type, on LOS of typical hip/knee replacement patient.



Notes: Variables kept at mean or most common values if categorical. The dotted line shows the effect of surgical occupancy at the start of day of admission as it moves from 0% to 100%, holding all other variables constant. The gray line shows the same, but for changes in medical occupancy.

Figure 2b. Effect of Occupancy on Day Before Expected Discharge, by type, on LOS of typical hip/knee replacement patient.



Notes: Variables kept at mean or most common values if categorical. The dotted line shows the effect of surgical occupancy at the end of the day before expected discharge as it moves from 0% to 100%, holding all other variables constant. The gray line shows the same, but for changes in medical occupancy.

TABLES

Table 1a. Hospital Level Summary Statistics

Number of Hospitals	243
Total Hospital Visits/Hospital (1/08-11/09)	26349 (15811)
Average Medical Patients/Day	71.3 (41.5)
Average Surgical Patients/Day	43.7 (36.5)
% Medical Patients	65.10% (11.2)
Patients in Sample	239,400
Average LOS (days)	4.84 (2.06)
Average Calculated expected LOS (days)	4.9 (1.94)

Standard Deviations in Parentheses

Table 1b. Summary Statistics of Controls (DRGs of Interest)

Continuous Variables (Averages)		
Procedure count	2.04	(1.53)
Number of comorbidities	7.93	(4.81)
Age	64.54	(16.36)
Scheduled admission census		
Day before expected discharge	40.05	(28.29)
Day of expected discharge	37.42	(28.79)
Categorical Variables (Percentages for most common categories)		
Day of Week - Weekday	88.53%	
Disposition - Home	46.41%	
Quarter of Year*		
First	25.86%	
Second	25.35%	
Third	24.67%	
Year - 2008**	50.94%	
Admit type - Emergency	53.02%	
Primary procedure day - Day of admission	69.18%	
Race		
White	66.73%	
Black	6.57%	
Hispanic	18.43%	
Sex - Female	59.42%	
Pay category		
Medicare	52.31%	
Medicaid	8.85%	
Private Insurance	29.81%	
Previously admitted	28.78%	
Primary diagnosis [by DRG]	No. Primary Diagnoses	Most Common Primary Diagnosis
Kidney/UTI w/o MCC [‡] [690]	21	UTI, site not specified
Nutritional and Metabolic Disorders w/o MCC [641]	49	Dehydration
Cellulitis w/o MCC [603]	68	Cellulitis, leg
Heart Failure w/CC [§] [292]	26	Congestive Heart Failure, unspecified
Pneumonia w/CC [194]	19	Pneumonia, organism unspecified
Laparoscopic Cholecystectomy w/o CDE ⁺ w/CC[418]	53	Acute cholecystitis, no obstruction
Spinal Fusion w/o MCC [460]	109	Lumbar or lumbosacral intervertebral disc
Hip/Knee Replacement w/o MCC [470]	210	Osteoarthritis, lower leg

*Based on visits in 2008; **Sample does not include 12/2009;

[‡] MCC = Major complications and/or comorbidities; [§]CC = Complications and/or comorbidities; ⁺CDE = Common duct exploration

Standard Deviations in Parentheses

Table 1c. Summary Statistics of Independent Variables
(Average % for DRGs of Interest)

Medical inpatient census		
Day of admission		75.72 (12.27)
Day before expected discharge		74.77 (12.29)
Surgical inpatient census		
Day of admission		70.81 (14.88)
Day before expected discharge		73.47 (14.55)
Medical admissions census		
Day of admission		64.50 (16.33)
Day before expected discharge		62.14 (16.61)
Day of expected discharge		62.03 (16.77)
Surgical admissions census		
Day of admission		60.49 (20.69)
Day before expected discharge		52.41 (22.06)
Day of expected discharge		51.01 (22.41)

Standard Deviations in Parentheses

Table 2. Correlation Table

			(1)		(2)		(3)		(4)		
			D	D+1	D	D+1	D	D+1	D	D+1	
Medical Occupancy	<i>Inpatient (1)</i>	D	1								
		D+1	0.858	1							
	<i>Incoming (2)</i>	D	0.237	0.529	1						
		D+1	0.248	0.25	0.202	1					
Surgical Occupancy	<i>Inpatient (3)</i>	D	0.234	0.249	0.154	0.183	1				
		D+1	0.249	0.26	0.153	0.169	0.854	1			
	<i>Incoming (4)</i>	D	0.198	0.168	0.132	0.126	0.229	0.534	1		
		D+1	0.186	0.203	0.15	0.172	0.228	0.237	0.261	1	

D=Day; D+1=Day after Day D; All values are significant at the p<0.05 level

Table 3. Hazard ratios associated with a change in inpatient occupancy (by type, medical or surgical) on the **day of admission** for the focal patient, where the focal patient is one of the eight DRGs listed.

			Same Type (Model 1)		Other Type (Model 2)	
DRG	n=		Medical	(Medical)^2	Surgical	(Surgical)^2
<i>Kidney/UTI (690)</i>	190,641		0.9945 (1.47)	1.0000 (1.03)	0.9975 (1.29)	1.0000 (1.07)
<i>Metabolic (641)</i>	137,221		0.9971 (0.80)	1.0000 (0.58)	0.9996 (0.15)	1.0000 (0.04)
Medical <i>Cellulitis (603)</i>	219,170		0.9919* (2.16)	1.0000+ (1.73)	0.9980 (0.96)	1.0000 (0.47)
<i>Heart Failure (292)</i>	178,604		0.9928+ (1.85)	1.0000 (1.60)	0.9960+ (1.75)	1.0000 (1.42)
<i>Pneumonia (194)</i>	227,273		0.9949 (1.46)	1.0000 (0.72)	1.0015 (0.67)	1.0000 (0.81)
DRG	n=		Surgical	(Surgical)^2	Medical	(Medical)^2
<i>Cholecystectomy (418)</i>	63,520		0.9944 (1.16)	1.0001 (1.62)	0.9704** (3.72)	1.0002** (3.48)
Surgical <i>Spinal Fusion (460)</i>	99,351		1.0003 (0.07)	1.0000 (0.34)	0.9806* (2.17)	1.0001+ (1.96)
<i>Hip and Knee (470)</i>	412,575		0.9948* (2.25)	1.0000* (2.49)	0.9674** (7.82)	1.0002** (7.26)

+p<0.1; *p<0.05; **p<0.01; Standard Errors in parentheses

Table 4. Average marginal change in hazard rates associated with a 1% increase in occupancy (by type) on day of admission on patients of each DRG, at baseline occupancies of N, where N=0, 25, 50, 65, 85, 92, and 100%

	N=	Medical					Surgical		
		Kidney/UTI (690)	Metabolic (641)	Cellulitis (603)	Heart Failure (292)	Pneumonia (194)	Cholecyst- ectomy (418)	Spinal Fusion (460)	Hip and Knee (470)
		190641	137221	219170	178604	227273	63520	99351	412575
Medical Occupancy	0	-0.0014 (0.0010)	-0.0008 (0.0010)	-0.0020+ (0.0010)	-0.0017 (0.0010)	-0.0011 (0.0008)	-0.0077** (0.0022)	-0.0043* (0.0021)	-0.0058** (0.0006)
	25	-0.0010 (0.0006)	-0.0005 (0.0006)	-0.0012* (0.0006)	-0.0011+ (0.0006)	-0.0009+ (0.0005)	-0.0046** (0.0014)	-0.0028* (0.0014)	-0.0043** (0.0005)
	50	-0.0007* (0.0003)	-0.0004 (0.0003)	-0.0008** (0.0003)	-0.0006* (0.0003)	-0.0007** (0.0002)	-0.0021** (0.0005)	-0.0015* (0.0007)	-0.0023** (0.0003)
	65	-0.0005** (0.0001)	-0.0002 (0.0002)	-0.0005** (0.0001)	-0.0003** (0.0001)	-0.0005** (0.0001)	-0.0009** (0.0003)	-0.0009** (0.0003)	-0.0012** (0.0001)
	85	-0.0002 (0.0002)	-0.0001 (0.0002)	-0.0001 (0.0002)	-0.0000 (0.0002)	-0.0004** (0.0001)	0.0004 (0.0003)	-0.0001 (0.0002)	0.0002 (0.0001)
	92	-0.0002 (0.0002)	-0.0001 (0.0003)	0.0000 (0.0002)	0.0001 (0.0002)	-0.0004* (0.0002)	0.0009* (0.0004)	0.0002 (0.0003)	0.0007** (0.0002)
	100	-0.0001 (0.0003)	0.0000 (0.0004)	0.0002 (0.0003)	0.0002 (0.0003)	-0.0003 (0.0002)	0.0015* (0.0006)	0.0005 (0.0005)	0.0013** (0.0003)
Surgical Occupancy	0	-0.0006 (0.0005)	-0.0001 (0.0006)	-0.0004 (0.0004)	-0.0009+ (0.0005)	0.0003 (0.0004)	-0.0010 (0.0010)	0.0001 (0.0008)	-0.0010* (0.0004)
	25	-0.0004 (0.0003)	-0.0001 (0.0004)	-0.0003 (0.0003)	-0.0006+ (0.0003)	0.0002 (0.0003)	-0.0005 (0.0006)	0.0002 (0.0006)	-0.0006* (0.0003)
	50	-0.0002+ (0.0001)	-0.0001 (0.0002)	-0.0003+ (0.0001)	-0.0003* (0.0001)	0.0000 (0.0001)	0.0000 (0.0003)	0.0003 (0.0003)	-0.0002 (0.0001)
	65	-0.0001 (0.0001)	-0.0001 (0.0001)	-0.0002** (0.0001)	-0.0002+ (0.0001)	-0.0000 (0.0001)	0.0003+ (0.0002)	0.0003* (0.0002)	0.0000 (0.0001)
	85	0.0000 (0.0001)	-0.0001 (0.0002)	-0.0002 (0.0001)	0.0000 (0.0002)	-0.0001 (0.0001)	0.0008* (0.0003)	0.0004+ (0.0002)	0.0003* (0.0001)
	92	0.0001 (0.0002)	-0.0001 (0.0002)	-0.0001 (0.0002)	0.0001 (0.0002)	-0.0002 (0.0002)	0.0009* (0.0004)	0.0005 (0.0003)	0.0004* (0.0002)
	100	0.0001 (0.0002)	-0.0001 (0.0003)	-0.0001 (0.0002)	0.0002 (0.0003)	-0.0002 (0.0002)	0.0022* (0.0005)	0.0005 (0.0004)	0.0005** (0.0002)

+p<0.1; *p<0.05; **p<0.01; Standard Errors in parentheses

Table 5. Hazard ratios associated with a change in inpatient occupancy (by type, medical or surgical) on the **day before expected discharge** for the focal patient, where the focal patient is one of the eight DRGs listed.

		Same Type (Model 1)		Other Type (Model 2)	
DRG	n=	<i>Medical</i>	<i>(Medical)^2</i>	<i>Surgical</i>	<i>(Surgical)^2</i>
<i>Kidney/UTI (690)</i>	190,641	1.0673** (13.24)	0.9995** (13.33)	1.0165** (4.13)	0.9999** (4.09)
<i>Metabolic (641)</i>	137,221	1.0648** (15.12)	0.9995** (16.08)	1.0129** (3.47)	0.9999** (3.60)
Medical					
<i>Cellulitis (603)</i>	219,170	1.0603** (17.08)	0.9996** (17.77)	1.0189** (5.44)	0.9999** (5.22)
<i>Heart Failure (292)</i>	178,604	1.0539** (11.45)	0.9996** (12.73)	1.0178** (4.91)	0.9999** (4.75)
<i>Pneumonia (194)</i>	227,273	1.0695** (17.41)	0.9995** (18.21)	1.0132** (3.22)	0.9999** (3.26)
DRG	n=	<i>Surgical</i>	<i>(Surgical)^2</i>	<i>Medical</i>	<i>(Medical)^2</i>
<i>Cholecystectomy (418)</i>	63,520	1.0278** (5.25)	0.9998** (5.85)	1.0639** (10.70)	0.9996** (10.16)
Surgical					
<i>Spinal Fusion (460)</i>	99,351	1.0201** (4.11)	0.9998** (4.96)	1.0622** (11.01)	0.9996** (11.23)
<i>Hip and Knee (470)</i>	412,575	1.0251** (4.05)	0.9998** (5.35)	1.0697** (14.43)	0.9996** (14.43)

*p<0.1; **p<0.05; ***p<0.01; Standard Errors in parentheses

Table 6. Average marginal change in hazard rates associated with a 1% increase in occupancy (by type) on day before expected discharge on patients of each DRG, at baseline occupancies of N, where N=0, 25, 50, 65, 85, 92, and 100%

	Medical					Surgical			
	<i>Kidney/UTI (690)</i>	<i>Metabolic (641)</i>	<i>Cellulitis (603)</i>	<i>Heart Failure (292)</i>	<i>Pneumonia (194)</i>	<i>Cholecyst- ectomy (418)</i>	<i>Spinal Fusion (460)</i>	<i>Hip and Knee (470)</i>	
<i>N=</i>	<i>190641</i>	<i>137221</i>	<i>219170</i>	<i>178604</i>	<i>227273</i>	<i>63520</i>	<i>99351</i>	<i>412575</i>	
Medical Occupancy	0	0.0113** (0.0007)	0.0121** (0.0007)	0.0077** (0.0003)	0.0078** (0.0006)	0.0090** (0.0004)	0.0019** (0.0002)	0.0019** (0.0007)	0.0023** (0.0002)
	25	0.0095** (0.0004)	0.0092** (0.0003)	0.0076** (0.0003)	0.0077** (0.0006)	0.0092** (0.0004)	0.0039** (0.0002)	0.0038** (0.0010)	0.0048** (0.0001)
	50	0.0035** (0.0001)	0.0029** (0.0001)	0.0034** (0.0001)	0.0033** (0.0002)	0.0036** (0.0001)	0.0033** (0.0003)	0.0032** (0.0006)	0.0039** (0.0003)
	65	0.0010** (0.0001)	0.0006** (0.0001)	0.0008** (0.0001)	0.0007** (0.0002)	0.0010** (0.0001)	0.0015** (0.0002)	0.0015** (0.0003)	0.0018** (0.0002)
	85	-0.0019** (0.0003)	-0.0022** (0.0003)	-0.0026** (0.0002)	-0.0028** (0.0003)	-0.0022** (0.0002)	-0.0016** (0.0003)	-0.0013** (0.0003)	-0.0013** (0.0002)
	92	-0.0031** (0.0003)	-0.0035** (0.0005)	-0.0038** (0.0003)	-0.0041** (0.0003)	-0.0034** (0.0003)	-0.0025** (0.0004)	-0.0022** (0.0005)	-0.0024** (0.0002)
	100	-0.0046** (0.0004)	-0.0052** (0.0006)	-0.0052** (0.0003)	-0.0055** (0.0004)	-0.0050** (0.0004)	-0.0034** (0.0004)	-0.0031** (0.0006)	-0.0035** (0.0003)
Surgical Occupancy	0	0.0034** (0.0007)	0.0029** (0.0008)	0.0031** (0.0005)	0.0031** (0.0006)	0.0022** (0.0006)	0.0029** (0.0003)	0.0025** (0.0006)	0.0033** (0.0004)
	25	0.0026** (0.0007)	0.0020** (0.0006)	0.0026** (0.0005)	0.0025** (0.0005)	0.0017** (0.0006)	0.0025** (0.0005)	0.0019** (0.0005)	0.0025** (0.0006)
	50	0.0013** (0.0004)	0.0008* (0.0003)	0.0014** (0.0003)	0.0013** (0.0003)	0.0008* (0.0003)	0.0009** (0.0003)	0.0006* (0.0003)	0.0009* (0.0004)
	65	0.0004+ (0.0002)	0.0000 (0.0002)	0.0005** (0.0002)	0.0004* (0.0002)	0.0002* (0.0002)	-0.0003 (0.0002)	-0.0002 (0.0002)	-0.0001 (0.0003)
	85	-0.0008* (0.0003)	-0.0010** (0.0004)	-0.0007** (0.0002)	-0.0008** (0.0003)	-0.0006* (0.0003)	-0.0018** (0.0003)	-0.0013** (0.0003)	-0.0015** (0.0002)
	92	-0.0012** (0.0004)	-0.0014** (0.0005)	-0.0011** (0.0003)	-0.0012** (0.0004)	-0.0009** (0.0003)	-0.0022** (0.0004)	-0.0017** (0.0004)	-0.0020** (0.0002)
	100	-0.0017** (0.0005)	-0.0018** (0.0006)	-0.0016** (0.0004)	-0.0016** (0.0005)	-0.0012** (0.0004)	-0.0026** (0.0004)	-0.0020** (0.0005)	-0.0024** (0.0003)

*p<0.1; *p<0.05; **p<0.01,; Standard Errors in parentheses

Table 7. Hazard ratios associated with a change in **incoming occupancy** (by type, medical or surgical) on the day before and day of expected discharge for the focal patient, where the focal patient is one of the eight DRGs listed.

		Day Before Expected Discharge		Day of Expected Discharge	
		Same Type (Model 1)	Other Type (Model 2)	Same Type (Model 3)	Other Type (Model 4)
DRG	n=	<i>Medical</i>	<i>Surgical</i>	<i>Medical</i>	<i>Surgical</i>
<i>Kidney/UTI (690)</i>	190,641	1.0009 (1.61)	0.9996 (0.77)	1.0002 (0.38)	1.0006 (1.16)
<i>Metabolic (641)</i>	137,221	1.0029** (4.25)	1.0002 (0.30)	1.0007 (1.15)	1.0005 (1.01)
Medical <i>Cellulitis (603)</i>	219,170	1.0010+ (1.86)	0.9997 (0.75)	1.0002 (0.53)	1.0000 (0.13)
<i>Heart Failure (292)</i>	178,604	1.0023** (3.67)	1.0003 (0.63)	1.0004 (0.72)	1.0003 (0.60)
<i>Pneumonia (194)</i>	227,273	1.0013+ (1.91)	1.0000 (0.08)	1.0017** (3.39)	1.0005 (1.01)
DRG	n=	<i>Surgical</i>	<i>Medical</i>	<i>Surgical</i>	<i>Medical</i>
<i>Cholecystectomy (418)</i>	63,520	1.0011+ (1.65)	1.0010 (1.21)	0.9999 (0.13)	1.0001 (0.22)
Surgical <i>Spinal Fusion (460)</i>	99,351	1.0006 (1.20)	0.9997 (0.49)	1.0006 (1.24)	1.0005 (1.10)
<i>Hip and Knee (470)</i>	412,575	1.0009* (2.44)	1.0002 (0.60)	1.0001 (0.41)	1.0000 (0.14)

+p<0.1; *p<0.05; **p<0.01; Standard Errors in parentheses

References

- Aiken, L. H., S. P. Clarke, D. M. Sloane, J. Sochalski, J. H. Silber. 2002. Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *J. Amer. Medical Assoc.* **288**(16) 1987-1993.
- Batt, R. J., C. Terwiesch, O. A. Soremekun. 2012. Docs under load: An empirical study of state-dependent service rate mechanisms. *Working Paper, Wharton School of Business, Pennsylvania.*
- Blow, O., L. Magliore, J. A. Claridge, K. Butler, J. S. Young. 1999. The golden hour and the silver day: Detection and correction of occult hypoperfusion within 24 hours improves outcome from major trauma. *J. Trauma-Injury, Infection, and Critical Care.* **47**(5) 964.
- Boudreau, J., W. Hopp, J. O. McClain, L. J. Thomas. 2003. On the interface between operations and human resources management. *M&SOM.* **5**(3) 179-202.
- Buist, M. D., G. E. Moore, S. A. Bernard, B. P. Waxman, J. N. Anderson, T. V. Nguyen. 2002. Effects of a medical emergency team on reduction of incidence of and mortality from unexpected cardiac arrests in hospital: Preliminary study. *BMJ.* **324**(7334) 387-390.
- Centers for Medicare & Medicaid Services Office of the Actuary National Health Statistics Group. 2012. *National health care expenditures data.* <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/downloads/tables.pdf>.
- Chalfin, D. B., S. Trzeciak, A. Likourezos, B. M. Baumann, R. P. Dellinger. 2007. Impact of delayed transfer of critically ill patients from the emergency department to the intensive care unit *Crit. Care Medicine.* **35**(6) 1477-1483.
- Chan, C. W., V. F. Farias, N. Bambos, G. J. Escobar. 2010. Maximizing throughput of hospital intensive care units with patient readmissions. *Working Paper, Columbia Business School, New York.*
- Clarke, A. 1996. Why are we trying to reduce length of stay? Evaluation of the costs and benefits of reducing time in hospital must start from the objectives that govern change. *Quality in Health Care.* **5**(3) 172.
- Dallery, Y., S. B. Gershwin. 1992. Manufacturing flow line systems: A review of models and analytical results. *Queueing Systems.* **12**(1) 3-94.
- Debo, L. G., L. B. Toktay, L. N. Van Wassenhove. 2008. Queuing for expert services. *Management Sci.* **54**(8) 1497-1512.
- Eddy, D. M. 1984. Variations in physician practice: The role of uncertainty. *Health Affairs.* **3**(2) 74-89.
- George, J. M., J. M. Harrison. 2001. Dynamic control of a queue with adjustable service rate. *Oper. Res.* **49**(5) 720-731.
- Gittell, J. H., K. M. Fairfield, B. Bierbaum, W. Head, R. Jackson, M. Kelly, R. Laskin, S. Lipson, J. Siliski, T. Thornhill, J. Zuckerman. 2000. Impact of relational coordination on quality of care, postoperative pain and functioning, and length of stay: A nine-hospital study of surgical patients. *Medical Care.* **38**(8) 807-819.
- Green, L., S. Savin, N. Savva. 2011. "Nurse vendor problem": Personnel staffing in the presence of endogenous absenteeism. *Management Sci.* (Forthcoming).
- Green, L. V., V. Nguyen. 2001. Strategies for cutting hospital beds: The impact on patient service. *Health Services Res.* **36**(2) 421-442.
- Ha, A. Y. 1998. Incentive-compatible pricing for a service facility with joint production and congestion externalities. *Management Sci.* **44**(12) 1623-1636.
- Hauck, K., X. Zhao. 2011. How dangerous is a day in hospital?: A model of adverse events and length of stay for medical inpatients. *Medical Care.* **49**(12) 1068-1075
- Hoetker, G. 2007. The use of logit and probit models in strategic management research: Critical issues. *Strat. Management J.* **28** 331-343.
- Hopp, W. J., S. M. R. Iravani, F. Liu. 2009. Managing white-collar work: An operations-oriented survey. *Prod. & Oper. Management.* **18**(1) 1-32.
- Hopp, W. J., S. M. R. Iravani, G. Y. Yuen. 2007. Operations systems with discretionary task completion. *Management Sci.* **53**(1) 61-77.

- Jenkins, S. P. 2004. *Survival analysis*. Institute for Social and Economic Research, University of Essex, Colchester, UK.
- Jouini, O., Y. Dallery, R. Nait-Abdallah. 2008. Analysis of the impact of team-based organizations in call center management. *Management Sci.* **54**(2) 400-414.
- KC, D., C. Terwiesch. 2009. Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *Management Sci.* **55**(9) 1486-1498.
- KC, D. S., C. Terwiesch. 2012. An econometric analysis of patient flows in the cardiac intensive care unit. *M&SOM*. **14**(1) 50-65.
- Kuntz, L., R. Mennicken, S. Scholtes. 2013. Stress on the ward: Evidence of safety tipping points in hospitals. *Working Paper, Cambridge Judge Business School, UK*.
- Little, J. D. C. 1961. A proof for the queuing formula: $L = \lambda w$. *Oper. Res.* **9**(3) 383-397.
- Meyer, H. 2011. At upmc, improving care processes to serve patients better and cut costs. *Health Affairs*. **30**(3) 400-403.
- Office of Inspector General. 2001. *Medicare hospital prospective payment system: How drg rates are calculated and updated*. <http://oig.hhs.gov/oei/reports/oei-09-00-00200.pdf>.
- Oliva, R., J. D. Serman. 2001. Cutting corners and working overtime: Quality erosion in the service industry. *Management Sci.* **47**(7) 894-914.
- Powell, A., S. Savin, N. Savva. 2012. Physician workload and hospital reimbursement: Overworked servers generate lower income. *M&SOM*. **14**(4) 512-528.
- Roberts, R. R., P. W. Frutos, G. G. Ciavarella, L. M. Gussow, E. K. Mensah, L. M. Kampe, H. E. Straus, G. Joseph, R. J. Rydman. 1999. Distribution of variable vs fixed costs of hospital care. *J. Amer. Medical Assoc.* **281**(7) 644-649.
- Schultz, K. L., D. C. Juran, J. W. Boudreau. 1999. The effects of low inventory on the development of productivity norms. *Management Sci.* **45**(12) 1664-1678.
- Schultz, K. L., D. C. Juran, J. W. Boudreau, J. O. McClain, L. J. Thomas. 1998. Modeling and worker motivation in jit production systems. *Management Sci.* **44**(12) 1595-1607.
- Schultz, K. L., J. O. McClain, L. J. Thomas. 2003. Overcoming the dark side of worker flexibility. *J. Oper. Management*. **21**(1) 81-92.
- Sebelius, K. 2013. *Affordable care act at 3: Paying for quality saves health care dollars* Health Affairs Blog.
- Shapiro, R. D. 1996. National cranberry cooperative. *Harvard Business School Case*. #688-122.
- Stidham, S., R. R. Weber. 1989. Monotonic and insensitive optimal policies for control of queues with undiscounted costs. *Oper. Res.* **37**(4) 611-625.
- Tan, T. F., S. Netessine. 2012. When does the devil make work? An empirical study of the impact of workload on worker productivity. *Working Paper, INSEAD, France*.
- Tucker, A. L., A. C. Edmondson. 2003. Why hospitals don't learn from failures: Organizational and psychological dynamics that inhibit system change. *Calif. Management Rev.* **45**(2).
- Tucker, A. L., S. J. Spear. 2006. Operational failures and interruptions in hospital nursing. *Health Services Res.* **41**(3p1) 643-662.

APPENDICES

Appendix 1. Estimation of LOS for Hip/Knee Replacement Patients

Linear regression

Number of obs = 124626

R-squared = 0.4147

Root MSE = .21477

(Std. Err. adjusted for 240 clusters in Hospital)

log_los	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
Day of Week						
[Sunday]						
1 (Monday)	-.0015075	.0080726	-0.19	0.852	-.0174099	.0143949
2	.0096599	.0082954	1.16	0.245	-.0066815	.0260013
3	.0149764	.0082331	1.82	0.070	-.0012423	.031195
4	.0493502	.0098866	4.99	0.000	.0298742	.0688262
5	.042038	.0076012	5.53	0.000	.0270641	.0570118
6	.0018502	.0076421	0.24	0.809	-.0132042	.0169046
Disposition						
[Unknown]						
Home	-.0921721	.0187588	-4.91	0.000	-.1291258	-.0552183
Acute-Hosp	-.2378591	.1804571	-1.32	0.189	-.5933485	.1176304
Acute-Other	-.1233339	.0263107	-4.69	0.000	-.1751644	-.0715034
SNIC-Hosp	-.1076678	.0334552	-3.22	0.001	-.1735725	-.041763
SNIC-Other	-.0501372	.0193279	-2.59	0.010	-.0882121	-.0120624
Residential	-.0992064	.0246622	-4.02	0.000	-.1477893	-.0506235
Home Health	-.0618525	.018814	-3.29	0.001	-.0989149	-.0247901
Other	-.0838973	.0201558	-4.16	0.000	-.1236029	-.0441917
Month						
[January]						
2	-.0079493	.0033187	-2.40	0.017	-.014487	-.0014116
3	-.01206	.003743	-3.22	0.001	-.0194334	-.0046866
4	-.0144458	.0039121	-3.69	0.000	-.0221525	-.0067392
5	-.0169812	.0039205	-4.33	0.000	-.0247043	-.0092581
6	-.0163335	.0032961	-4.96	0.000	-.0228265	-.0098404
7	-.0150674	.0038035	-3.96	0.000	-.02256	-.0075747
8	-.024063	.0037844	-6.36	0.000	-.031518	-.016608
9	-.0238179	.0036074	-6.60	0.000	-.0309243	-.0167116
10	-.0275002	.0038148	-7.21	0.000	-.0350152	-.0199853
11	-.0286159	.0042595	-6.72	0.000	-.0370068	-.020225
12	-.0153144	.004293	-3.57	0.000	-.0237714	-.0068575
Admit type						
[Scheduled]						
Unscheduled	.0203388	.008754	2.32	0.021	.0030939	.0375837
Unknown	.0294281	.0301144	0.98	0.329	-.0298954	.0887516
Proc. Count	.0351415	.0020497	17.14	0.000	.0311037	.0391794
Diag. Count	.0129505	.0004799	26.99	0.000	.0120052	.0138959
Hospital ID	Included					
Primary Diag.	Included					
Primary Proc.	Included					
Proc. Day						
[Day Before]						
1	.086574	.0232324	3.73	0.000	.0408076	.1323404
2	.1817506	.0238901	7.61	0.000	.1346885	.2288127
3+	.3481239	.0237305	14.67	0.000	.3013762	.3948715
Race						
[Unknown]						
White	-.0051303	.0091705	-0.56	0.576	-.0231957	.0129351
Black	.0319338	.0105252	3.03	0.003	.0111999	.0526678
Hispanic	.0106101	.010503	1.01	0.313	-.0100802	.0313004
Pacific	.018364	.0117236	1.57	0.119	-.0047308	.0414589
Nat. Amer.	.0299414	.0175769	1.70	0.090	-.0046841	.0645669

Other	.0181231	.0113127	1.60	0.110	-.0041622	.0404084
Age	.0010847	.000143	7.58	0.000	.000803	.0013664
Sex						
[Male]						
Female	.0412991	.0026962	15.32	0.000	.0359878	.0466104
Other	.1704533	.0105226	16.20	0.000	.1497244	.1911823
Unknown	.0298172	.0195533	1.52	0.129	-.0087017	.068336
Ins. pay type						
[Blank]						
Medicare	-.0138867	.0524626	-0.26	0.791	-.1172348	.0894615
Medicaid	.0562161	.0529284	1.06	0.289	-.0480496	.1604819
Private	-.0307179	.0523652	-0.59	0.558	-.1338742	.0724383
Work. Comp	.0341272	.052001	0.66	0.512	-.0683117	.1365661
County Ind.	.0338987	.0540748	0.63	0.531	-.0726254	.1404228
Other Gov.	-.0049211	.057483	-0.09	0.932	-.1181591	.1083169
Other Ind.	.0232248	.0602885	0.39	0.700	-.0955399	.1419895
Self Pay	-.0054752	.0546648	-0.10	0.920	-.1131615	.1022111
Other	.0105268	.0526042	0.20	0.842	-.0931004	.1141539
Revisit						
[No revisit]						
1	.0104333	.0093399	1.12	0.265	-.0079658	.0288323
2+	.0115101	.0096493	1.19	0.234	-.0074986	.0305187
Constant	2.130724	.0671367	31.74	0.000	1.998469	2.26298

 *[Baseline]

Appendix 2. Values used to produce survival functions for hip and knee replacement patients

Variable Name	Value
Day of week of admission	Tuesday
Disposition	Home Health Service
Quarter	2
Year	2008
Race	White
Sex	Female
Admission type	Scheduled
Primary Diagnosis	Osteoarthritis, lower leg
Insurance type	Medicare
Revisit	1 st visit
Primary procedure day	Day of admission
Number of procedures	1
Number of diagnoses	4
Hospital	Cedars Sinai Medical Center
Medical Admissions (Day of admission)	73%
Medical Admissions (Day before expected d/c)	69%
Medical Admissions (Day of expected d/c)	71%
Surgical Admissions (Day of Admission)	73%
Surgical Admissions (Day before expected d/c)	63%
Surgical Admissions (Day of expected d/c)	64%
Scheduled Admissions (Day before expected d/c)	60%
Scheduled Admissions (Day of expected d/c)	69%%
Medical Occupancy (Day of admission)	87% [0,1,50,51,85,86,99,100]
Medical Occupancy (Day before expected d/c)	85% [0,1,50,51,85,86,99,100]
Surgical Occupancy (Day of Admission)	81% [0,1,50,51,85,86,99,100]
Surgical Occupancy (Day before expected d/c)	83% [0,1,50,51,85,86,99,100]

Note: For the occupancy values, the first number is the value during analysis if it is not the variable of interest, while the numbers in [] are the values used when the specific occupancy is adjusted to show changes in LOS related to the variable of interest

Appendix 3. Hazard Ratios for Medicare Only Focal Patients

		Medical					Surgical		
MS-DRG		690	641	603	292	194	418	460	470
	N=	121,875	89,715	84,941	126,080	150,338	23,515	38,462	251,460
Occupancy: Day of Admission	Medical	0.9890 (2.48)*	0.9980 (0.37)	0.9957 (0.69)	0.9895 (2.32)*	0.9980 (0.49)	0.9661 (2.26)*	0.9779 (1.57)	0.9675 (6.64)**
	(Medical) ²	1.0001 (1.88)+	1.0000 (0.13)	1.0000 (0.41)	1.0001 (2.09)*	1.0000 (0.28)	1.0002 (2.25)*	1.0001 (1.41)	1.0002 (5.96)**
	Surgical	0.9983 (0.71)	0.9989 (0.38)	0.9984 (0.48)	0.9983 (0.62)	1.0017 (0.63)	0.9751 (2.38)*	1.0000 (0.01)	0.9933 (2.72)**
	(Surgical) ²	1.0000 (0.50)	1.0000 (0.20)	1.0000 (0.01)	1.0000 (0.46)	1.0000 (0.92)	1.0002 (2.57)*	1.0000 (0.35)	1.0001 (2.93)**
	Medical	1.0760 (13.26)**	1.0682 (12.67)**	1.0658 (11.93)**	1.0532 (9.83)**	1.0690 (15.05)**	1.0805 (7.28)**	1.0635 (7.33)**	1.0730 (14.85)**
	(Medical) ²	0.9995 (14.59)**	0.9995 (13.08)**	0.9995 (12.37)**	0.9996 (10.79)**	0.9995 (15.40)**	0.9995 (7.15)**	0.9996 (7.11)**	0.9995 (14.48)**
Expected Discharge	Surgical	1.0081 (1.61)	1.0102 (2.15)*	1.0164 (3.38)**	1.0182 (4.33)**	1.0157 (3.51)**	1.0274 (2.74)**	1.0262 (3.49)**	1.0263 (4.16)**
	(Surgical) ²	0.9999 (1.65)+	0.9999 (2.29)*	0.9999 (2.90)**	0.9999 (4.10)**	0.9999 (3.44)**	0.9998 (2.79)**	0.9998 (4.09)**	0.9998 (5.30)**
	Medical	1.0005 (1.26)	1.0000 (0.03)	1.0007 (1.26)	1.0000 (0.06)	1.0003 (0.83)	1.0020 (1.48)	0.9999 (0.08)	0.9993 (1.77)+
Admissions: Day of Admission	Surgical	1.0003 (0.86)	1.0004 (1.21)	1.0002 (0.63)	1.0004 (1.05)	0.9999 (0.27)	0.9986 (1.28)	1.0016 (1.55)	1.0021 (3.86)**
	Medical	1.0017 (2.22)*	1.0025 (3.02)**	1.0014 (1.62)	1.0024 (3.22)**	1.0015 (1.94)+	0.9995 (0.36)	0.9998 (0.23)	1.0005 (1.14)
Expected Discharge	Surgical	0.9991 (1.28)	1.0012 (1.44)	1.0001 (0.10)	0.9998 (0.28)	0.9999 (0.18)	1.0008 (0.65)	1.0016 (1.75)+	1.0006 (1.46)
	Medical	1.0000 (0.04)	1.0016 (2.24)*	1.0003 (0.46)	0.9999 (0.22)	1.0018 (2.76)**	0.9983 (1.27)	0.9997 (0.42)	1.0002 (0.52)
Admissions: Day of Expected Discharge	Surgical	1.0014 (2.13)*	1.0003 (0.47)	0.9990 (1.52)	1.0011 (1.81)+	1.0003 (0.61)	1.0005 (0.42)	1.0000 (0.04)	0.9997 (0.93)

†p<0.1; *p<0.05; **p<0.01; Standard Errors in parentheses