Stress on the Ward: Evidence of Safety Tipping Points in Hospitals

Ludwig Kuntz
Faculty of Management, Economics and Social Sciences, University of Cologne kuntz@wiso.uni-koeln.de

Roman Mennicken
Rheinisch-Westfälisches Institut für Wirtschaftsforschung roman.mennicken@rwi-essen.de

Stefan Scholtes
Judge Business School, University of Cambridge s.scholtes@jbs.cam.ac.uk

Do hospitals experience safety tipping points as utilization increases and if so, what are the implications for hospital operations management? We argue that safety tipping points occur when managerial escalation policies are exhausted and workload variability buffers are depleted. Front-line clinical staff is forced to ration resources and, at the same time, becomes more error-prone as a result of elevated stress hormone levels. We confirm the existence of safety tipping points for in-hospital mortality using the discharge records of 82,280 patients across six high-mortality-risk conditions from 256 clinical departments of 83 German hospitals. Focusing on survival during the first seven days following admission, we estimate a mortality tipping point at an occupancy level of 92.5%. Among the 17% of patients in our sample who experienced occupancy above the tipping point during the first seven days of their hospital stay, high occupancy accounted for one in seven deaths. The existence of a safety tipping point has important implications for hospital management. First, flexible capacity expansion is more cost-effective for safety improvement than rigid capacity, as it will only be used when occupancy reaches the tipping point. In the context of our sample, flexible staffing saves more than 40% of the cost of a fully staffed capacity expansion, while achieving the same reduction in mortality. Second, reducing the variability of demand by pooling capacity in hospital clusters can greatly increase safety in a hospital system, as it reduces the likelihood that a patient experiences occupancy levels beyond the tipping point. Pooling the capacity of nearby hospitals in our sample reduces the number of deaths due to high occupancy by 34%.

Key words: Hospital mortality; occupancy; utilization; variability buffers; stress; pooling; flexible staffing
History: Revised version November 2013

1. Introduction

Avoidable deaths occur in hospitals as the direct or indirect consequence of avoidable adverse events, such as medication errors, infections, delayed treatments, or technical complications during operations. The scale of the problem was prominently highlighted by the influential Harvard Medical Practice Study in the early 1990s (Brennan et al. 1991, Leape et al. 1991), which estimated that 6,895 of the 2,671,863 patients hospitalized in New York State in 1984 had died in hospital as a consequence of preventable adverse events – more than three times the New York State traffic death toll in the same year (US Dept. of Transportation 2012). This study ignited a major global effort to improve hospital safety, much of which was focused on the prevention of individual human errors through process redesign and the use of technology (Kohn et al. 2000). For example, Bates et al. (1999) report an 81% reduction in medication errors through the introduction of computerized physician order entry systems. Leape et al. (1999) and Kucukarslan et al. (2003) report that the inclusion of pharmacists in ward rounds on ICUs and general medicine wards led to a reduction of 66% and 78% respectively, in preventable adverse drug reactions. Comprehensive surgical safety checklists reduced the total number of surgical complications in a Dutch hospital from 27.3% to 16.7% (Eefje et al. 2010). Despite such impressive improvements, advances appear to be isolated and system-wide progress in reducing avoidable death rates remains slow (Leape and Berwick 2005, Landrigan et al. 2010).

The mortality effects of intervention at the level of a hospital or a hospital system are less well understood than those of intervention at the process level, despite the fact that the former are less dependent on the specific organizational context and therefore more easily scalable and more likely to help achieve the desired system-wide progress. An example in point is hospital capacity pooling, where nearby hospitals agree to short-term admission diversion or staff relocation protocols in response to local occupancy surges. While pooling does not affect average hospital utilization in the system, it reduces the variability in utilization levels of individual hospitals. Does capacity pooling reduce avoidable death rates in the system? The answer to this question depends on the nature of the relationship between occupancy levels and mortality. If this relationship is linear, then avoidable deaths in the system are not affected by pooling: Any reduction in mortality by avoiding further occupancy increases in a busy hospital is offset by an increase in mortality due to increased occupancy in the hospital that admits the diverted patients. However, if the relationship is nonlinear, capacity pooling can have a significant effect on avoidable deaths. We will argue in this paper that bed occupancy has a highly nonlinear effect on mortality: Mortality remains unaffected by occupancy up to a tipping point, beyond which it deteriorates rapidly with further increased occupancy levels. As a consequence, capacity pooling has the potential to reduce avoidable death rates across a health system as patient diversion reduces the propensity of a hospital to exceed the safety tipping point, while the mortality in a less busy diversion hospital is not affected as long as its occupancy level remains below the tipping point.

In order to provide evidence for a safety tipping point, we have to overcome three methodological challenges. First, avoidable in-hospital deaths are rare events. One has to combine data from multiple hospitals and multiple patient segments to assemble a large sample of patients and obtain sufficient statistical power. In this study we use patient-level data from 83 acute hospitals across six patient segments with high mortality risk.

Second, occupancy studies usually consider occupancy aggregated at the corporate hospital level. However, patients experience occupancy at the level of individual wards. Operational decision in response to occupancy, such as the allocation of patients to specific wards, are mostly taken at the level of the clinical departments. Therefore, departmental occupancy is more relevant for patient care than aggregate hospital occupancy. Unfortunately, hospital departments, as managerial units, are not standardized and not recorded as part of US or UK administrative patient records, which form the basis of most empirical studies of hospital operations and it is therefore not possible to reconstruct department-level occupancy from such data. Departments are, however, recorded in a standardized way in the German discharge records that we use in this study.

The third methodological challenge is of a statistical nature: when occupancy is high, doctors may choose to discharge relatively healthy patients earlier than they would normally do to make space for newly arriving patients. They may well select patients for early discharge on health-related factors that are not recorded in the discharge record and thus not observable to the researcher. As a consequence, patients in the hospital during periods of high occupancy are more unwell in an unobserved way and the mortality risk among these patients is higher - not as a consequence of high occupancy per se but as a consequence of a changed risk set due to early discharge decisions in response to high occupancy. We account for this endogeneity in our econometric models.

Using a sample of 82,280 patients with high mortality risk, we estimate a tipping point at an occupancy level of 92.5%; 17.4% of these patients experienced occupancy levels above the tipping point. We estimate that 78 of the 4,247 deaths in our sample are due to high occupancy, and therefore avoidable, accounting for one in 55 deaths in the sample. However, 82.6% of patients in the sample never experienced occupancy above the tipping point. The effect on those patients who experienced occupancy above the tipping is therefore considerably larger: occupancy accounts for one in seven deaths (14.4%) among these patients. This is clinically highly significant.

We discuss potentially beneficial interventions in tipping point systems, specifically the effect of flexible staffing and capacity pooling. In our sample, flexibly staffed capacity increase turns out

to be 43% cheaper than rigid staffing for a commensurate mortality improvement. To estimate the effect of capacity pooling, we combine near-by hospitals in our sample to hospital clusters and estimate that 34.4% of the deaths due to occupancy in our sample could have been avoided if nearby hospitals had pooled their capacity.

2. Related Literature

The nature of the association between system load and service quality in general, and hospital occupancy and mortality in particular, is not yet well understood and the literature has hitherto produced inconsistent results. In an early paper, Oliva and Sterman (2001) built a dynamic model of a service organization to illustrate the complex interactions between service demand and managerial response and how these can lead to an erosion of service quality over time. More recently, a series of empirical studies have focused on operational efficiency and throughput in the context of hospitals. These studies acknowledge that a focus on throughput alone can have harmful consequences for clinical quality and therefore recommend also investigating effects on clinical outcomes such as mortality rates (KC and Terwiesch 2009, 2012, Long and Mathews 2013, Berry Jaeker and Tucker 2012, Kim et al. 2013). KC and Terwiesch (2009), for example, complemented a throughput analysis of a sample of cardiothoracic patient records from a US hospital with a study of the effect of workload on mortality. They found a significant effect of fatigue but could not identify a significant effect of bed occupancy on mortality.

In contrast to these throughput studies, this paper focuses on system load as a cause of quality deterioration and complements recent studies that consider workload effects at the level of individual workers. Powell et al. (2012) show that doctors' discharge coding behavior is affected by workload, with detrimental effect for hospital reimbursement. Green et al. (2012) show that absenteeism rates are correlated with anticipated future nurse workload. Drawing on the theory of stress, Tan and Netessine (2012) show that workload has a curvilinear effect on waiters' performance and illustrate that a reduction in staffing can in fact lead to an increase in revenues. In the same vein, Hopp et al. (2007) illustrate within a queuing model that when a server has discretion over service time in response to workload, increasing the number of servers may worsen congestion. We will incorporate these insights into the development of the tipping point hypothesis, which integrates the effects of excess capacity, managerial actions, and individual worker responses to explain the organization-level effect of variation in system utilization on service quality.

Several recent studies in the medical literature have identified a link between hospital activity levels and mortality (see Kane et al. (2007) for a review). Schilling et al. (2010) explored the effects of hospital occupancy levels on admission, annual nurse staffing levels, and seasonal factors on hospital mortality in a retrospective study of 166,920 emergency patients with high-risk conditions admitted to 39 Michigan hospitals between 2003 and 2006. The study found that admission on days when the hospital is in the top tertile of its occupancy range is associated with an elevated mortality risk. Needleman et al. (2011) studied the effect of below-target nurse staffing levels using 197,861 patient records from 43 clinical units of a US medical center and concluded that registered nurse staffing below target levels is associated with increased mortality. These studies model workload with a dichotomous "high/low" variable, which does not rule out a linear relationship between occupancy levels and mortality. Our study goes beyond these papers in that we estimate a continuous nonlinear occupancy model that supports the existence of a safety tipping point.

In summary, the main contribution of this paper is to point out that there are good operational reasons to expect the effect of occupancy on mortality to exhibit a threshold phenomenon: Occupancy has no discernable effect on mortality up to a tipping point, beyond which it affects mortality significantly. By not taking this phenomenon into account previous studies either could not detect an effect or overestimated effects at low occupancy levels and underestimated the severity of very high utilization. We provide evidence for the existence of safety tipping points and discuss managerial implications.

3. The Tipping Point Hypothesis

In most countries hospital planning focuses on bed capacity as the primary metric for sizing hospital departments; requisite staffing and other resources are largely calculated on a per-bed basis, using ratios that depend on departmental characteristics (Rechel et al. 2010). Consequently, as the volume of patients in a department approaches full bed capacity, workload pressure builds up across the unit as all resources become stretched. Capacity utilization – measured as the percentage of beds occupied – is therefore a useful aggregate measure of workload pressure in hospital departments.

Occupancy levels in acute hospitals show significant variation as demand for urgent care is unpredictable. Such variation is managed by drawing on variability buffers. The first and most obvious buffer is built-in excess capacity: Hospital plans are typically based on average bed occupancy levels in the order of 85% to 90%, thus providing a capacity buffer for demand peaks (Green 2004). A second class of buffers relates to managerial actions when occupancy levels rise. Managers can ask staff to work overtime, deploy flexible staff from elsewhere in the hospital or hire temporary staff from nursing banks or medical locum agencies. In 2004–2005 UK hospitals spent 9.4% of their nursing budget on temporary nursing staff (Department of Health (UK) 2006). In addition, hospitals can manage demand by canceling scheduled elective cases at short notice in response to unexpected surges in emergency admissions. UK hospitals, for example, cancel between 0.7% and 1.0% of elective patient admissions at the last minute for reasons unrelated to the circumstances of the scheduled patient (Department of Health (UK) 2012). The third class of buffers relates to responses by front-line staff. Doctors and nurses are willing to work harder and for longer in times of crisis (Scott et al. 2006). In fact, many nurses choose their profession based on an intrinsic motivation to care for people in need. Doctors undergo a substantial socialization process during their long professional training (Laine and Davidoff 1996). The values and professional norms of healthcare workers instil a strong motivation and willingness to "go the extra mile", which provides an important human variability buffer.

As workload increases, healthcare professionals are forced to ration access to care and, in doing so, will give priority to sicker patients. KC and Terwiesch (2012) and Long and Mathews (2013) provide evidence of active rationing from busy intensive care units (see also Berk and Moinzadeh (1998) and Padma et al. (2004)). While rationing can have a negative effect on the less sick, the ability to prioritize is an important variability buffer, helping to shelter the most critically ill. All these variability buffers are drawn on simultaneously as system load increases and allow a hospital department to cope with a wide variation in occupancy levels while safeguarding the most critical aspect of clinical care: the avoidance of death.

However, as occupancy levels continue to rise the organization's variability buffers become depleted; all beds are filled and additional patients need to "board" in other departments, no more elective patients can be canceled at short-notice, and qualified agency staff are scarce or resources to hire them are limited. Yet demand for hospital care is at times unrelenting. Acute care hospitals cannot turn emergency patients away. When the variability buffers are depleted, resources need to be rationed more aggressively; doctors and nurses begin to cut corners even for more seriously ill patients, using service quality as an implicit variability buffer (Oliva and Sterman 2001, Hopp et al. 2007).

In addition to cutting corners as a conscious response to excessive workload, doctors and nurses are exposed to workload-related stress, which causes their performance to deteriorate. Lazarus and Folkman (1984) point out that stress results from an "imbalance between demands and resources" and occurs when "pressure exceeds one's perceived ability to cope"; this is precisely the case when workload becomes excessive and the ability to cope by exploiting buffers reaches its limits. This effect was pointed out by Piquette and Reeves (2009), who observed that, in the context of critical care, "individual distress occurred in unexpectedly high demands unmatched by appropriate resources." At the biological level, workload stress leads to elevated stress hormone levels, specifically those of cortisol (Dickerson and Kemeny 2004, Sonnentag and Fritz 2006), which impairs

workers' cognitive abilities, especially memory and attention, and the quality of their decision-making (Lupien et al. 2007). In addition, teamwork deteriorates: Piquette and Reeves (2009) observed that "emotional distress was strongly contagious to other team members, [...] disruptive for teamwork and deleterious for individual and collective performance." The consequent negative impact of stress on clinical outcomes is well documented in the medical literature. Dugan et al. (1996) show for example that nursing-related stress is strongly associated with the propensity of adverse incidents; Buckley et al. (1997) show that haste and stress were causative factors in 17% of 281 critical incidents.

In summary, as system utilization increases to moderately high levels, managers respond by exploiting resource buffers and well-motivated employees work harder. Quality of care can largely be maintained and safety is not negatively affected. However, at very high utilization levels, variability buffers are depleted and managerial response is inhibited. If utilization exceeds this critical tipping point, managers are unable to respond. The pressure is passed on to front-line staff, who are unable to escape it. They then respond in two ways: First, by consciously cutting corners, using quality as an implicit variability buffer; and second, by subconsciously committing more errors as a result of elevated stress hormone levels. As a consequence, quality of care and safety will deteriorate during periods of high utilization. The following empirical study provides evidence for this tipping point phenomenon.

4. Empirical Study

4.1. Data from German Hospital Departments

We use data from German hospitals, which are particularly suited for a multi-hospital department-level analysis for two reasons. First, the organizational structure of German hospitals is firmly regulated, leading to rigid and fairly homogeneous departmental organization across hospitals. The structural similarity begins at the top: Almost all German hospitals have the same top management team structure, consisting of a commercial director, a medical director, and a nursing director, each with well-defined roles and obligations across the hospital. The departmental structure below the top team is also standardized, including general services, such as kitchen and laundry, large diagnostic divisions, such as radiology and pathology, and the clinical departments, including general surgery and general medicine, as well as specialist departments. These bed-bearing clinical departments are the focus of our study and system utilization is measured at the level of these organizational units. Importantly, every department has a clinical director – the *Chefarzt* – who, as lead physician, has ultimate clinical responsibility for all patients and is the superior of all doctors in the department. The clinical director also has budgetary responsibility for her department. Although there is a cautious trend toward the use of interdisciplinary beds, the system remains rigid and the vast majority of patients and resources are managed at the level of these clinical departments. In particular, any responses to occupancy variations are most likely to be managed at departmental rather than hospital level.

Second, in contrast to the US and UK, German hospital discharge records contain standardized department codes, including departmental referrals during a hospital episode. Earlier multi-hospital mortality studies concentrated on the effect of aggregate hospital occupancy (e.g. Schilling et al. (2010)). In contrast, we measure daily occupancy at the level of departments within hospitals. This is important because managers and clinicians are most likely to respond to occupancy levels in their department. Aggregate hospital occupancy measures the departmental load that is relevant for a particular patient with significant error, which leads to attenuation bias and underestimated effect sizes (Wooldridge 2002).

Our initial database consists of standardized discharge records of 101 German hospitals, covering all patients discharged from the hospital during a specific period. The database contains data for 12 months for 72 hospitals, covering 1 January to 31 December of either 2004 or 2005, while 29 hospitals are observed for 24 months, from 1 January 2004 to 31 December 2005, totalling 1,415,754 patient episode records across 624 hospital departments. Since high occupancy is unlikely to impact the mortality risk of relatively healthy patients, we select a subsample of patients with high mortality risk and focus on patients with six primary diagnoses identified by the US Department of Health as conditions "*for which mortality has been shown to vary substantially across institutions and for which evidence suggests that high mortality may be associated with deficiencies in the quality of care*" (Agency for Healthcare Research and Quality 2006): Acute myocardial infarction (AMI), congestive heart failure (CHF), gastrointestinal hemorrhage (GIH), hip replacement after fracture (HIP), pneumonia (PNE) and stroke (STR). To increase the homogeneity of the sample, we remove all hospitals that do not have emergency admissions, such as rehabilitation clinics, as such hospitals are unlikely to have critically ill patients. For departments with a small volume of these high-risk patients, mortality is rare and department fixed effects together with patient covariates can predict survival perfectly. This leads to numerical instability of the maximum likelihood optimization procedure. We therefore exclude all departments for which the department is a perfect predictor of patient survival.

The fact that the entire patient population for each hospital department is included allows us to calculate daily midnight patient counts for each hospital department from 1 January. However, at the end of the observation period we do not have data about patients who are admitted but not discharged during the observation period. As a consequence, calculated occupancy rates start dropping during the final month of a hospital's observation period. We therefore restrict our sample to patients admitted between 1 January and 30 November of a hospital's observation period. The removal of patients who were admitted during the final month of hospital observation is prudent in light of an average length of stay of 12 days for our chosen subsample. The remaining sample consists of 82,280 patients in 256 departments of 83 hospitals.

It would be preferable to conduct analyses at the level of single hospital departments and single conditions. However, such subsamples have insufficient size to lead to statistically significant results. Although mortality is relatively high amongst the conditions in our sample, the event of interest – *avoidable* death due to high occupancy – is rare. Estimates of rates of avoidable deaths range widely in the medical literature, depending on the context; Gruen et al. (2006) studied a sample of trauma deaths and estimated that 2.64% were avoidable; Healey et al. (2002) investigated deaths in several surgical departments and found that between 19.0% and 44.1% were avoidable.

A sample size estimation reveals the magnitude of the associated statistical challenge. The simplest specification is a logistic regression model of the form $\text{logit}(Y_i) = \alpha + \beta X_i + \gamma Z_i$, where Y_i is a dichotomous variable indicating the death of patient i, X_i is the occupancy covariate, and Z_i is a vector of control covariates. For ease of interpretation, we assume that X_i is a dichotomous variable with value 1 if patient i experienced occupancy levels above a threshold beyond which we believe mortality will be affected. The logistic sample size formula of Hsieh et al. (1998) can then be used for a power analysis for a significant two-tailed Wald test for the null hypothesis $H_0 : \beta = 0$. The required sample sizes for 80% power at a 5% significance level depends on the unknown effect size, i.e. the excess mortality of those patients who experience occupancy levels above the tipping point. If a patient population mortality of 5% rises to 5.5% for the subpopulation above the tipping point and if 15% of patients experience such high occupancy levels, then the required sample size exceeds 85,000. The required sample size increases further if the occupancy variable X_i and the controls Z_i are correlated (Hsieh et al. 1998) or if the population mortality is lower than 5%. The required sample size decreases if the effect size is larger or if more patients are exposed to occupancy levels above the tipping point. In conclusion, mortality studies of the type we conduct

in this paper require large samples of high-risk patients. Since only a relatively small proportion of any single hospital's patients will have a sufficiently high mortality risk, we have to combine data from multiple hospitals and multiple patient segments.

4.2. The Need for a Survival Analysis

Figure 1 gives an initial indication of the tipping point on the basis of raw data and illustrates the importance of accounting for time already spent in the hospital in the statistical analysis. The figure is based on a patient-day data set, where each observation corresponds to a record of one full in-patient day for a particular patient, including an appropriate measure of occupancy experienced by the patient up to this observation day. Figure 1 is based on peak occupancy, defined as the maximum of all midnight occupancy levels experienced by the patient up to the beginning of the observation day. We will discuss occupancy measures in more detail in Section 4.3. The left-

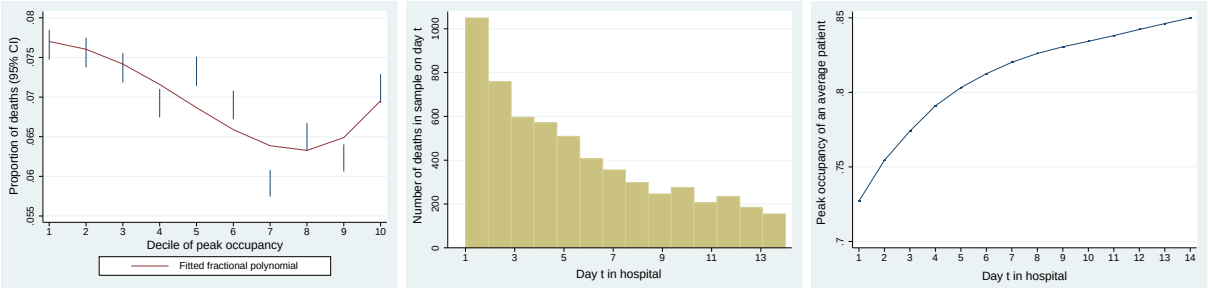

Figure 1 Uncontrolled association between day of stay, peak occupancy, and mortality

hand graph of Figure 1 shows how mortality varies across the deciles of peak occupancy, initially decreasing but with a surprisingly marked increase at the 10th decile. The fact that mortality decreases with peak occupancy is to be expected: First, as the middle graph in Figure 1 shows, the mortality risk decreases with the length of stay as most deaths occur during the first few days of the stay, when a patient is most critically ill. Second, as illustrated by the right-hand graph, peak occupancy increases over time: The longer a patient stays in hospital the more likely it is that she will be exposed to high occupancy levels at some point during her stay. The combination of decreasing mortality and increasing peak occupancy with time results in a decrease in mortality with peak occupancy, as indicated in the left-hand graph up to the ninth decile of peak occupancy. This expected trend, however, is starkly reversed at the tenth decile, in accordance with the tipping point hypothesis.

The simplest mortality model focuses on patient episodes as the units of observation, survival as a dichotomous episode outcome, and an appropriate aggregate measure of occupancy during the patient episode as the independent variable of interest (e.g. KC and Terwiesch (2009)). This model is problematic in our context as it requires the aggregation of occupancy over a patient's episode. The natural aggregate occupancy metric is the average occupancy over the patient's stay. The standard deviation of this metric, however, depends on the length of stay of the patient: The fewer days the patient spends in the hospital, the larger the variation of the average occupancy over these days will be. Therefore the distribution of the episode-averaged occupancy metric across the patients in the sample will contain an over-proportional number of patients with short length of stay in both tails. At the same time, mortality is higher for patients with a short length of stay for two reasons. First, death curtails length of stay and, second, as shown in Figure 1, patients are most likely to die early on in their stay. Combining these two facts – the higher frequency of patients with short length of stay in both tails of the distribution of average occupancy and the higher mortality amongst patients with a shorter length of stay – leads to a non-linear effect of occupancy

on mortality. This nonlinear effect, however, is a consequence of the aggregation mechanism – the averaging of occupancy – not of occupancy per se and would lead to the identification of a spurious tipping point. Note that one cannot use length of stay as a control variable in a patient-level model of mortality because mortality curtails length of stay, which leads to reverse causality. It is important to use a model that incorporates the time-varying nature of occupancy and survival analysis is therefore the natural modeling choice. We have chosen a discrete-time survival model based on patient-days as units of observation, which takes account of the effect of time spent in the hospital and treats occupancy as a time-varying covariate. We explain the survival model in more detail in Section 5.

The dependent variable of interest in the discrete survival model is a patient's probability of death during a full day in hospital; the causal variable of interest is peak midnight occupancy experienced by the patient prior to the observation day. It is important to note that the admission day is not a full day in hospital. A patient who arrives in hospital with a stroke at 6pm is, ceteris paribus, less likely to die on the admission day than on the first day of her hospital stay because she spends only six hours in hospital on that first day. In fact, the time spent in hospital on the admission day varies by patient. This makes the inclusion of the admission day problematic because we cannot estimate the probability of death on a full day for these patients. In addition to this methodological concern with the inclusion of the admissions day, there is also a contextual concern. Death on the day of admission is less likely to be caused by the conditions in the admitting department and more likely by the conditions in the emergency department or operating theaters if an emergency operation is necessary. We do not have information about these conditions. In view of these methodological and contextual concerns, we discard the patient's admission day, beginning patient observation at midnight after admission, and study the adverse effect of departmental occupancy on the population of patients who survive their admission day.

Occupancy may affect patients differently at different stages of their hospital stay. Specifically, patients are likely to be more critically ill during the early phase while they recover during the later phase of their stay. Convalescent patients are not as care-intensive with regard to monitoring requirements and will be less vulnerable to deviations from optimal care. However, when patients stay considerably longer than expected, this is an indication that they are more severely ill. Since peak occupancy up to the t-th day of the hospital stay is a monotone function of t, this change in the risk profile of the remaining patients at high values of t could potentially cause the "Bathtub"-curve of mortality as a function of peak occupancy, as observed in the left-most graph in Figure 1. To rule out this cause and increase the homogeneity of the patient-day sample, we restrict our survival analysis to the first week of a patient's hospital stay, which, as shown in the death rate histogram in Figure 1, is the most critical period of a patient's stay. The first seven days is a prudent estimate of this critical phase, in light of an average length of stay of 12 days.

All admitted patients are observed daily at midnight over the first seven days of their stay or up to their death or discharge if this occurs before the end of the seventh day in hospital. After the seventh day, patients are not observed further. Such prescribed follow-up periods are common in clinical and epidemiological studies and are known as administrative or type I censoring (Klein and Moeschberger 1997). Table 1 contains summary statistics of the patient-day sample.

4.3. Occupancy as a Time-varying Covariate

As occupancy refers to the percentage of used capacity, we need to first measure capacity for hospital departments. The natural measure is the number of beds in operation; however, this number is rarely available as public documents refer to the number of *certified* hospital beds. Interviews with hospital managers revealed that this number can deviate significantly from the number of beds in operation that are fully resourced and readily available for patients, and is therefore not a reliable measure of operational capacity. In addition, certified bed numbers, while

Table 1 Descriptive statistics of sample

Condition	Patients	Percentage of full sample	Emergency admissions	Age (mean)	Mortality	Length of stay (days)	7-day mortality	7-day discharges	7-day patient-days
AMI	12,811	15.6%	53.1%	68.1	9.4%	10.0	5.9%	39.8%	69,985
CHF	17,852	21.7%	44.5%	74.9	9.0%	11.4	4.8%	29.7%	109,089
GIH	9,029	11.0%	48.5%	64.9	5.2%	9.2	2.8%	50.9%	48,128
HIP	7,974	9.7%	66.8%	73.5	5.3%	17.0	2.4%	10.9%	52,795
PNE	14,610	17.8%	50.5%	63.8	11.5%	11.7	6.2%	27.2%	91,179
STR	20,004	24.3%	61.9%	69.7	10.5%	13.5	6.5%	20.6%	126,404
Full sample	82,280	100.00%	53.7%	69.4	9.1%	12.1	5.2%	29.1%	497,580

available in aggregate for hospitals, are not available at the department level, where we wish to measure occupancy. In the absence of reliable operational bed numbers, we therefore use the maximal daily midnight patient count in the department over the department's observation period as a measure of the department's capacity. For each day of the patient's stay we then calculate the daily capacity utilization as the ratio of the midnight patient count at the beginning of the day and the department's capacity.

In the survival analysis framework, where we observe patients daily, occupancy is a time-varying covariate. A critical question is how exposure to varying occupancy levels over time should be measured for an individual patient. The midnight patient count at the beginning of day t would appear to be a natural candidate to affect mortality on day t. However, with this occupancy metric we would only capture the immediate effect on the observation day; lagged effects, two or three days hence, would be discarded. In view of the rareness of avoidable deaths, it is unlikely that we will have enough power in our data to detect this immediate occupancy effect. A second candidate is the average occupancy experienced up to the beginning of day t. However, the use of this metric is problematic within a tipping point model and is likely to lead to the spurious detection of a nonlinear effect discussed in Section 4.2. The time dummies in the survival model do not control for this effect because they only affect the intercept of the occupancy curve but not its shape. We would therefore estimate a nonlinear effect for time-averaged occupancy, even when occupancy has no effect.

We choose the *maximal* midnight occupancy level up to the beginning of day t as the measure of the occupancy experienced to date by a patient in the hospital on day t. This *peak occupancy* metric has the advantage of being monotonically increasing over the stay in the hospital, which captures an important lag effect: Exposure to high occupancy on day t can lead to death at a later day and cannot be "undone" by low occupancy after day t. This monotonicity property of the time-varying exposure also introduces a positive correlation with time, which is itself negatively correlated with mortality. Since the resulting correlation between peak occupancy and mortality is negative, the detection of a positive effect of peak occupancy on mortality beyond a tipping point will only become more difficult, which renders significant estimates conservative.

4.4. Control Variables

The need for risk-adjustment of patient-level data is comprehensively discussed in the literature (Iezzoni 2003). Our discharge records contain several variables that allow us to control for patient heterogeneity. Beside the primary medical condition and the individual risk factors age, gender, and emergency admission, the presence of secondary diagnoses is an important source of heterogeneity. To account for these comorbidities we follow Needleman et al. (2011) and use indicator variables for a list of coexisting conditions (Elixhauser et al. 1998), adapted to the German context following Quan et al. (2005). In addition, we control for admission from another hospital with a dichotomous variable and for departmental transfers within the hospital prior to the observation day with a

time-varying exposure dummy that takes the value 1 on all days following the first departmental transfer within the hospital.

We include day-of-stay dummy variables to model the baseline mortality hazard over the patient stay. To account for differential baseline risk across the six conditions, we interact the day-of-stay with the primary condition. Seasonal effects must also be controlled for as there might be times of the year when certain conditions occur more frequently or in a more severe way, e.g. through the winter months, and when occupancy in hospitals is also higher. Time-of-year can therefore confound results. To control for potential temporal correlations, we include dummy variables for the month of the year and for the observation year 2005. We also control for the weekday of the admission to account for the so-called weekend-effect discussed in the medical literature (e.g. Bell and Redelmeier (2001)): Patients who are admitted on weekends have a higher mortality risk relative to weekday admissions, even after controlling for their individual risk factors. Finally, we control for the weekday of the observation day.

We use department dummy variables to control for organizational heterogeneity in an aggregate way, as departments will have differences in case-mix, size, and staff endowment. We use department rather than hospital fixed effects because departments are fairly autonomous units in our context, as explained in Section 4.1. To account for potential correlations between the error terms of patients in the same department we cluster standard errors at the department level. The table in the Appendix summarizes the control variables included in the models and shows their correlations with daily mortality and peak occupancy.

5. Econometric Specification

We wish to estimate the association between the occupancy levels that patients experience during their hospital stay and the probability of in-hospital survival. As occupancy levels are most reliably calculated on the basis of midnight counts, a discrete-time survival analysis using patient-day observations is a natural modeling framework for this purpose. The population of interest in this study consists of patients who are admitted to hospital with one of the six high-risk conditions discussed in Section 4.1, and who survive until midnight on their day of admission to the hospital. By beginning our observation at midnight following admission we ensure that all observation periods have equal length. We follow up patients for seven days after admission and wish to estimate patient i's discrete mortality hazard on day t after admission

$$h_{it} = P[T_i = t \mid T_i > t-1, X_{it}], t = 1, \ldots, 7, \tag{1}$$

where T_i denotes the day of death of patient i, counted from the day of admission, and X_{it} is a covariate vector that is observable at the beginning of day t. The time-varying covariate vector X_{it} includes dummy variables for each period t, which captures a baseline hazard model as a time-dependent intercept, as well as a component X_{jit} for peak occupancy experienced by patient i up to the beginning of day t. The most common logit, probit, and cloglog specifications for the discrete time hazards (Singer and Willett 2003) give very similar results. We report results for the probit specification

$$P[Y_{it} = 1 \mid X_{it}] = \Phi(X_{it}\beta), \tag{2}$$

where Y_{it} is the observed dichotomous mortality variable, taking the value 1 if patient i dies on day t and 0 otherwise, and Φ is the standard normal cumulative distribution function.

5.1. The Tipping Point Model

We use a piecewise linear specification to estimate a potential tipping point with respect to occupancy. Specifically, assuming the peak occupancy that patient i experienced during their stay up

to day t is stored in the jth component X_{jit} of the covariate vector X_{it}, we use the parametric specification

$$\beta_{j1} X_{jit} + \beta_{j2} \max\{X_{jit} - \beta_{j3}, 0\} \tag{3}$$

to model the tipping point β_{j3}. Here, β_{j1} is the slope of the line to the left of the tipping point and β_{j2} captures the change in slope of the line as X_{jit} exceeds the tipping point β_{j3}. We estimate all three parameters β_{j1}, β_{j2}, and β_{j3}.

The tipping point model is parsimonious with the minimum number of required parameters – one for the tipping point, and one for the behavior of the function on either side of the tipping point – and has several advantages over the more common polynomial specification of a nonlinear effect: First, it treats the tipping point explicitly as a parameter, which will allow us to estimate confidence intervals for the tipping point; second, its estimates have an immediate interpretation; and third, the shape of the piecewise linear function can be asymmetric, with different slopes on either side of the tipping point. In contrast, polynomial models, with maxima and minima as candidates for tipping points, exhibit symmetric second-order behavior and therefore symmetric shapes in the vicinity of these tipping points. The disadvantage of the piecewise linear model is that the term $\beta_{j2} \max\{X_{jit} - \beta_{j3}, 0\}\}$ in (3) renders the probit maximum likelihood problem non-concave. Fortunately, concavity is restored once the tipping point β_{j3} has been fixed in (3). In order to optimize the likelihood function, we first estimated the remaining parameters repeatedly for a range of tipping points β_{j3} and then used a procedure suggested by Muggeo (2003) to check optimality and estimate the standard error of the tipping point estimate.

5.2. Average Partial Effects

In view of the difficulty of interpreting coefficient estimates in generalized linear models, it has become customary to base statistical inference on average partial effects (APE) (Wooldridge 2002). Given probit estimates $\hat{\beta}$ based on (2), individual patient-day-level partial effect estimates

$$\nabla_X P[Y_{it} = 1 | X_{it}] = \phi(X_{it}' \hat{\beta}) \hat{\beta}$$

are aggregated to average partial effects

$$\mathrm{APE}(\hat{\beta}) = \frac{1}{N} \sum_{(i,t)} \phi(X_{it}' \hat{\beta}) \hat{\beta}, \tag{4}$$

where $\phi(z) = \Phi'(z)$ is the standard normal density and the sum is taken over all N patient-days (i,t) in the sample. The APE has a natural population-based interpretation as the proportional effect of a unit increase in a covariate across *all* patient-days (Wooldridge 2002). The asymptotic variance-covariance matrix of the APE can be obtained via the delta method and is of the form $M \hat{V} M'$, where $\hat{V}$ is the variance-covariance matrix of the probit coefficient estimates,

$$M = \frac{1}{N} \sum_{(i,t)} \phi(X_{it}' \hat{\beta}) [I - X_{it}' \hat{\beta} \hat{\beta}' X_{it}], \tag{5}$$

and I is the identity matrix (see Chapter 2.6.6. of Green and Hensher (2010)). In the context of the tipping point model, we are interested in the average partial effect of peak occupancy below and above the tipping point. In order to compute these average partial effects, we have to average over the appropriate subsample of patient-days with peak occupancy above and below the tipping point, rather than over the entire sample. Note, however, that model (3) does not provide a direct estimate of the slope above the tipping point β_{j3}; instead this slope is the sum of the two correlated estimates β_{j1} and β_{j2}. Furthermore the slope β_{j1} applies to occupancy below and above the tipping

point. We therefore re-estimate the model for the already optimized tipping point β_{j3} using the following reparametrization of (3):

$$\tilde{\beta}_{j1} \min\{x_{jit}, \beta_{j3}\} + \tilde{\beta}_{j2} \max\{x_{jit} - \beta_{j3}, 0\}. \tag{6}$$

In this model $\tilde{\beta}_{j1} = \beta_{j1}$ estimates the slope below the tipping point β_{j3}, while $\tilde{\beta}_{j2} = \beta_{j2} + \beta_{j1}$ estimates the slope above the tipping point. We cannot use this parametrization of the tipping point model for the estimation of the tipping point as Muggeo (2003)'s standard error estimation for the tipping point does not apply to this model. However, once the tipping point has been estimated on the basis of (3), we can fix it and re-estimate the remaining parameters using the new parametrization (6). We then calculate average partial effects associated with $\tilde{\beta}_{j1}$ and $\tilde{\beta}_{j2}$ via (4) and (5) by averaging over the relevant subsamples of patient-days with peak occupancy below or above the tipping point.

5.3. Discharge as a Competing Risk

Observations of patients are censored in two ways in our model. First, we follow patients only during the first week of their stay in hospital and discard observations beyond the first week. This so-called type I censoring is non-informative, i.e. censoring of a patient provides no information about the survival probability of this patient beyond the censoring time (Klein and Moeschberger 1997). Standard survival analysis methods allow for such uninformative right-censoring. However, in our context there is another form of censoring, namely discharge from hospital. This censoring mechanism is informative; knowing that a patient has been discharged home provides information about her health status and thus her survival prospects. In order to treat discharge properly as informative censoring, we use a competing risk model, with discharge as the competing risk, and estimate the subdistribution hazard, a concept introduced by Fine and Gray (1999), which has found widespread applications in biostatistics and epidemiology (Lau et al. 2009). Rather than censoring discharged patients on the day of discharge, the Fine-Gray approach maintains the records of the discharged patients in the data beyond the time of discharge. This is operationalized by duplicating the patient's records on the day of discharge and censoring the patient at the end of the seven-day follow-up period. Estimation is achieved by applying model (2) to the expanded data set.

The expansion of the risk set on day t by all patients who were discharged prior to day t changes the hazard definition. The standard mortality hazard – the probability that a patient survives day t in the hospital, conditional on having survived up to the end of day $(t-1)$ – is estimated by $h(t) = \frac{N_t}{R_t}$, where N_t is the number of patients in the sample who die on day t of their hospital stay and R_t is the number of patients in the sample who are still in the hospital at the beginning of day t of their stay. If we keep the records of discharged patients in the data set after their discharge date, as proposed, we enlarge the risk set on day t and estimate $h_s(t) = \frac{N_t}{R_t + D_t}$ instead, where D_t is the number of patients who were discharged prior to the t-th day of their stay. This quantity estimates the Fine-Gray subdistribution hazard, which is the probability of dying on day t given that either (a) the patient is still in hospital at the beginning of day t *or* (b) the patient has been discharged prior to day t (Lau et al. 2009). The fact that the discharged patients are maintained in the risk set is equivalent to assuming that all patients discharged before day seven of their hospital stay would have survived in the hospital up to the end of the follow-up period of seven days had they not been discharged, and would not have been exposed to further variation in occupancy levels. Using the subdistribution hazard in our context therefore leads to a conservative estimate of the mortality risk, accounting for the unknown correlation between discharge risk and mortality risk. We refer the reader to the original paper Fine and Gray (1999) and the survey article Lau et al. (2009) for more details on the method. We explain and analyze an alternative to the Fine-Gray approach in Section 7.1 as part of our robustness checks.

6. Results

We estimated the discrete subdistribution model (2) using the *probit* command in STATA, Version 13. Table 2 shows the estimation results for the variables of interest. The Linear Model includes peak

Table 2 Seven-day in-hospital survival estimates for peak occupancy before observation day

	Linear Model		Tipping Point Model	
	Coefficient estimate	Average partial effect	Coefficient estimate	Average partial effect
Peak occupancy	0.127* (0.0561)	0.0024* (0.0011)		
Tipping point			0.925 (0.0184)	
Peak occupancy below tipping point			0.0374 (0.0647)	0.0007 (0.0012)
Peak occupancy above tipping point			1.772*** (0.442)	0.0337*** (0.0091)
Patient days	557,828		557,828	
Number of parameters	366		368	
Log-liklihood	-22,143.46		-22,137.26	
p-value of LR test			0.002	

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, robust standard errors in parentheses, clustered by hospital departments

occupancy up to the observation day as a linear covariate. The estimated coefficient (beta=0.127, $p = 0.024$) is statistically significant at the 5% level. The slope estimates in the Tipping Point Model refer to the parametrization (6). The tipping point and its standard error were estimated using the alternative parametrization (3), as proposed in Muggeo (2003). We first estimated the model for fixed tipping points with a 1% spacing at 85%, 86%,..., 98%, resulting in a maximal likelihood at 92%, and then used the procedure suggested by Muggeo (2003) to further optimize locally in the vicinity of 92% and to calculate the standard error of the final estimate. The tipping point was estimated at 92.5%, with a 95% confidence interval of [88.9%, 96.0%]. In contrast to the linear model, the tipping point model renders peak occupancy below the tipping point statistically insignificant (beta=0.0374, $p = 0.56$). The effect of peak occupancy above the tipping point, however, is statistically highly significant (beta=1.772, $p < 0.001$). The overall model fit is significantly improved relative to the linear model (deviance = 12.4, $p = 0.002$). These estimations support the tipping point hypothesis.

6.1. Effect Size Estimations

Of the 557,828 patient-days in our sample, 71,510 (12.8%) were associated with historical peak occupancy above the estimated tipping point. The average historical peak occupancy on these days was 95.8% and the average daily mortality rate on these days was 0.00757, i.e. on average 757 of 100,000 patients who had experienced occupancy above the tipping point in the past died each day. Table 2 reports the average partial effects for the Tipping Point model, as explained in Section 5.2. The average partial effect estimate of peak occupancy above the tipping point is 0.0337 with a 95% confidence interval of $[0.016, 0.052]$. This suggests that a reduction of peak occupancy *above* the tipping point by one percentage point, reducing average occupancy from 95.8% to 94.8%, would reduce the daily mortality rate on these days by 0.000337 (95% CI=$[0.00016, 0.00052]$). On aggregate, the daily loss of 757 patients per 100,000 patients would be reduced by 39 patients (95% CI=$[16, 52]$): a reduction of the death rate by 6.2% (95% CI=$[2.9\%, 9.5\%]$). This is a clinically

significant effect associated with a modest one percentage point reduction in peak occupancy above the tipping point. A comparison of the average partial effects in Table 2 shows that the linear model overestimates the effect of a reduction of occupancy below the tipping point and severely underestimates the effect above the tipping point.

The tipping point model suggests an alternative estimation of the total number of avoidable deaths in the sample: Our model allows us to predict the mortality hazard on each observed patient-day for varying levels of peak occupancy. For each patient-day with a peak occupancy value above the estimated tipping point level of 92.5% we can compare the predicted mortality hazard h_{it} (1) with the predicted hazard h'_{it} if peak occupancy is reduced to the tipping point level. Summing up the differences between these two hazards over all patient-days in the sample provides an estimate of lives saved if no patient had experienced peak occupancy above the tipping point. For our sample this results in 78 lives saved out of 4,247 deaths: an overall reduction in mortality by 1.8%. In other words, one in 55 deaths in the sample is accounted for by occupancy above the tipping point.

Since peak occupancy can only have an effect on *avoidable* deaths, it is useful to relate this estimate to avoidable death estimates in the medical literature. Such estimates vary significantly, depending on the context(see Lessing et al. (2010) for a review). Combining an extreme estimate of a 44.1% avoidable death rate (Healey et al. 2002) with the estimated one in 55 deaths due to high occupancy results in a combined estimate of 4.2% avoidable deaths due to high occupancy. Combining our estimate with a more realistic assumption that one in 10 deaths are avoidable in our high-risk sample would suggest that 18% of these avoidable deaths are associated with peak occupancy above the safety tipping point.

So far, we have only discussed the size of the effect of high occupancy on *all* patients. However, 82.6% of patients in our sample did not experience occupancy above the tipping point. If we relate the estimated 78 saved lives to the 541 deaths among the 14,321 patients who experienced occupancy levels above the tipping point, the size of the effect becomes substantially larger: 14.4% of deaths among these patient could have been avoided if no patient had been exposed to occupancy above the tipping point.

7. Robustness Checks

7.1. Discharge as a Selection Problem

Patients are discharged as their health status improves and this discharge trigger may well be affected by occupancy levels. When occupancy is high, doctors make space for new patients by discharging some patients earlier (KC and Terwiesch 2012). There is some evidence that doctors choose less ill patients for early discharge (Long and Mathews 2013). We have dealt with discharge as a competing risk, based on the subdistribution approach of Fine and Gray (1999). In this section we analyze an alternative way of dealing with discharge by extending the probit model (2) to account for endogenous discharge decisions in a bivariate probit model with selection (Green 2003). As in the standard probit model, we assume that death occurs on day t when a latent sickness status Y^*_{it} becomes positive. Analogously, discharge on day t is triggered by a second latent variable S^*_{it}, which is interpreted as the difference between the clinician's utilities from keeping patient i in hospital beyond day t or discharging her on day t. If $S^*_{it} > 0$, then the patient is kept in hospital beyond day t, otherwise she is discharged on day t. Both latent variables are assumed to depend on individual covariates and random errors

$$\begin{aligned} Y^*_{it} &= X_{it}\beta + \epsilon_{it} \\ S^*_{it} &= Z_{it}\gamma + \nu_{it}. \end{aligned} \tag{7}$$

As in the probit model (2), both error terms are assumed to follow standard normal distributions. However, some of the unobserved factors that explain the latent sickness status may also affect the discharge decision. We therefore allow for correlated errors and assume them to be sampled

from a bivariate normal distribution with a correlation coefficient ρ that needs to be estimated. By allowing for correlated errors in the simultaneous equations, we lift relevant unobserved information from the discharge equation to the mortality equation.

In a standard bivariate probit model, one observes each of the four combinations of the two events, discharge and death. In our context, however, we do not observe the combined outcome – death and discharge – but only one of three event combinations:

1. patient i was discharged on day t
2. patient i was not discharged on day t and died on day t
3. patient i was not discharged on day t and survived day t.

This leads to a bivariate probit model with selection (see Chapter 21.6.4 of Green (2003)). This model allows us to test for confounding by discharge in our sample by testing whether ρ is significantly different from zero.

Although the bivariate model is, in principle, identified by its bivariate normality assumption, robust identification benefits greatly from covariates that satisfy an exclusion restriction (Wooldridge 2002). Such variables are significant predictors in the discharge equation but are excluded from the mortality equation in the sense that they have vanishing coefficients in the population model of this equation. The former condition is testable, while the latter is an untestable assumption. We make use of two covariates for which we believe an exclusion restriction is plausible. The first covariate is a dichotomous variable that takes the value 1 if the observation day is a Sunday. Patients are unlikely to be discharged on Sundays because administrative staff is unavailable and staffing is generally lower and focused on clinical care. It is plausible to assume that, after controlling for the control variables, patients are not more or less likely to die on Sundays than on weekdays.

The second variable with an exclusion restriction is the discharge rate of the other patients (i.e. excluding the observed patient) on the day of observation. This variable captures the general discharge behavior of clinicians and should be related to the observed patient's discharge probability. We believe it is plausible to assume that this variable satisfies an exclusion restriction for the outcome equation, i.e. that an individual patient's probability of dying on day t does not depend on the rate at which other patients in the department are discharged, after controlling for the other variables in the mortality equation. To make discharge comparable between departments, we calculated the z-score of the number of discharges for each day in each department by normalizing daily discharge numbers with respect to the average number of discharged patients and the associated standard deviation in the department during the observation period for the department. Formally, we calculate the following variable for patient i in department j on day t of her stay:

$$z_{ijt} = \frac{d_{ijt} - (\bar{d}_j - 1)}{\sigma_j}, \tag{8}$$

where d_{ijt} is the number of other patients (excluding patient i) discharged on day t of patient i's stay in department j, and $\bar{d}_j$ and σ_j are the average and standard deviation of the daily discharge numbers in department j during the observation period. We use $\bar{d}_j - 1$ instead of $\bar{d}_j$ in the numerator because the variable relates to all patients except patient i.

We estimated the bivariate model using the *heckprob* command in STATA, Version 13. The occupancy tipping point estimate of 92.4% and the slope estimates for peak occupancy in the mortality equation are similar to the estimations in Table 2, with an insignificant left slope (beta=-0.0518, $p = 0.45$) and a highly significant right slope (beta=1.568, $p < 0.001$). The selection effect, as identified by the correlation coefficient between the error terms in the two equations, is significant (rho=0.272, $p < 0.01$), as are the two variables with exclusion restrictions in the selection equation, the Sunday indicator variable (beta=0.256, $p < 0.001$), and the discharge rate of the other patients (beta=-0.213, $p < 0.001$). In summary, the selection model provides very similar results to the subdistribution hazard model.

7.2. Proportional Hazards

The survival model assumes that the effect of peak occupancy before the observation day is the same for all observation days during a patient's stay in the hospital. To test this proportionality assumption, we divide the seven-day observation window into an early phase, from the first to the third day of stay, and a late phase, from the fourth to the seventh day, and code phase-dependent slopes on either side of the tipping point using a phase dummy variable (Singer and Willett 2003). The tipping point estimate remains at an occupancy level of 92.5% and the estimated coefficients for peak occupancy below the tipping point are 0.032 (sd=0.074, $p > 0.1$) for the early phase and 0.0436 (sd=0.089, $p > 0.1$) for the late phase, while the coefficient estimates for peak occupancy above the tipping point are 1.957 (sd=0.623, $p < 0.01$) for the early phase and 1.662 (sd=0.636, $p < 0.01$) for the late phase. A Wald test does not reject the equal coefficient hypothesis for the two phases ($p = 0.95$). We are therefore satisfied that the proportionality assumption is tenable in our case.

7.3. Expanded Follow-up Period

Recall that we chose the seven-day follow-up period because we expect unobserved heterogeneity amongst patients to increase with time spent in hospital. While most patients with the considered conditions are severely ill during the early phase of their stay, they separate into two groups during the later phase: those who are convalescent and therefore not as vulnerable and not as care-intensive, and those who were particularly ill at the outset and are therefore still in hospital. For very long lengths of stay, the latter, sicker patients may in fact dominate the patient pool and, in conjunction with the monotonically increasing peak occupancy, cause a tipping point. As a robustness check, we estimated a model over a 14-day follow-up period. The model estimates a significant tipping point at 92.1% with left slope 0.126 (sd=0.0579, $p = 0.03$) and right slope 1.36 (sd=0.326, $p < 0.001$). To test the proportionality assumption, we then allowed for different slopes in the first and second week of the patient's stay, as in Section 7.2. The tipping point changed to 93.0%. A Wald test failed to reject the hypothesis of equal slopes only marginally (p=0.07), indicating that the proportional hazards assumption will become violated as we expand the follow-up period. This provides additional justification for our choice of a seven-day follow-up period.

7.4. Multiple Tipping Points and Smooth Splines

To test whether models with multiple tipping points or smooth curves would fit the data better, we used a model selection procedure suggested by Royston and Sauerbrei (2007), implemented in the STATA command *uvrs*, which chooses amongst alternative spline models with multiple breakpoints. We allowed for a maximum of 10 spline pieces with nine breakpoints located at the 10th, 20th,..., 90th percentiles of peak occupancy. The algorithm chooses the best-fitting model by comparing successively increasingly complex spline models, i.e. models with an increasing number of breakpoints across the possible locations, with the most complex model, i.e. the model with nine breakpoints. The algorithm stops when this most complex model does not provide a significantly better fit at the 5% significance level, based on the chi-square statistic of log-likelihood differences. If all goodness-of-fit tests are significant, the most complex model is chosen. We first estimate linear spline models, i.e. continuous models with linear pieces between the breakpoints. This method identifies the piecewise linear model with a single tipping point at the 90th percentile of peak occupancy as the best fit. Repeating this procedure with cubic instead of linear splines allows for nonlinearity between breakpoints but forces smoothness at breakpoints. The resulting best-fitting cubic spline also has a single tipping point at the 90th percentile of peak occupancy, but has lower likelihood than the piecewise linear model: The advantage of nonlinearity between breakpoints is insufficient to compensate for the forced smoothness at the breakpoints. We are therefore satisfied that the piecewise linear model with a single tipping point is appropriate.

8. Managerial Implications of the Tipping Point

Our empirical study provides evidence that occupancy levels above a tipping point are associated with a substantial increase in in-hospital mortality. If the tipping point is reached frequently, the hospital will experience a sustained quality problem, which may even lead to its closure (Ruef and Scott 1998). Two natural managerial levers are capacity increases and capacity pooling with nearby hospitals; the first will reduce occupancy levels across the board, while the second will reduce the variability of occupancy levels. Both actions imply that fewer patients exceed the occupancy tipping point. We estimate the effects of these interventions on the basis of our data and discuss the value of flexibility in the context of capacity increase.

8.1. Rigid Versus Flexible Capacity

In this section we analyze the effect of a 1% increase in hospital system capacity on mortality, and the associated cost. We consider two options: a rigid capacity expansion with fully staffed beds, and a semi-flexible capacity expansion, where beds are fully resourced with the exception of staffing, which is flexibly deployed in response to occupancy surges. Increasing system capacity by 1% reduces peak occupancy for all patient-days by a factor of 1/1.01. Our model allows us to predict the corresponding changes in daily mortality hazards, which we sum up across the sample to obtain the number of saved lives in our sample. Increasing capacity across the sample by 1% reduces the number of patients who were exposed to occupancy above the tipping point from 14,321 to 12,039: a reduction by 15.9%. The model predicts that 21 lives could have been saved with a 1% increase in capacity, amounting to 3.9% of the 541 patients who died after having experienced occupancy above the tipping point. Note that these 21 saved patients account for 26.9% of the 78 patients that could have been saved if *no* patient had been exposed to occupancy levels above the tipping point (see Section 6.1). For the cost-benefit analysis, we annualize the number of saved deaths, accounting for the fact that observation periods differ by departments in the sample. This results in 22.13 lives saved per annum in the hospitals in our sample.

We estimate the annual costs of a 1% increase in capacity using national average costs (German Bureau of Statistics 2013) and department-specific staffing information from published hospital reports. We differentiate between clinical staff costs, related to doctors and nurses, and other infrastructure and overhead costs of capacity, such as beds, space or support services. We consider two options: fixed staffing and flexible staffing. The fixed staffing option assumes that both clinical staff and other infrastructure costs in the department are increased by 1%, while for the flexible staffing option only infrastructure costs are increased by 1% and clinical staff costs are only increased by 1% on days when occupancy is above the tipping point.

For the fixed capacity option, we first increase medical staffing in all departments by 1%. For the 14 departments where we did not have staffing information, we used the mean of the staffing unit of per capacity of the other departments of the same type. On aggregate, a 1% increase in clinical staffing in all departments in our sample requires 52 doctors and 164 nurses. Based on national average costs this results in total costs of 13.8M Euros. Taking account of departmental capacities and department-type specific national cost averages, we calculated costs for support services (radiology, pathology, anaesthesia) of 4.6M Euros, administrative overheads and logistics (e.g. kitchen services, energy, building maintenance) of 10.0M Euros, and capital costs of 2.0M Euros. Capital costs are based on investment costs of 0.2M Euros per bed and a depreciation period of 25 years (Bavarian Ministry of Finance 2013). All other on-costs were calculated on the basis of national average costs of the three most frequent conditions (diagnosis-related groups) in the department (InEK GmbH 2013). In summary, the estimated total annual cost of a 1% capacity expansion in the sample departments amounts to 30.4M Euros, of which 13.8M Euros are costs of additional departmental medical staff. In relation to the 22.13 saved lives associated with a 1% capacity increase, this amounts to a cost of 1.37M Euros per live saved. This is a very conservative

estimate of the benefits of a 1% capacity increase, as it is likely that for each avoidable death there are many more adverse events that result not in death but in harm and associated additional medical, legal, and reputational costs. In addition, the capacity expansion will not only benefit the patients with one of the six conditions considered in this paper but all patients in the department.

The tipping point phenomenon suggests installing semi-flexible capacity and employing this capacity only when occupancy reaches the safety tipping point. We can estimate the associated costs by assuming that the required infrastructure is installed but its departmental medical staffing remains flexible. The total costs of capacity without medical staffing is 16.6M Euros. Since only 4.0% of departmental days in the sample had occupancy above the tipping point, this reduces staffing costs to 0.8M Euros, giving a total cost of 17.4M Euros, or 0.79M Euros per saved life; semi-flexible capacity is 42.7% cheaper than fixed capacity and achieves, due to the tipping point characteristic, the same mortality reduction.

8.2. Capacity Pooling

As capacity expansion is associated with high costs, we study the effect of pooling as a potentially less costly alternative. Pooling reduces the variability of demand and therefore of occupancy levels, which in turn reduces the propensity of a patient experiencing occupancy levels above the tipping point. Pooling is often implemented at the hospital level through cooperation agreements. Such agreements can include transfers of patients before or following admission, as well as transfers of staff to cover shortages at a partnering hospital. Hitherto, the main rationale for such cooperation is cost-reduction; the safety aspect, due to reducing the proportion of days with occupancy levels above the tipping point, is less appreciated. We can estimate this effect within our sample.

In order to achieve synergies from pooling, especially for emergency patients, pooled hospitals should be in close proximity so that ambulance diversions in response to high occupancy levels do not cause inappropriate delays. We use German zip-codes to estimate the distances between the hospitals in our sample. This allows us to group the 83 hospitals into clusters. We do this step-wise, starting from single hospital clusters by merging two clusters if they contain two hospitals that are less than 30 km apart. This leads to 43 hospital clusters for our sample, of which the largest consists of 18 hospitals. The maximum distance between any two hospitals within any of the clusters is 53 km. We then recalculate daily occupancy levels for each department type across the hospitals in the clusters by pooling departmental capacities: For each day of the year we added midnight patient counts across departments of the same type in the cluster and divided these by the sum of the capacities of these departments to obtain cluster occupancy levels. Pooling reduces the 71,510 patient-days with peak occupancy above the estimated tipping point to 50,302 patient-days: a reduction of 30%. At the patient level, 14,321 patients were exposed to occupancy above the tipping point on some day of their first seven days in the hospital. After pooling, this number is reduced to 10,114 patients. We can calculate the number of saved lives by calculating for each patient-day the difference between the model-predicted mortality hazard for the realized peak occupancy level and for the pooled peak occupancy and summing up the differentials over all patient-days. This resulted in an estimated 27 lives saved by pooling. As we had estimated earlier that 78 patients could have been saved if *no* patients had been exposed to occupancy above the tipping point, this occupancy effect can be reduced by 34.4% by pooling alone.

These estimated benefits of pooling are only indicative and discard important costs, such as additional cost of patient transport. Also, even if pooling of hospitals is desirable, it may not be easy to implement. Hospitals that compete fiercely, as is often the case in large cities, are less likely to cooperate; insurance companies may have different agreements with different hospitals, which can lead to reimbursement problems; patients will often be involved in the choice of hospital and may wish to go to a specific hospital, even if it is running at high capacity. Despite these limitations, the analysis above suggests that safety can be a powerful rationale for pooling in addition to the prevalent cost argument.

8.3. Daily staffing variation

Needleman et al. (2011) show in the context of a single medical center that staffing below target during a patient's stay affects mortality. We do not have daily departmental staffing data, which is a limitation that our study shares with other multi-center studies of occupancy effects. The inclusion of seasonal variables, such as admission day of the week, controls to some extend for systematic variations in staffing. However, it is conceivable that current or anticipated future occupancy levels affect daily staffing levels. Management may react to current high occupancy or anticipated future high occupancy by scheduling additional staff, either from less busy units in the hospital or from nursing pools or nursing agencies. This managerial response would lead to increased staffing when occupancy rises. At the same time, staff response to high occupancy might lower staffing levels. Green et al. (2012) report within the context of an emergency department that while absenteeism was not correlated with current workload after controlling for fixed effects, high anticipated future workload was associated with an increased rate of absenteeism. This potential causal effect of current or anticipated future occupancy on daily staffing levels does not invalidate our estimations of the occupancy effect, but rather points to staffing variation as a potential *mechanism* through which occupancy affects mortality; occupancy is still the root cause of mortality (see e.g. Section 3.2.3 in Angrist and Pischke (2009)). Nonetheless, it would be important to know whether daily staffing variation mediates the occupancy effect, as this would provide additional insight into how staffing after an occupancy peak can be used to reduce mortality. While we do not expect significant correlation between daily occupancy and daily staffing, after controlling for seasonal and department fixed effects, in the German hospital context during our observation period 2004/05, such correlations may well be present today as working conditions have deteriorated in the wake of cost-cutting efforts (Zander et al. 2013) and managerial response mechanisms, such as the use of nursing pools, have improved over the past decade.

9. Conclusion

Hospitals cannot turn away patients with acute conditions and therefore have to deal with surges in demand, leading to spikes in occupancy levels. When occupancy is very high, the managerial ability to respond by exploiting variability buffers becomes constrained as these buffers become depleted. The strain is passed on to employees, who are forced to ration limited resources to cope with excessive demand, while stress impairs their cognitive abilities. In combination, these effects lead to safety tipping points in hospitals. Neither the organization nor its clinical staff are able to absorb a further increase in occupancy beyond the safety tipping point without significant deterioration in the quality of care. Our empirical analysis demonstrates that such tipping points exist. In our sample, a patient's mortality risk begins to increase significantly with occupancy when occupancy levels exceed a tipping point of 92.5%. Our results provide ammunition for operations managers when their finance colleagues argue that capacity can be reduced while activity levels are maintained. When this is done, more patients will experience an unsafe day in the hospital, i.e. a day when occupancy levels exceed the safety tipping point. In our sample 17.4% of patients experienced days with occupancy above the estimated tipping point and one in seven deaths among these patients in our sample is accounted for by high departmental occupancy.

The existence of safety tipping points is important. Earlier studies had neglected this phenomenon and either did not find a relationship between occupancy and mortality or exaggerated the effect at low occupancy levels and underestimated the effect at high levels. This is particularly relevant in the debate about capacity pooling as the associated reduction in occupancy variability reduces the propensity of a patient to experience occupancy above the tipping point. In a simulation of capacity pooling, we have estimated that in our sample 34.4% of the deaths that are accounted for by occupancy could have been avoided if capacity had been pooled. This significant safety effect of capacity pooling is not apparent in a linear occupancy model, where a gain from

avoiding high occupancy in one hospital is offset by the loss in increasing occupancy in another. This adds an important safety dimension to the cost-reduction benefits of the capacity pooling of healthcare services.

A further important implication of the safety tipping point is that it has a marked effect on the value of semi-flexible capacity, specifically capacity with flexible medical staffing. Within the context of our data, we have argued that an increase in capacity with flexible staffing, triggered by occupancy levels, may be more than 40% cheaper than rigid capacity and achieve the same safety improvement in terms of mortality reduction.

We have pointed out the statistical challenges in estimating occupancy tipping points with respect to mortality, and specifically the fact that avoidable mortality is a rare event that requires large samples. It was necessary to assemble a multi-hospital department-based dataset for our analysis. There are, however, other less severe but more frequent indicators of quality deterioration, such as readmission to hospital, operating theater, or ICU, patient falls, medication errors, or patient complaints, which are routinely recorded by hospitals (KC and Terwiesch 2012, Kim et al. 2013, Long and Mathews 2013). While they are rarely associated with a patient's death, these events are indicative of poor clinical quality and are likely to be affected by occupancy levels. These events are sufficiently frequent to have potential for department-level tipping point analyses, similar to the analysis conducted in this paper. The strength of a department-level analysis would be further enhanced by including data on day-to-day variations of staffing levels, which is currently not available in a sufficiently standardized form for multi-hospital studies. Results of such analyses could be very powerful in providing sorely needed evidence to guide the design of departmental escalation policies and process re-engineering efforts, as summarized by feedback we received from the clinical director of a large medical department: "The tipping point is the point at which further reductions in staff are associated with worse outcomes. If we could identify what factors altered the tipping point, we might be some way to understanding how to improve outcomes with less staff – to increase efficiency. What are those factors - cultural, technological, skill mix, experience?" More research is required to answer this question.

Acknowledgments

We received valuable feedback on earlier versions of the manuscript from Carri Chan, Michael Freeman, Paul Kattuman, Stelios Kavadias, Christoph Loch, Nicos Savva and Christian Terwiesch. We would also like to thank Christian Rossbach of Activa GmbH for his support in assembling the staffing data.

References

Agency for Healthcare Research and Quality. 2006. Inpatient quality indicators. URL `http://www.qualityindicators.ahrq.gov/downloads/iqi/2006-Feb-InpatientQualityIndicators.pdf`.

Angrist, J.D., J.S. Pischke. 2009. *Mostly harmless econometrics: an empiricist's companion*. Princeton University Press, Princeton.

Bates et al. 1999. The impact of computerized physician order entry on medication error prevention. *J Am Med Inform Assoc* **6** 313–321.

Bavarian Ministry of Finance. 2013. Hospital support in Bavaria (in German). URL `http://www.stmf.bayern.de/kommunaler_finanzausgleich/allgemeines/krankenhausfoerderung`.

Bell, C.M., D.A. Redelmeier. 2001. Mortality among patients admitted to hospitals on weekends as compared with weekdays. *New England Journal of Medicine* **345**(9) 663–668.

Berk, E., K. Moinzadeh. 1998. The impact of discharge decisions on health care quality. *Management Science* **44**(3) 400–415.

Berry Jaeker, J., A.L. Tucker. 2012. Hurry up and wait: Differential impacts of congestion, bottleneck pressure, and predictability on patient length of stay. Working paper, Harvard Business School.

Brennan et al. 1991. Incidence of adverse events and negligence in hospitalized patients: Results of the Harvard Medical Practice Study I. *New England Journal of Medicine* **324** 370–376.

Buckley, T.A., T.G. Short, Y.M. Rowbottom, T.E. Oh. 1997. Critical incident reporting in the intensive care unit. *Anaesthesia* **5** 403–409.

Department of Health (UK). 2006. Improving the use of temporary nursing staff in nhs acute and foundation trusts. URL `http://www.nao.org.uk/publications/0506/improving_the_use_of_temporary.aspx`.

Department of Health (UK). 2012. Cancelled elective operations. URL `http://www.dh.gov.uk/en/Publicationsandstatistics/Statistics/index.htm`.

Dickerson, S.S., M.E. Kemeny. 2004. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychological Bulletin* **130**(3) 355–391.

Dugan et al. 1996. Stressful nurses: The effect on patient outcomes. *J. Nursing Care Quality* **10**(3) 46–58.

Eefje et al. 2010. Effect of a comprehensive surgical safety system on patient outcomes. *N Engl J Med* **363** 1928–1937.

Elixhauser, A., C. Steiner, D.R. Harris, R.M. Coffey. 1998. Comorbidity measures for use with administrative data. *Medical Care* **36**(1) 8–27.

Fine, J.P., R.J. Gray. 1999. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association* **94**(446) 496–509.

German Bureau of Statistics. 2013. Cost analysis of hospitals 2011 (in German). URL `https://www.destatis.de/DE/Publikationen/Thematisch/Gesundheit/Krankenhaeuser/KostennachweisKrankenhaeuser2120630117004.pdf`.

Green, L., S. Savin, N. Savva. 2012. "Nursevendor problem": Personnel staffing in the presence of endogenous absenteeism (Working paper, London Business School).

Green, L.V. 2004. Capacity planning and management in hospitals. M.L. Brandeau, F. Sanifort, W. Pierskalla, eds., *Operations Research and Health Care*. Springer, 15–41.

Green, W.H. 2003. *Econometric Analysis*. Prentice Hall.

Green, W.H., D.A. Hensher. 2010. *Modelling Order Choices: A Primer*. Cambridge University Press.

Gruen et al. 2006. Patterns of errors contributing to trauma mortality: Lessons learned from 2594 deaths. *Annals of Surgery* **244**(3) 371–378.

Healey et al. 2002. Complications in surgical patients. *Arch Surg.* **137** 611–618.

Hopp, W.J., S.M.R. Iravani, G.Y. Yuen. 2007. Operations systems with discretionary task completion. *Management Science* **53**(1) 61–77.

Hsieh, F.Y., D.A. Boch, M.D. Larsen. 1998. A simple method for sample size calculation for linear and logistic regression. *Statistics in Medicine* **17** 1623–1634.

Iezzoni, L. 2003. *Risk adjustment for measuring health care outcomes*. Health Administration Press.

InEK GmbH. 2013. G-DRG System 2011 (in German). URL `http://www.g-drg.de/cms/G-DRG-System_2011/Abschlussbericht_zur_Weiterentwicklung_des_G-DRG-Systems_und_Report_Browser/Report-Browser_2009_2011`.

Kane et al. 2007. The association of registered nurse staffing levels and patient outcomes: Systematic review and meta-analysis. *Medical Care* **45** 1195–1204.

KC, D.S., C. Terwiesch. 2009. Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *Management Science* **55**(9) 1486–1498.

KC, D.S., C. Terwiesch. 2012. An econometric analysis of patient flows in the cardiac ICU. *Manufacturing and Service Operations Management* (forthcoming).

Kim, S.H., C.W. Chan, M. Olivares, G. Escobar. 2013. Icu admission control: An empirical study of capacity allocation and its implication on patient outcomes. Working paper, Columbia Business School.

Klein, J.P., M.L. Moeschberger. 1997. *Survival Analysis Techniques for Censored and Truncated Data*. Springer Verlag.

Kohn, L.T., J.M. Corrigan, M.S. Donaldson. 2000. *To err is human: building a safer health system*. Institute of Medicine, National Academy Press.

Kucukarslan et al. 2003. Pharmacists on rounding teams reduce preventable adverse drug events in hospital general medicine units. *Arch Intern Med.* **163** 2014–2018.

Laine, C., F. Davidoff. 1996. Patient-centered medicine. *JAMA* **275** 152–156.

Landrigan et al. 2010. Temporal trends in rates of patient harm resulting from medical care. *New England Journal of Medicine* **363** 2124–2134.

Lau, B., S.R. Cole, S.J. Gange. 2009. Competing risk regression models for epidemiologic data. *Am. J. Epidemiol.* **170** 244–256.

Lazarus, R.S., S. Folkman. 1984. *Stress, Appraisal and Coping*. Springer New York.

Leape, L.L., D.M. Berwick. 2005. Five years after To Err Is Human - What have we learned? *JAMA* **293**(19) 2384–2390.

Leape et al. 1991. Incidence of adverse events and negligence in hospitalized patients: Results of the Harvard Medical Practice Study II. *New England Journal of Medicine* **324** 377–384.

Leape et al. 1999. Pharmacist participation on physician rounds and adverse drug events in the intensive care unit. *JAMA* **282** 267–270.

Lessing, C., A. Schmitz, B. Albers, M. Schrappe. 2010. Impact of sample size on variation of adverse events and preventable adverse events: Systematic review on epidemiology and contributing factors. *Qual Saf Health Care* **19** 1–5.

Long, E.F., K.S. Mathews. 2013. Patients without patience: A priority queuing simulation model of the intensive care unit. Working paper, Yale School of Management.

Lupien et al. 2007. The effects of stress and stress hormones on human cognition: Implications for the field of brain and cognition. *Brain and Cognition* **65** 209–237.

Muggeo, V. 2003. Estimating regression models with unknown break-points. *Statistics in Medicine* **22** 3055–3071.

Needleman et al. 2011. Nurse staffing and inpatient hospital mortality. *New England Journal of Medicine* **364**(11) 1037–1045.

Oliva, R., J.D. Sterman. 2001. Cutting corners and working overtime: Quality erosion in the service industry. *Management Science* **47** 894–914.

Padma et al. 2004. International differences in evolution of early discharge after acute myocardial infarction. *Lancet* **363** 511–517.

Piquette, D., S. Reeves. 2009. Stressful intensive care unit medical crises: How individual responses impact on team performance. *Critical Care Medicine* **37**(4) 1251–1255.

Powell, A., S. Savin, N. Savva. 2012. Physician workload and hospital reimbursement: Ooverworked physicians generate less revenue per patient. *Manufacturing and Service Operations Management* (forthcoming).

Quan et al. 2005. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Medical Care* **43**(11) 1130–1139.

Rechel, B., S. Wright, J. Barlow, M. McKee. 2010. Hospital capacity planning: from measuring stocks to modelling flows. *Bulletin of the World Health Organization* **88** 632–636.

Royston, P., W. Sauerbrei. 2007. Multivariable modeling with cubic regression splines: a principled approach. *STATA Journal* **7**(1) 45–70.

Ruef, M., W.R. Scott. 1998. A multidimensional model of organizational legitimacy: hospital survival in changing institutional environments. *Administrative Science Quarterly* **43**(4) 877–904.

Schilling, P.L., D.A. Campbell Jr., M.J. Englesbe, M. M. Davis. 2010. A comparison of in-hospital mortality risk conferred by high hospital occupancy, differences in nurse staffing levels, weekend admission, and seasonal influenza. *Medical Care* **48**(3) 224–232.

Scott et al. 2006. Effects of critical care nurses work hours on vigilance and patients safety. *American Journal of Critical Care* **15** 30–37.

Singer, J.D., J.B. Willett. 2003. *Applied Longitudinal Data Analsyis*. Oxford University Press, Oxford, UK.

Sonnentag, S., C. Fritz. 2006. Endocrinological processes associated with job stress: Catecholamine and cortisol responses to acute and chronic stressors. P. Perrewé, D. Ganster, eds., *Employee Health, Coping and Methodologies*. Emerald, 1–59.

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).

US Dept. of Transportation. 2012. New York drunk driving statistics. URL `http://www.alcoholalert.com/drunk-driving-statistics-new-york.html`.

Wooldridge, J.M. 2002. *Econmetric analysis of cross section and panel data*. MIT Press, Cambridge, Massachusetts.

Zander, B., L. Dobler, R. Busse. 2013. The introduction of drg funding and hospital nurses changing perceptions of their practice environment, quality of care and satisfaction: Comparison of cross-sectional surveys over a 10-year period. *International Journal of Nursing Studies* **50** 219–229.

Appendix: Control Variables

		Correlations	
	Mean (SD)	**Died**	**Peak occupancy**
Peak occupancy	.791 (0.125)	-0.0053***	
Conditions			
AMI (reference)	0.153	0.0058***	0.0194***
CHF	0.218	-0.0032***	-0.0330***
GIH	0.111	-0.0145***	-0.0509***
HIP	0.099	-0.0159*	-0.0498***
PNE	0.177	0.0081***	-0.0066***
STR	0.242	0.0126***	0.0934***
Age categories			
< 10	0.029	-0.0082***	-0.0613***
$>= 10$ and < 20	0.007	-0.0056***	-0.0141***
$>= 20$ and < 30	0.010	-0.0065***	-0.0084***
$>= 30$ and < 40	0.020	-0.0096***	-0.0087***
$>= 40$ and < 50	0.052	-0.0124***	0.0154***
$>= 50$ and < 60	0.088	-0.0138***	0.0165***
$>= 60$ and < 70	0.194	-0.0163***	0.0360***
$>= 70$ and < 80 (reference)	0.289	-0.0049***	0.0104***
$>= 80$ and < 90	0.244	0.0260***	-0.0175***
$>= 90$	0.068	0.0313***	-0.0241***
Elixhauser comorbidities			
Congestive heart failure	0.239	0.0114***	-0.0033*
Cardiac arrhythmias	0.282	0.0100***	0.0210***
Valvular disease	0.107	-0.0105***	0.0103***
Pulmonary circulation disorders	0.029	0.0009	0.0223***
Peripheral vascular disorders	0.067	0.0034*	0.0229***
Hypertension, uncomplicated	0.432	-0.0204***	0.0572***
Hypertension, complicated	0.082	-0.0090***	0.0175***
Paralysis	0.150	0.0085***	0.0556***
Other neurological disorders	0.116	0.0074***	0.0173***
Chronic pulmonary disease	0.118	-0.0027*	0.0100***
Diabetes, uncomplicated	0.157	-0.0024	0.0184***
Diabetes, complicated	0.110	0.0000	-0.0120***
Hypothyroidism	0.033	-0.0074***	0.0082***
Renal failure	0.143	0.0072***	0.0194***
Liver disease	0.032	0.0031*	0.0038**
Peptic ulcer disease excluding bleeding	0.007	-0.0038**	-0.0048***
AIDS/HIV	0.001	-0.0010	0.0045***
Lymphoma	0.006	0.0005	0.0149***
Metastatic cancer	0.013	0.0167***	0.0025
Solid tumor without metastasis	0.025	0.0121***	0.0059***
Rheumatoid arthritis/collagen, vascular diseases	0.013	-0.0051***	0.0109***
Coagulopathy	0.030	0.0123***	0.0258***
Obesity	0.100	-0.0160***	0.0198***
Weight loss	0.023	0.0188***	-0.0148***
Fluid and electrolyte disorders	0.166	0.0192***	-0.0094***
Blood loss anemia	0.011	-0.0029*	-0.0025
Deficiency anemias	0.019	-0.0039**	-0.0035**
Alcohol abuse	0.038	-0.0041**	0.0051***
Drug abuse	0.005	-0.0037**	0.0056***
Psychoses	0.006	-0.0027*	-0.0017
Depression	0.036	-0.0111***	0.0033*
Day of stay			
Day 1 (reference)	0.148	0.0246***	-0.0143***
Day 2	0.145	0.0083*	-0.0680***
Day 3	0.144	-0.0008	-0.0174***
Day 4	0.142	-0.0020	0.0219***
Day 5	0.141	-0.0054***	0.0500***
Day 6	0.140	-0.0112***	0.0711***
Day 7	0.139	-0.0141***	0.0877***
Admission period			
January (reference)	0.096	-0.0004	-0.0059***
February	0.093	0.0019	0.1021***
March	0.103	0.0049***	0.1094***
April	0.089	-0.0004	-0.0147***
May	0.090	-0.0008	-0.0275***
June	0.085	-0.0014	-0.0127***
July	0.083	-0.0008	-0.0444***
August	0.082	-0.0003	-0.0765***
September	0.082	-0.0001	-0.0241***
October	0.088	-0.0007	0.0065***
November	0.088	-0.0013	0.0226***
December	0.021	-0.0020	-0.0966***
Year 2005	0.556	-0.0007	0.1069***
Admission day of the week			
Monday	0.178	-0.0036**	0.0580***
Tuesday	0.161	-0.0047***	0.0400***
Wednesday	0.152	-0.0001	0.0149***
Thursday	0.156	-0.0018	-0.0077***
Friday	0.147	-0.0007	-0.0756***
Saturday	0.103	0.0063***	-0.0547***
Sunday (reference)	0.104	0.0070***	0.0130***
Observation day of the week			
Monday	0.139	0.0003	-0.0553***
Tuesday	0.147	0.0001	0.0124***
Wednesday	0.148	-0.0017	0.0391***
Thursday	0.148	0.0002	0.0468***
Friday	0.154	-0.0031*	0.0337***
Saturday	0.136	0.0026	-0.0237***
Sunday (reference)	0.128	0.0018	-0.0592***
Other controls			
Gender (male=1)	0.498	-0.0113***	0.0284***
Emergency admission	0.534	0.0178***	0.0280***
Admission from another hospital	0.064	-0.0080***	0.0276***
Departmental transfer within hospital	0.057	0.0086***	0.0549***
255 department dummies		not reported	
30 interactions of day-of-stay and conditions		not reported	
Patient days	557,828	557,828	557,828

$^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$