

An Examination of Early Transfers to the ICU Based on a Physiologic Risk Score

Wenqi Hu, Carri W. Chan, José R. Zubizarreta

Decision, Risk, and Operations, Columbia Business School
whu17@gsb.columbia.edu, cwchan@columbia.edu, zubizarreta@columbia.edu

Gabriel J. Escobar

Kaiser Permanente Division of Research, gabriel.escobar@kp.org

Unplanned transfers of patients from general medical-surgical wards to the Intensive Care Unit (ICU) may occur due to unexpected patient deterioration. Such patients tend to have higher mortality rates and longer lengths of stay than direct admits to the ICU. A new predictive model, the EDIP2, was developed with the intent to identify patients at risk for deterioration, which in some cases could trigger a proactive transfer to the ICU. While it is conceivable that proactive transfers could improve individual patient outcomes, they could also lead to ICU congestion. In this work, we utilize a retrospective dataset from 21 Kaiser Permanente Northern California hospitals to estimate the potential benefit of proactive ICU transfers. In order to increase the robustness of our estimation results, we make a number of design choices to strengthen the instrumental variable and reduce model dependence. Using our empirical results to calibrate a simulation model, we find that proactively admitting the most severe patients could reduce mortality rates without increasing ICU congestion. However, being too aggressive with proactive transfers could degrade quality of care and, ultimately, patient outcomes. Thus, while we find evidence that proactive transfers could be effective, they would need to be used judiciously.

Key words: Intensive Care Units, Empirical Models, Matching

1. Introduction

Intensive Care Units (ICUs), which provide care for critically ill patients, often operate near full capacity (Green 2002). ICU admissions have increased by 48.8% from 2002 through 2009 (Mullins et al. 2013), and the usage of ICUs will likely continue to rise with the population aging (Milbrandt et al. 2008). The high cost of ICU care and rising use of ICUs make it of increasing interest to develop a better understanding of the ICU admission decision. In this work, we focus our attention on the ICU admission decisions for patients in general medical-surgical wards and the Transitional Care Unit (TCU), because unplanned transfers to the ICU from these units are associated with worse patient outcomes than direct admissions (Barnett et al. 2002, Ensminger et al. 2004, Simpson et al. 2005, Luyt et al. 2007). Indeed, these patients typically have over three times higher mortality than expected and have longer length-of-stay (LOS) by 10 days

(Escobar et al. 2011). In this work, we use a new real-time physiologic risk score for patients staying in the general ward and the TCU to develop an understanding of the potential benefits and costs of proactively transferring patients to the ICU based on the risk score before they experience rapid deterioration.

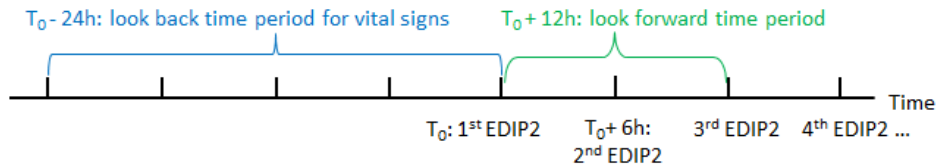
Recognizing the risks associated with unplanned transfers, the Institute for Healthcare Improvement has advocated the development of early warning systems to support the work of rapid response teams (RRTs) with the hope that this would reduce catastrophic medical events that can lead to unplanned transfer to the ICU or in-hospital death on the ward or TCU (Duncan et al. 2012). A rapid response team is a team of clinicians who bring critical care expertise to the bedside of the patient who exhibits early signs of clinical deterioration. No standard detection mechanism exists for RRTs. Some teams employ manually assigned scores such as the Modified Early Warning Score (MEWS) (Stenhouse et al. 2000) and the National Early Warning Score (NEWS) (Royal College of Physicians 2012). Unfortunately, these scores are quite coarse and can suffer from high false positive and false negative rates (Escobar et al. 2012, Gao et al. 2007).

The setting for our work is Kaiser Permanente Northern California (KPNC), an integrated health care delivery system that routinely uses severity of illness and longitudinal comorbidity scores for internal quality assurance. As is the case with some university hospitals (Kollef et al. 2014), KPNC is starting to embed predictive models into the electronic medical record (EMR). In November of 2013, KPNC began a two hospital early warning system pilot project that provides clinicians in the emergency department (ED) and general medical-surgical wards with a severity of illness score (Laboratory-based Acute Physiology Score, version 2, LAPS2) (Escobar et al. 2013), a longitudinal comorbidity score (COMorbidity Point Score, COPS2) (Escobar et al. 2013), as well as a real-time in-hospital deterioration risk estimate (Early Detection of Impending Physiologic Deterioration score, version 2, EDIP2) (Escobar et al. 2012). The real time scoring system provides clinicians with deterioration estimates every 6 hours.

The EDIP2 score predicts the probability of unplanned transfer from the medical-surgical ward or the TCU to the ICU or death on the ward for patients who are ‘full code’ (i.e., those who desire full resuscitation efforts in the event of a cardiac or respiratory arrest) within the next 12 hours, and is updated every 6 hours at 4am, 10am, 4pm and 10pm, as seen in Figure 1. The EDIP2 score utilizes vital signs, vital signs trends, and laboratory tests from the past 24-72 hours as well as patient diagnoses and demographics to determine a patient’s EDIP2 score. The EDIP2 score is more than twice as efficient as the manually assigned MEWS, i.e., the EDIP2 score results in less than half the number of “false alarms” as compared with the MEWS model for identifying the same proportion of all transfers to the ICU (Escobar et al. 2012).

The main premise of the EDIP2 score is to alert clinicians of a patient’s risk of deterioration so that they may consider discrete interventions. We focus on the decision to transfer patients to the ICU based on

Figure 1 Timeline for the EDIP2 score



their EDIP2 scores, which we will refer to as a ‘*proactive ICU transfer*’ throughout this paper. Despite the improved predictive power of the EDIP2 score, concern exists that, if every alert (or a preponderance of the alerts) led to proactive transfer, there would be an increase in ICU congestion. Our goal is to develop an understanding as to whether such a fear is well-founded. To that end, we utilize a dataset of nearly 300,000 hospitalizations to estimate the potential benefit of proactive ICU transfers for individual patients. Because it is not feasible to conduct randomized controlled trials which explore the benefit of ICU admissions, we utilize a comprehensive retrospective dataset. Unfortunately, a common challenge with using such datasets is that there are often unobserved confounders which can increase the likelihood of both ICU admission and adverse patient outcomes (i.e., endogeneity is present). In order to address this problem, we utilize an instrumental variable approach and make a number of design choices to improve the reliability of our estimates. Next, we use a simulation model to examine how various proactive ICU transfer policies might impact patient flow and outcomes at the system level. To the best of our knowledge, our work is the first to consider proactive ICU transfers initiated by a real-time deterioration probability estimate. Our main contributions can be summarized as:

- We utilize an extensive dataset consisting of 296,381 hospitalizations across 21 KPNC hospitals to estimate the impact of proactive ICU transfers on patient mortality risk and patient length of stay to patients of varying levels of severity, as measured by the EDIP2 score. Our dataset is very comprehensive and includes both real-time severity scores (EDIP2), longitudinal patient trajectories (bed histories), as well as patient demographics; these allow us to better model the complex setting for proactive ICU transfers.
- Our empirical approach is guided by design choices to make the study more robust to unobserved confounders and model misspecification. Specifically, we restrict the analysis to the night-time period where our instrument is stronger (and thus is less sensitive to violations in the exclusion restriction) and use recent developments in multivariate matching to reduce model dependence in the outcome analyses (in this way, we avoid extrapolating results to regions of the covariate space where we do not have enough data).
- We conduct a simulation study of patient arrivals to the general medical wards and ICU and find that proactively transferring patients to the ICU could reduce mortality rates, but may increase demand-driven discharges from the ICU and ICU readmissions if done too aggressively.

The rest of the paper is structured as follows. We finish this section with a brief summary of related literature. In Section 2 we present our study setting and describe our data. In Section 3, we describe the empirical challenges we face as well as our approach to estimating the impact of proactive ICU transfers on mortality and LOS. We present our results, including some robustness checks, in Section 4. In Section 5, we describe our simulation model and results. Finally, we provide some concluding remarks and discussion in Section 6.

1.1. Related Literature

Our work is related to two broad areas of research: healthcare operations management and empirical methodologies.

Using empirical and simulation models, we consider using a predictive model (the EDIP2 score Escobar et al. (2012)) to make ICU transfer decisions. The use of predictive models to improve operational decisions has been considered in the emergency department setting (Peck et al. 2012, 2013, Xu and Chan 2015) and in call centers (Gans et al. 2015). Similar to Peck et al. (2012, 2013), we use simulation models to explore the impact of proactive transfer policies. There have been a number of simulation studies examining the impact of congestion on patient delays and diversions (e.g. Bountourelis et al. (2011, 2012) among others). To the best of our knowledge, we are the first to incorporate the possibility of proactive ICU transfers. Moreover, we utilize our empirical findings that rely on causal models to calibrate our simulation model.

A number of works have examined the flow of critical patients through the ICU. One area of focus has been on the discharge of patients from the ICU and the fact that patients are more likely to be discharged when the unit is congested. In turn, these ‘demand-driven’ discharged patients are more likely to be readmitted. Kc and Terwiesch (2012) provides rigorous empirical evidence for this phenomenon while Chan et al. (2012) considers theoretically and via simulation the impact of various discharge strategies. In contrast to this body of work, we consider the transfer of patients into the ICU.

While this work is specifically focused on the transfer of patients from the medical-surgical wards to the ICU, there have been a number of works examining the impact of congestion on ICU admissions (e.g. Shmueli et al. (2003, 2004), Kim et al. (2015)). Both Shmueli et al. (2004) and Kim et al. (2015) use congestion in the ICU as an instrumental variable to address the endogenous nature of the ICU admission decision and to estimate the impact of denied ICU admission on patient outcomes. While we also use ICU congestion as an instrument, our work differs from these in two key ways. 1) We estimate the impact of *proactive* transfers to the ICU. We utilize a unique and dynamic severity measure, the EDIP2 score, to examine how admitting a patient to the ICU *early* based on the likelihood the patient might need ICU care later as given by the EDIP2 score. 2) We make certain design choices to i) strengthen the instrumental

variable and ii) reduce model dependence. The strength of an instrumental variable is measured by its correlation to the endogenous variable and it is considered to be “weak” if the correlation is low. A common approach is to rely on the strength of the instrumental variable as given directly by the data, which is the approach taken in Shmueli et al. (2004), Kim et al. (2015) to estimate the models for patient outcomes. Unfortunately, it is common for instrumental variables (IVs) to be weak in healthcare settings, where the impact of the IV may vary due to the complexity in patient conditions and treatment processes, and this can lead to inference problems. Our approach provides a more robust defense against biases due to unobserved confounders.

Empirical works using instrumental variables often find the partial correlation between the instruments and the included endogenous variable to be low; that is, the instruments are weak (Staiger and Stock 1997). Two crucial problems associated with the use of weak instrumental variables are (1) IV estimates can be largely biased even with a slight violation of the exclusion restrictions and (2) confidence intervals may be misleading and IV estimates tend to be biased in the same way that ordinary least squares estimates are biased (Bound et al. 1995). Nelson and Startz (1990) show that when the instrument is weak and the number of observations is small, the asymptotic approximation to the distribution of the instrumental variable estimate is poor and can result in misleading standard errors and confidence intervals. Imbens and Rosenbaum (2005) address the second problem of incorrect asymptotic approximations for confidence intervals by using permutation inferences to obtain the correct and wider confidence intervals for the IV estimates in order to reflect that the instrument is weak and not informative. While the second problem can be mitigated by increasing the sample size and using other inference methods such as permutation inferences, the first problem persists even with a large sample size.

To address both problems of weak instruments, we draw upon the literature on design of observational studies (Rosenbaum 2010, 2015) and use recent advancements in the methodology of near-far matching (Baiocchi et al. 2010, Zubizarreta et al. 2013, Yang et al. 2014). In particular, we restrict the analyses to the night time, where the effect of the instrument on the treatment is stronger and violations to the exclusion restriction are less likely to occur than during the rest of the day. Also, we use near-far matching to match observations that are near in the covariates (and thus, reduce model dependence) and far on the instrument (potentially strengthening the instrument).

2. Study Setting

In this work, we consider a retrospective dataset of all 296,381 hospitalizations which began at one of 21 hospitals in a single hospital network. We utilize patient level data assigned at the time of hospital admission as well as data which are updated during the patient’s hospital stay.

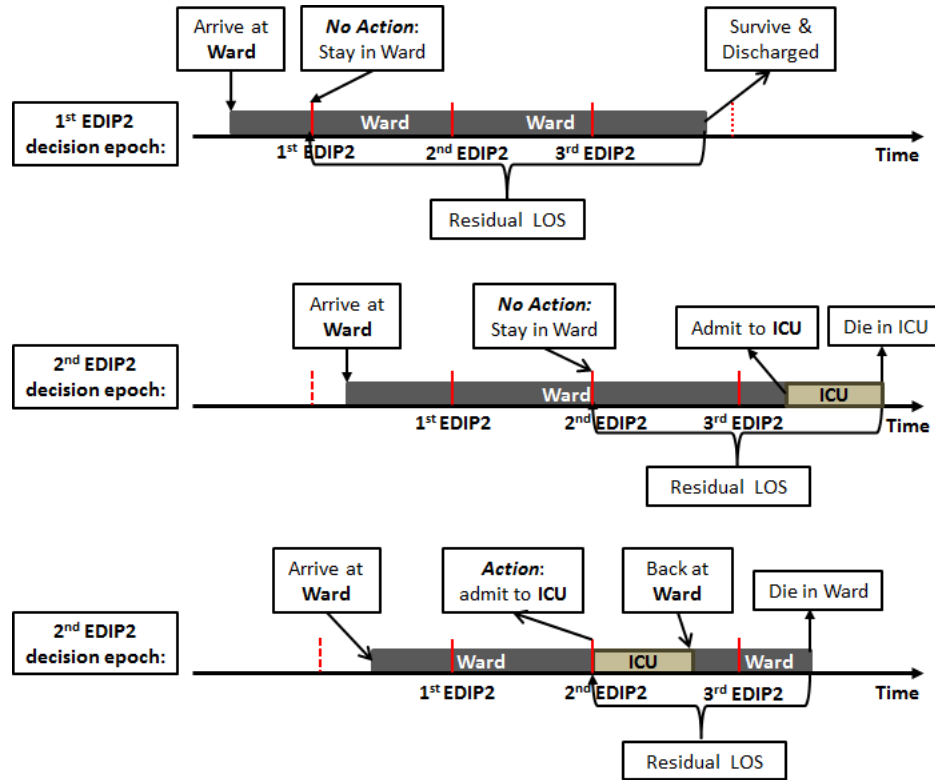


Figure 2 Examples of patient pathways

For every hospitalization episode, we have patient level admission data which includes the patient's age, gender, admitting hospital, admitting diagnosis, classification of diseases codes, and three severity of illness scores. The *COMorbidity Point Score 2 (COPS2)* score captures the patient's burden from chronic diseases. The *Laboratory Acute Physiology Score 2 (LAPS2)* score which is based on laboratory tests captures illness severity. Finally, a *composite hospital mortality risk score (CHMR)* is a predictor for in-hospital death that includes *COPS2*, *LAPS2* and other patient level indicators (see Escobar et al. (2013) for more information on these scores).

During a patient's hospital stay, he may be admitted to and discharged from multiple different units. Our data provides the admission and discharge date and time for each unit stayed in as well as the unit's level of care. In the hospital system which we study, the units are specified as being either the ICU, Transitional Care Unit (TCU), general medical-surgical ward, the operating room (OR), or the Post-anesthesia care unit (PACU).

In addition, all patients in our dataset have EDIP2 scores assigned every 6 hours while in the ward or TCU (scores are not assigned to patients in other units). Figure 2 depicts a few hypothetical patient pathways.

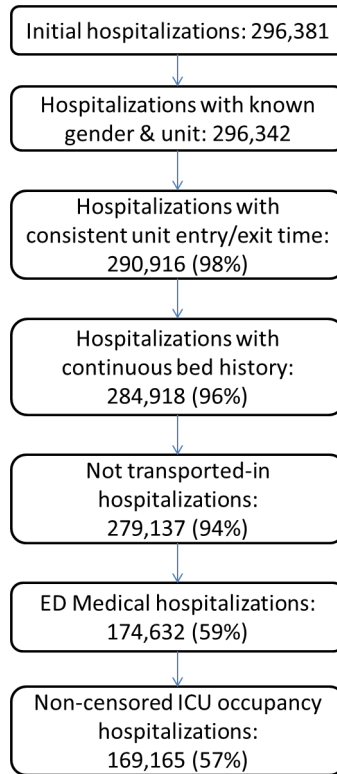


Figure 3 Development of final study cohort

2.1. Data Selection

We utilize data from all 296,381 hospitalizations to derive the maximum capacity and hourly occupancy level of the ICU in each of the 21 hospitals. We found that the maximum ICU occupancy varied from 6 to 34 for the 21 hospitals over our study period. In the patient flow data, 39% of the total ICU arrivals come from ED, 8% are from outside the hospital, 31% come from OR and 22% are from the medical-surgical wards and the TCU.

We now describe our data selection process for our final study cohort, which is depicted in Figure 3. We first eliminate 39 hospitalizations with unknown patient gender or missing inpatient unit code. Next, we eliminated 5,426 hospitalizations because there were inconsistent records for the inpatient unit entry/exit times (e.g. discharge took place prior to admission). 5,998 patients were missing unit admission and discharge times during their hospital stay. We dropped 5,781 hospitalizations for patients who experienced hospital transfers. We focus our study on patients who are admitted to a Medical service via the Emergency Department as this comprises the largest proportion of admitted patients (> 60%). Finally, we remove the episodes admitted in the first and last month of our dataset in order to avoid censored estimates of the ICU occupancy level.

The final study cohort consists of 174,632 hospitalizations from 21 hospitals, among which 13% are admitted to the ICU at least once. Out all of the hospitalizations, 6.7% experience a transfer to the ICU from the ward or TCU. The patient characteristics of the final study cohort are summarized in Table 1.

Table 1 Characteristics of the final study cohort, N=174,632

	Mean	Median	Std. Dev.
First EDIP2	.012	.006	.022
Sex (Female=1)	53.80%		
CHMR	4.04%	1.55%	7.39%
COPS2	45.00	29.00	43.03
LAPS2	73.24	69.00	36.51
Age	67.34	70.00	17.71

2.2. Actions

We consider every EDIP2 time point (4am, 10am, 4pm, 10pm) as a decision epoch. Note that this requires the patient is in the ward or TCU, otherwise, an EDIP2 score would not be recorded and this would not be considered a decision epoch. If a patient is transferred from the ward (or TCU) to the ICU before the next EDIP2 time point, we define this to be an **action**. On the other hand, if the patient remains in the ward (or TCU) until the next EDIP2 time point, we consider it to be **no action**.

Figure 2 depicts a few patient pathways. For the first pathway, if we consider the 1st EDIP2 decision epoch, no action is taken. In fact, for this particular patient, there are three EDIP2 time points, so three decision epochs—no action is taken at all three. The patient in the second pathway also has three EDIP2 time points. When considering the second EDIP2 decision epoch, no action is taken. However, if we were to consider the third EDIP2 decision epoch, we would count the ICU admission as an action.

2.3. Patient Outcomes

In this study, we focus on two measures of patient outcomes: (1) in-hospital death (*Mortality*) and (2) length-of-stay (*LOS*). Because an action can occur at any EDIP2 decision epoch, our measure of LOS is defined as the remaining hospital LOS from the EDIP2 decision epoch. See Figure 2 for examples of how the LOS is measured depending on the EDIP2 decision epoch.

Table 2 summarizes the statistics for in-hospital mortality and hospital residual length-of-stay considering the first EDIP2 decision epoch. On average, a patient stays in the ward/TCU for 21.2 hours before being admitted to the ICU. Patients who are transferred to the ICU have an inpatient mortality rate of 9.5% and stay in the hospital for an average of 149.1 hours following the first EDIP2 time point. Patients who are never transferred to the ICU have an in-hospital mortality rate of only 2.2% and an average residual LOS of 81.0 hours.

Table 2 Summary statistics for 2 patient outcomes, N=174,632

	Mean	Min	Median	Max
Mortality Rate				
All	3.2%			
Transferred to ICU	9.5%			
Never transferred to ICU	2.2%			
Hospital LOS since first EDIP2 (hours)				
All	90.5	0.03	60.8	13,050
Transferred to ICU	149.1	0.03	77.6	13,050
Never transferred to ICU	81.0	0.08	58.9	5,820

3. Empirical models and approach

Our goal is to estimate the potential benefit of proactive admissions to the ICU for patients of different severity as measured by the EDIP2 score. In this section, we describe the empirical challenges in addressing this question and our solution approach.

3.1. Empirical challenges

In our study, we utilize the retrospective patient dataset described in Section 2. While this data is quite rich, we are faced with a number of estimation challenges.

Endogeneity: There are a number of factors a physician will take into consideration when deciding whether to admit a patient to the ICU. Many of these factors, such as age, severity of illness, and primary condition of admission, are observable in our data. The EDIP2 score also provides a severity of illness score which is updated every 6 hours. While we can (and will) utilize this information to adjust for heterogeneous patient severity in our models, it is possible there are unobservable severity factors that influence both the admission decision and a patient's outcome, which can lead to biased inferences when ignoring this potential source of endogeneity. For instance, sicker patients are more likely to be admitted to the ICU, but they are also more likely to stay in the hospital longer and/or die, which would suggest that proactive ICU admission results in *worse* patient outcomes. To address this concern, we utilize an instrumental variable approach. Similar to Kim et al. (2015), we use ICU occupancy as an instrument, with the hypothesis that increasing occupancy will decrease the likelihood of ICU admission.

Weak instruments: While instrumental variables can be effective at mitigating endogeneity biases, problems can arise if the instrument is not strongly correlated with the endogenous variable. This would be the case if high ICU occupancy only has a small effect on reducing the likelihood of ICU admissions. If an instrument is weak, the confidence intervals formed using the asymptotic distribution for two stage least squares may be misleading and IV estimates can be biased in the same way that OLS estimates are biased (Bound et al. 1995, Staiger and Stock 1997). Additionally, the IV estimates based on weak instruments are

highly sensitive to small violations of the exclusion restriction (Bound et al. 1995, Small and Rosenbaum 2008), even with a large sample size. To address this problem, we restrict the analysis to the night time period, where ICU congestion exerts a much stronger influence on ICU admissions than during the rest of the day.

Effect modification: Our goal is to develop an understanding of how proactive admissions to the ICU will benefit patients of different severity as measured by the EDIP2 score. For this we not only need to estimate the causal effect of proactive admissions to the ICU, but also assess this effect at different values of the EDIP2 score. In order to achieve this goal, we resort to parametric statistical models. However, to fit these models, we want to make sure that there is sufficient overlap in the covariate distributions across IV groups, so that the predictions of the models are an interpolation and not an extrapolation, and that their results are less dependent on specific parametric assumptions (Rosenbaum 2010).

3.2. Design choices to strengthen the instrument and reduce model dependence

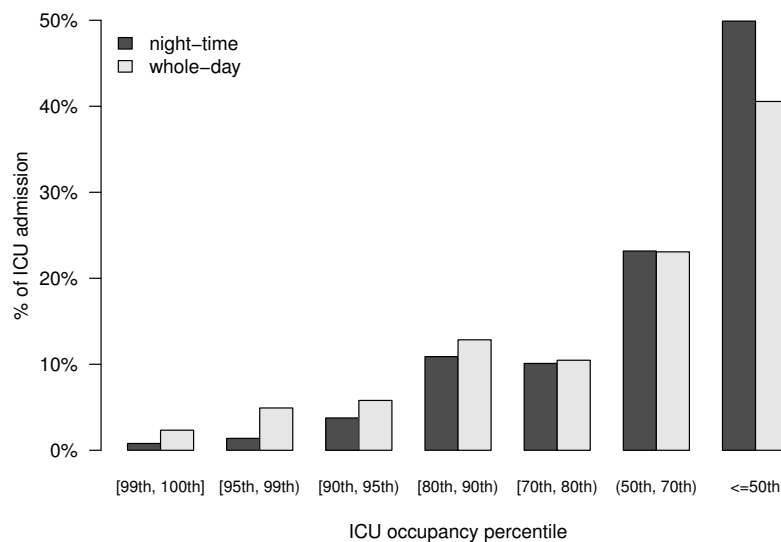
In observational studies of treatment effects, one can draw a sharp distinction between the design and analysis stages of the study (Rubin 2008). Typically, the design stage involves all those decisions and examinations that do not require using the outcomes, whereas the analysis stage relies on the outcomes in some way. This distinction is important to avoid manipulation of the data and preserve the levels of the tests. In our study, to strengthen the instrument and reduce model dependence, we make two design choices. First, we restrict the analysis to the night-time period, where the instrument has a stronger effect on the treatment and violations to the exclusion restriction are less likely. Second, we use recent advancements in multivariate matching mainly to reduce model dependence in the outcome analyses. Naturally, these two choices will result in a smaller sample for analysis; however they enhance the robustness of the findings to unobserved confounders. For instance, Small and Rosenbaum (2008) demonstrates that a smaller study cohort with a stronger instrument is more robust to unobserved biases than a larger study cohort with a weak instrument.

3.2.1. Night time analyses In our setting, there are four EDIP2 decision epochs each day: 4am, 10am, 4pm, and 10pm. There is evidence that ICU admission decisions may vary by day of the week and time of the day (Sheu et al. 2007, Barnett et al. 2002, Cavallazzi et al. 2010), so it is natural to consider whether the impact of ICU occupancy on ICU admissions also vary by time of day.

In KPNC hospitals, nurse staffing is relatively constant across the day for a given unit, with a minimum of one registered nurse for every two patients for the ICU, while the minimum for the ward is 1:4, with TCU staffing ranging between 1:2.5 to 1:3. On the other hand, physician staffing on the ward and TCU can

change dramatically over a 24 hour period, particularly outside regular work hours (7:30 AM to 5:30 PM). Because the physician coverage decreases at night, physicians may be more likely to transfer ‘borderline’ patients to the ICU where they will receive more constant monitoring. As such, the differential impact of a busy ICU on deterring ICU admissions will be *more substantial* at night time. Figure 4 depicts variation in the percentage of ICU transfers by the extent of ICU congestion (as measured by the ICU occupancy percentile) when considering all four EDIP2 time points (whole-day) versus just the 10pm EDIP2 time point (night-time). We can see that the difference between (very) high occupancy (e.g. $\geq 90^{th}$ percentile) and low occupancy ($\leq 50^{th}$ percentile) is much greater when restricting to the night time EDIP2 decision epoch versus considering all four. In comparing the average EDIP2 score of patients at 10pm to the average EDIP2 score across all 4 EDIP2 time points via a t-test, we find that patients at night are less sick at significant level $p < .0001$, which corroborates the idea that patients are more likely to be admitted to the ICU at night, thereby making the average EDIP2 score of patients in the ward or TCU lower. This suggests the instrument may be stronger when only considering the night time decision epoch. When considering the ICU occupancy for all four EDIP2 time points, we will refer to this as the “whole-day instrument”; when considering the ICU occupancy at 10pm, we refer to this as the “night-time instrument”. Note that we do not include the 4am decision epoch into the night time instrument, because the time for action associated with this decision epoch spans 4:00am-9:59am, which includes a few hours where the physician staffing is at day-time levels. Finally, we find that the night time effect is strongest during the first four EDIP2 scores.

Figure 4 Percentage of ICU transfer by ICU occupancy during night-time and whole-day



To the best of our knowledge, we are the first to leverage the different impact of our instrumental variable, ICU occupancy, depending on the time of day. In order to mitigate the inferential problems associated with weak instruments, we restrict our analysis to night-time observations for the first four EDIP2 time points.

3.2.2. Multivariate matching In observational studies, matching methods are often used to adjust for covariates (Stuart 2010, Lu et al. 2011). In these settings, the typical goal of matching is to remove the part of the bias in the estimated treatment effect due to differences or imbalances in the observed covariates across treatment groups. In order to achieve this aim, matching methods select a subset of the observations that have balanced covariate distributions. Generally, matching methods are used to estimate the effect of treatment under the identification assumption of “ignorability” or “unconfoundedness”, which states that all the relevant covariates have been measured (in other words, that there is selection on observables (Imbens and Wooldridge 2009)). More recently, matching methods have been extended to estimation with instrumental variables, which do not require all the relevant covariates to be measured and whose identification assumptions are thus typically considered to be weaker (Baiocchi et al. 2010).

In instrumental variable settings, the goal of matching is to find a matched sample that is balanced on the observed covariates and imbalanced (or separated) on the instrument. The first goal attempts to reduce biases due to imbalances in observed covariates and model misspecification, whereas the second goal aims at strengthening the instrument. This is achieved by near-matching on the covariates and far-matching on the instrument (Baiocchi et al. 2010). We implement this method using integer programming as in Zubizarreta et al. (2013) and Yang et al. (2014). See Appendix A for details.

3.3. Parametric models

We now introduce the parametric models we use to estimate the potential benefits of proactive ICU transfers.

In all of our models, we use the ICU occupancy as an instrumental variable. In order for ICU occupancy to be a valid instrument, it needs to satisfy two main assumptions: 1) it must have a significant impact on the decision to admit, our possibly endogenous treatment, and 2) it must affect the outcome only through the treatment (this is the so-called “exclusion restriction” (Joshua D. Angrist 1996)). To examine the first assumption, we use logistic regression to see how the ICU occupancy impacts the ICU transfer decision when controlling for the patient’s EDIP2 score, age, gender, and other patient level and seasonality controls. We find that the ICU occupancy level is significant at the 5% level. Next, we consider whether ICU congestion is correlated with patient severity. If, for instance, high ICU congestion coincided with the

arrival of high severity patients, one could erroneously attribute poorer patient outcomes to the lack of ICU transfer due to high occupancy rather than to the fact that patients already had higher risk of bad outcomes. This could happen if there is an epidemic or a severe accident which would increase hospital occupancy levels and also increase the severity of patients. We see little evidence that this could be an issue. In particular, we run a linear regression of ICU occupancy on observed patient severity scores—*COPS2*, *LAPS2* and *EDIP2* scores—as well as other patient risk factors, and find that these variables are not relevant to ICU occupancy; also, the coefficient of determination of the model is very low ($R^2 < 0.001$). Assuming that observed patient risk factors are reasonable proxies for unobservable risk measures, ICU occupancy is unlikely to be related to unobservable risk measures.

Formally, we define an ICU to be “busy” when the ICU occupancy is above the 90th percentile of its occupancy distribution. An ICU is “not-busy” when the ICU occupancy is below 70th percentile of its occupancy distribution. The larger the separation between these two thresholds, the more variation there will be in the propensity to transfer a patient to the ICU, thereby increasing the strength of the instrument. However, this comes at the cost of eliminating observations which can be used in the analysis because the ICU occupancy level falls between the two thresholds, i.e. all observations with ICU occupancy in (70th, 90th) percentiles will be dropped. Comparing with other potential cutoffs, the {70th, 90th} definition strikes a good balance in achieving a relatively large difference in ICU transfer rates while dropping a relatively small sample size. We examine other cutoffs as robustness tests in Section 4.3.1.

Remaining Hospital LOS (LOS): We now present our econometric model for LOS, which is defined as the remaining hospital LOS following the *EDIP2* decision epoch in question (see Figure 2). We use a standard two-stage-least-squares (2SLS) method with probit regression in the first stage to account for the binary ICU transfer decision.

We let T_i be the ICU transfer decision, Z_i be the instrument of ICU busyness, and X_i be the patient severity factors and operational controls. Additionally, we define T_i^* as the corresponding latent variable capturing the likelihood of ICU transfer. We have that:

$$\begin{aligned} T_i^* &= X_i^T \beta_1 + \beta_2 Z_i + \epsilon_i \\ T_i &= \mathbb{1}\{T_i^* > 0\} \end{aligned} \tag{1}$$

where we assume ϵ_i to be a normally distributed error term. The second equation of our two-stage model is then:

$$\log Y_i = X_i^T \beta_3 + \beta_4 T_i + \nu_i \tag{2}$$

where ν_i is assumed to be normally distributed and correlated with ϵ_i , so that (ϵ_i, ν_i) follows a bivariate normal distribution with correlation coefficient ρ ,

$$\begin{pmatrix} \epsilon_i \\ \nu_i \end{pmatrix} \Big| X_i \sim \mathcal{N} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right]. \quad (3)$$

A likelihood ratio test can be used to determine whether ρ is significantly different from zero, i.e. whether T_i is indeed endogenous. Note that we take a natural logarithmic transformation for the hospital length-of-stay because the distribution of LOS is skewed as is shown in Table 2. Finally, we fit the model using the entire cohort described in Section 2. This includes patients who do not survive to hospital discharge, but our results are robust to excluding them.

Mortality: We now present our econometric model for mortality. Because *Mortality* is a binary outcome, it is more efficient to model the joint determination of mortality and the ICU transfer decision by a bivariate probit model and use maximum likelihood estimation rather than by two-stage least squares (Wooldridge 2010, Greene 2011).

The treatment equation is the same as before in equation (1). For the binary outcome *Mortality*, the second equation is:

$$\begin{aligned} Y_i^* &= X_i^T \beta_5 + \beta_6 T_i + \nu_i \\ Y_i &= \mathbb{1}\{Y_i^* > 0\}. \end{aligned} \quad (4)$$

Similarly, (ϵ_i, ν_i) follows a bivariate normal distribution with correlation coefficient ρ .

4. Empirical Results

We now present our main empirical results. First, we consider the impact of our design choices. Namely, does restricting to the night-time instrument strengthen our instrument? Additionally, were we able to reduce model dependence by restricting our sample to a well-balanced cohort? Finally, we present our main estimation results.

4.1. Design Choices

4.1.1. Night-time Instrument Our first step in preprocessing is restricting our analysis to the night-time EDIP2 decision epoch. To compare the strength of the instrument when restricting to night-time versus the whole-day instrument, we consider the results of the transfer decision, which is the first stage in the econometric models presented in Section 3.3. The results are summarized in Table 3. Despite the fact that the first night-time EDIP2 sample has only 40% of the number of observations in the first whole-day

EDIP2 sample, we see that the coefficient estimate for the ICU occupancy (*IV*) is much larger and has higher statistical significance, as measured by its p-value. Additionally, when comparing the first-stage partial F-statistic as recommended in Stock et al. (2002), we see that the night-time instrument seems to be much stronger. Moreover, when we examine the average marginal effect—defined as the relative difference in likelihood of ICU admission when the ICU is busy—we see the effect at night-time is nearly triple that of the whole-day. This provides additional support that the night-time instrument has a much larger impact on ICU transfer decisions than the whole-day instrument. With a stronger instrument in the first stage of regression, we can be more confident that the second stage estimation results are less likely to suffer from unobservable biases.

Table 3 Strength of the first night-time *IV* and whole-day *IV* in probit regression models

	Sample Size	<i>IV</i>	(Std. Err.)	P-val.	F-stat.	Pct. Incr. in Prob (<i>Admit</i>)
Night-time	65,845	0.255	(0.084)	0.002	16.208	122%
Whole-day	168,351	0.098	(0.039)	0.012	9.091	34%

4.1.2. Near-Far Matching The next step involves using near-far matching to balance covariates and reduce model dependence (near matching), and separate the matched groups on the instrument and potentially strengthen the instrument (far matching). We first examine the quality of the matched sample by looking at balance tables for all covariates used in near matching. We matched encouraged to discouraged with a 1 : 5 matching ratio, matching in total 85,208 observations (15,149 discouraged observations; 88% of all the available discouraged before matching in the data set).

Covariate balance Tables 9-11 in Appendix A.2 show the covariate balance achieved for patient-risk, day-of-week and calendar-month covariates after matching. We can see that for each patient risk covariate and for each day-of-week, the absolute standardized differences are all less than or equal to 0.1. For calendar month, we slightly relax the restriction on the standardized difference in doing mean balance on calendar month. We do this because several months have more congested ICUs (e.g. January) while other months have less congested ICUs (e.g. September), which makes it difficult to achieve such strict mean balance across the encouraged and discouraged groups. For most of the 12 months, however, the number of observations admitted in each calendar month is close in the encouraged and discouraged groups. As such, we still find the mean balance to be appropriate for calendar month.

We also verify that the number of observations for each hospital in the encouraged and discouraged groups is highly similar with maximum difference of 0.003 (see Figure 11 in Appendix A.2). The number of males and females in the two groups are similarly balanced as well.

In summary, we find that our matched sample is well-balanced, thereby reducing model dependence and allowing for more robust estimates of the effect of ICU admission at different values of the EDIP2 score.

Instrument strength To examine the impact of matching on the strength of the night-time instrument, we need to compare the estimation results using the matched sample and the before-match sample. Note that now we are already restricting to night-time decision epochs. A direct comparison of the p-values of the estimation results from the two samples (as we did when examining the night-time versus whole-day instrument) is unfair due to the difference in sample sizes. Because the standard error of the coefficient estimate is inversely proportional to the square root of the sample size, the p-value of the before-match sample coefficient estimate will be larger than that of the after-match sample even when the estimated coefficients are the same for both samples simply because the after-match sample size is only 53% that of the before-match sample. In order to make a more fair comparison, we randomly draw observations from the before-match data so that there are the same number of observations and the same proportion of observations across each of the first 4 night-time EDIP2 decision epochs as in the after-match sample. We select 1,000 such random draws and report the sample median coefficients and the corresponding p-values. We obtain the 95% confidence interval for the coefficient estimates in the 1,000 random samples using the 2.5 and 97.5 percentiles for the 1,000 estimated coefficients. We also report the median first-stage F-statistic for the IV by applying a two-stage linear probability model to the 1,000 random samples. The 95% confidence interval for the coefficient estimates in the after match sample is calculated using the corresponding estimated standard deviation.

Table 4 Effect of the night-time IV on ICU transfer before and after matching:
N=84,870

	<i>IV</i>	95% CI	P-val.	F-stat.	Pct. Incr. in Prob (<i>Admit</i>)
Before-match	0.237	[0.130, 0.357]	0.004	13.927	114%
After-match	0.201	[0.060, 0.343]	0.005	10.774	95%

Table 4 summarizes the first stage results for the matched sample and the 1,000 random samples of the before-match data. Here, matching was successful in balancing the covariates but it did not increase the strength of the instrument as measured by the F-statistic (actually, it decreased it somewhat, but the instrument is strong by most econometrics standards (Stock et al. 2002)). A possible explanation for this is that matching slightly changes the composition of the sample for estimation, which impacts how many observations comply with the instrument. Table 5 compares the patient severity factors before and after matching and we find that the after match sample is slightly healthier. ICU transfer is very rare (6.7% of hospitalizations experience an ICU transfer), and it is even more rare for healthier patients. As such, the

differential impact of a congested ICU will be smaller than when considering patients who have a base-line likelihood of ICU transfer which is larger. Still, we are comfortable with the estimation results after matching because 1) we already have strengthened the instrument through using the night-time instrument and 2) we are able to reduce model dependence by matching.

Table 5 Mean of patient severity risk factors

	Before-match	After-match
# of observations	159,475	85,208
EDIP2	0.008	0.007
COPS2	45.576	44.922
LAPS2	73.691	72.386
CHMR	4.077%	3.716%

4.2. Estimation Results: Effect of Proactive ICU transfers on Mortality and LOS

Table 6 summarizes the estimation results for *Mortality* and *Residual LOS* models for our final, pre-processed data. Note that because we are using full MLE to estimate these models, the coefficients in the first-stage are slightly different than that of Table 4.

For both outcomes, the instrument is highly significant at 1% level. Being encouraged for ICU transfer (when the ICU is not busy) increases the probability of transfer by 86% and 85% on average for the mortality model and residual LOS model, respectively. We find that ICU transfer has a highly significant impact in reducing mortality risk: proactive transfer to the ICU reduces the average estimated in-hospital mortality from 2.6% to 0.1%. Note that our estimates are for the average effect. While proactive ICU admission may have very little (if any) effect on low risk patients, the effect may be quite substantial for high risk patients. Because the mortality rate for patients on the ward and TCU is very low, this average effect seems quite large. We also see that ICU transfer reduces the average LOS by 33 hours.

Table 6 Estimation results using the night-time IV after matching

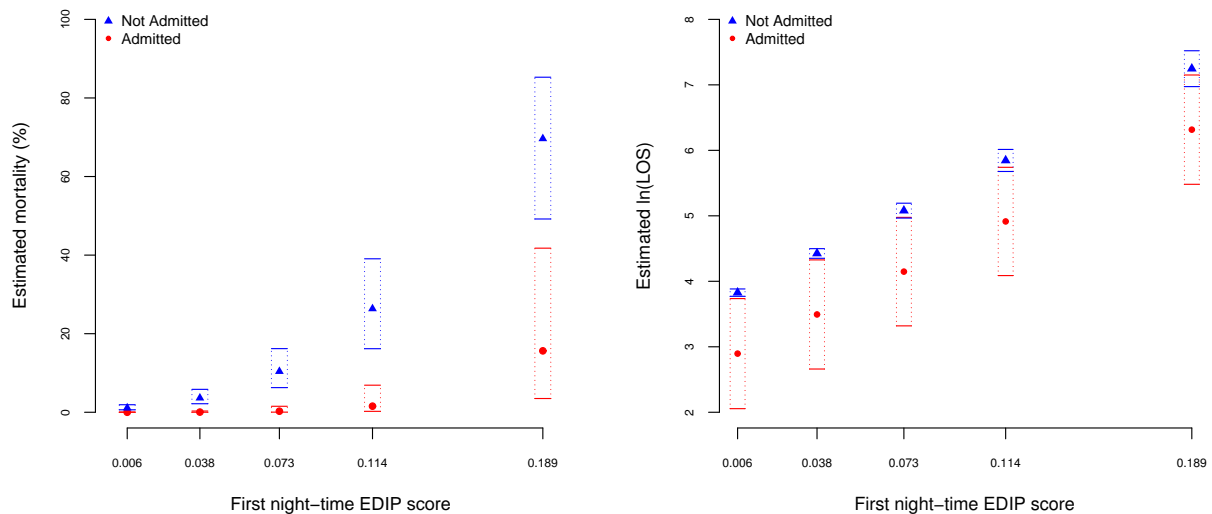
Y	IV (SE)	Pct. Incr. in Prob (<i>Admit</i>)	<i>Admit</i> (SE)	$\Delta\bar{Y}$
Mortality	0.187*** (0.069)	86%	-1.524*** (0.228)	-2.5%
Residual LOS	0.184*** (0.072)	85%	-0.861*** (0.336)	-33 hrs

*** Significance at the 1% level

Our results suggest that proactive ICU transfers can improve patient outcomes on average. However, we wish to gain a better understanding of these improvements for patients of varying severity, as measured by their EDIP2 score. To do this, we obtain the estimated mortality and residual length-of-stay (LOS) when proactively transferred or not transferred to the ICU for each EDIP2 score. For illustrative purposes,

we focus on patients who are admitted to the hospital with the most hospitalizations. Additionally, we consider patients with a primary diagnosis of Neurological Diseases as this was the largest specific group of patients who have their first EDIP2 alarm during the night time in the final dataset. We select five EDIP2 scores to compare; these 5 scores were selected based on recommendations from our medical collaborators. Figure 5 depicts the estimated in-hospital mortality and residual LOS for each EDIP2 score while fixing other covariates at their means. Note that in order to compare the five groups simultaneously, we use the Bonferroni method (Dunn (1961) and Šidák (1967)) to adjust the significance level and obtain the correct 95% confidence intervals for multiple comparisons.

Figure 5 Estimated mortality and log(LOS in hours) for neurological patients in one big hospital, by first night-time EDIP2 with 95% confidence intervals



For all five EDIP2 groups, the 95% confidence intervals do not overlap when considering in-hospital mortality. Thus, we find that proactive ICU transfer has a statistically significant impact (at the 95% confidence interval level) in reducing in-hospital mortality for patients from all five EDIP2 groups. The benefit is largest for the highest three EDIP2 groups. The confidence interval is wider for higher EDIP2 groups due to the fact that less than 1% of our observed sample falls in the highest EDIP2 group, which results in the difficulty in estimating the benefits of ICU transfer on in-hospital mortality precisely for patients with high EDIP2 scores.

For hospital residual LOS, we find that ICU transfer has a statistically significant impact in reducing hospital residual LOS for patients in the two lowest EDIP2 groups, though the magnitude of the gains are

small. For the top three EDIP2 groups, the confidence intervals are relatively large, resulting in overlap. Therefore we cannot conclude on the LOS benefits of proactive ICU transfer for these patients.

4.3. Robustness Checks

We now consider modifications to our initial empirical models to examine the robustness of our results to these alternative specifications.

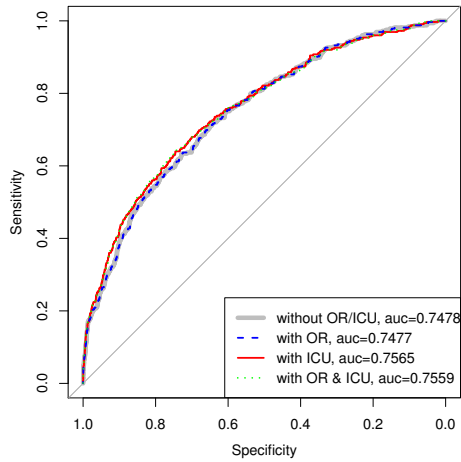
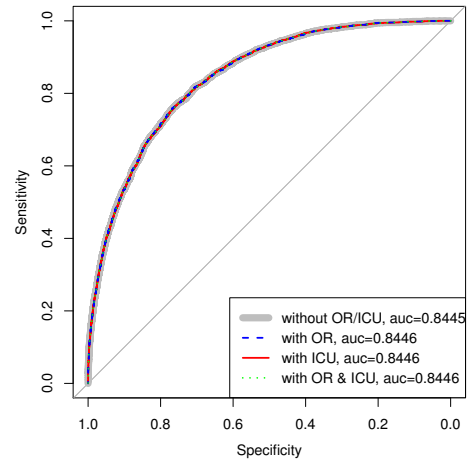
4.3.1. Alternative IV Definition In defining the binary instrumental variable from the continuous ICU occupancy levels, we use the 90th percentile and 70th percentile of the ICU occupancy distribution for each hospital as the threshold for “busy” and “not-busy”. We also tried different combinations of the thresholds, including the 65th, 67.5th, 72.5th and 75th percentiles as the “not-busy” threshold, and 92.5th and 87.5th percentiles as the “busy” threshold. The estimation results are similar with only slight changes in the coefficient estimates.

4.3.2. Additional Covariates In the econometric models for the two patient health outcomes, we have included both patient severity factors and seasonality controls. We also considered including two other risk factors: indicators of whether a patient had been admitted to the ICU or OR before being admitted to an inpatient unit. We fitted a logistic regression of the ICU transfer decisions on all patient severity risk factors and seasonality controls, including the two additional indicators, and use the fitted values to construct a receiver operating characteristic (ROC) curve. An ROC curve is usually used for model comparisons as it depicts relative trade-offs between true positive (benefits) and false positive (costs) for different cut-offs of the parameter (Zweig and Campbell 1993, Pepe 2004). The area under the ROC curve (AUC) is a measure of how well a parameter can distinguish between the admitted and not admitted groups.

Figure 6 shows the ROC curves for the ICU transfer model with and without the two additional risk factors. The four curves almost coincide with each other and the DeLong et al. (1988) test on the difference between any two AUCs shows no significant difference between any two models at the 5% significance level. Thus, it seems that adding these covariates does not significantly improve the estimation model for ICU transfers. We conducted a similar study for mortality to see if these covariates could improve the estimates for the patient outcomes. As before and as seen in Figure 7, the models are practically identical. As such, to avoid over-fitting, we opted not to include the two additional covariates as controls.

5. System level effect of proactive admissions

Thus far, we have focused on the impact of proactive ICU transfer on *individual* patients. Our empirical findings provide evidence that such transfers could improve patient outcomes though the impact that varies depending on a patient’s severity. Given the limited ICU resources, it is reasonable for physicians to have

Figure 6 ROC for night-time ICU transfers**Figure 7** ROC for mortality of night-time patients

concerns about causing ICU congestion by proactively transferring patients ‘before they *really* need it’, which may also create access issues for other, more critical patients who may arrive later. We utilize a simulation model to examine the system level impact of proactive ICU admissions on patient flow and patient outcomes.

5.1. Model of patient flows

We consider a system with two levels of inpatient care: ICU and non-ICU, where the non-ICU units include the general medical-surgical ward and a TCU if the hospital has one. In this work, we focus specifically on the proactive ICU admission decision and for simplicity of exposition, we will refer to the non-ICU units as the wards, with the understanding that this includes the TCU if one exists. Note that this does not account for transfers from the general medical-surgical ward to the TCU (if the hospital has one), which is a transfer whose consideration that, in theory, could be triggered by the EDIP2 score in KPNC. In order to focus on the physicians’ concern of creating unnecessary over-congestion in the ICU (and because the ICU is often the bottleneck), we assume the ward has ample capacity, but explicitly account for the limited number of ICU beds, which we denote by N .

Patients can arrive to the ICU as transfers from the wards or via an external arrival stream of direct admits (e.g. following surgery or directly from the ED). We model the arrivals of the external arrivals as a Poisson process with rate λ^e . These patients have a hospital LOS which is lognormally distributed with mean μ^e and standard deviation σ^e . We assume that a proportion p^e of the patient’s hospital LOS is spent in the ICU. These patients survive to hospital discharge with probability $1 - d^e$.

The second way patients can be admitted to the ICU is via transfer from the wards. To capture the varying level of severity of patients on the ward, we consider C patient classes. The arrival (and admission

to the ward) pattern for patients of type i follows a Poisson process with rate $\lambda_i^w, i = 1, 2, \dots, C$. Every 6 hours, a patient's EDIP2 score is updated, so patient i 's class will now be $j \in \{1, 2, \dots, C\}$. Alternatively, three other possible events may occur: the patient 1) may 'crash' and require immediate ICU admission 2) fully recover and leave the hospital or 3) die and leave the hospital. Because we are focused on the impact of proactive transfers, which can occur at each EDIP2 decision epoch, we model the evolution of a patient's state on the ward via a discrete time Markov Chain with transition matrix \mathbf{T} with each time-slot corresponding to 6 hours. If a patient from EDIP2 group i requires immediate ICU transfer due to *crashing* on the ward, he will have a hospital LOS which is lognormally distributed with mean μ_i^C and standard deviation σ_i^C . We assume that a proportion p_i^C of the patient's hospital LOS is spent in the ICU. These patients survive to hospital discharge with probability $1 - d^C$.

External (direct admits) arrivals and patients who crash receive the highest prioritization for ICU admission. If there are no available ICU beds available at the time of arrival (or crash), the current ICU patients with the shortest remaining service time will be "demand-driven discharged" (Kc and Terwiesch 2012, Chan et al. 2012). Demand-driven discharged patients have an ICU readmission rate of r^D . External arrival and crashed patients who are not demand-driven discharged have an ICU readmission rate of r^e . We do not incorporate the impact of demand-driven discharges on in-hospital mortality because, while some studies find that mortality risk increases with high ICU occupancy at discharge (e.g. Chrusch et al. (2009)), others do not find evidence of an impact (e.g. Iwashyna et al. (2009), Chan et al. (2012)).

Any patient in the ward can be proactively transferred to the ICU at each EDIP2 decision epoch. Such proactive transfers can only occur if there is an available ICU bed for the transferred patient. If there are not enough available beds in the ICU to accommodate all proactive ICU transfer requests, the most severe patients will be given priority. If a patient from EDIP2 group i is proactively transferred to the ICU, he will have a hospital LOS which is lognormally distributed with mean μ_i^A and standard deviation σ_i^A . We assume that a proportion p_i^A of the patient's hospital LOS is spent in the ICU. These patients survive to hospital discharge with probability $1 - d^A$. If this patient is naturally discharged from the ICU (as opposed to demand-driven discharged), his probability of readmission to the ICU is r_i^A . Otherwise, it is r^D .

5.2. Model calibration

We now calibrate our simulation model using the available data described in Section 2 and our empirical results from Section 4.

We start by considering external arrivals. On average, 3 patients are admitted to the ICU directly from the ED, OR or outside the hospital every day. The average hospital LOS for these patients is 5.45 days, of which 47% is spent in the ICU. The mortality rate for these patients is 9.6%. The ICU readmission rate is

14%, which is similar in magnitude to that found in Kc and Terwiesch (2012). Because patients who are demand-driven discharged exhibit higher readmission rate than those naturally discharged, we set r^D to 16% (Kc and Terwiesch 2012, Chan et al. 2012).

Table 7 Simulation parameters for external arrivals

Parameter	Description	Values
λ^e	Direct admits arrival rate to ICU (per day)	2.76
μ^e	Mean hospital LOS (days)	5.45
σ^e	Standard deviation of the hospital LOS	3.88
p^e	Proportion of hospital LOS as ICU LOS	0.47
d^e	Mortality rate	9.60%
r^e	Readmission rate, naturally discharged	14.00%
r^D	Readmission rate, demand-driven discharged	16.00%

We consider five EDIP2 groups, so that $C = 5$. The ranges for the EDIP2 score in each group are: $[0, 0.03)$, $[0.03, 0.06)$, $[0.06, 0.10)$, $[0.10, 0.14)$, $[0.14, 1]$. As mentioned in Section 4, these groups were based on the recommendation of our medical collaborators. The median EDIP2 score for these 5 groups corresponds to the EDIP2 scores in Figure 5. We use the complete data set before preprocessing to calibrate: (1) arrival rates for each EDIP2 group and (2) the Markovian transition matrix \mathbf{T} (see Appendix B). On average, there are 12 patients arriving at the ward per day, 90% of which belong to the lowest EDIP2 group (i.e. EDIPs $\in [0, 0.03)$). As seen in Figure 5, patients who are proactively admitted to the ICU have lower mortality risk, which we incorporate into our model by leveraging our empirical results. For each EDIP2 group, the mortality rate for proactive transfer is given by the average predicted mortality under proactive transfer for patients whose EDIP2 scores fall into that group. Similarly, the mortality rate for crashed patients is given by the average predicted mortality for no action. Since the reduction in residual LOS in Figure 5 is mostly statistically insignificant when doing multiple comparisons, we do not differentiate between proactive and crashed patients when considering the hospital LOS. We denote the mean and standard deviation of the remaining LOS by μ_i and σ_i , respectively. Similarly, p_i denotes the proportion of hospital LOS which takes place in the ICU. These parameters are summarized in Table 8.

Finally, we utilize ICU readmission rates from the literature to determine r_i^A . The literature suggests a range of readmission rates between 1% and 12%, which are clustered around 4-7% (see, for example, Chen et al. (1998), Durbin and Kopel (1993), Rosenberg and Watts (2000), Baker et al. (2009), Lee et al. (2009), Chan et al. (2012), Brown et al. (2013), Kramer et al. (2013)).

Table 8 Simulation parameters for ward patients by EDIP2 group

Parameter	Description	EDIP2 group				
		1	2	3	4	5
λ_i^w	Arrival rate (patients per day)	11.17	0.65	0.16	0.05	0.05
p_i	Proportion of remaining hospital LOS as ICU LOS	0.39	0.48	0.52	0.52	0.55
μ_i	Mean remaining hospital LOS since ICU transfer, unplanned	8.36	10.98	10.80	15.73	12.11
σ_i	Standard deviation of the remaining hospital LOS since ICU transfer (days)	15.63	18.65	17.78	40.21	25.29
d_i^C	Mortality rates since ICU transfer, unplanned	15.42%	27.34%	34.02%	37.57%	41.70%
d_i^A	Mortality rates since ICU transfer, proactive	0.01%	0.06%	0.37%	1.95%	30.00%
r_i^A	ICU readmission rates, proactive and not demand-driven discharged	1.50%	3.00%	4.50%	6.00%	7.50%

5.3. Results

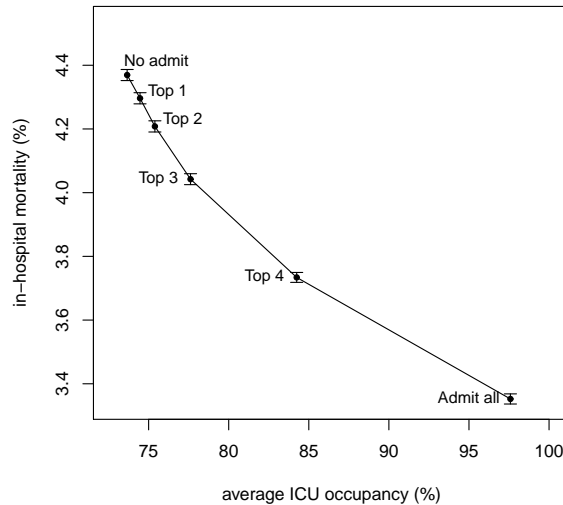
In the hospitals we study, the first quartile for ICU sizes is 11, while the median size is 15 beds. As such, we consider four different ICU sizes $N = 11, 13, 15, 17$. We simulate 1 year with 1 month of warm-up period, over 1,000 iterations.

Proactive ICU transfer policies: Similar to Kim et al. (2015), we consider threshold policies, which have been shown to be optimal admission strategies in a number of settings (e.g. Altman et al. (2001)). Thus, we define a threshold T_{EDIP2} , such that if a patient is in EDIP2 group $i \geq T_{EDIP2}$, the patient will be proactively transferred. If the EDIP2 score is below the threshold, the patient will remain on the ward. For completeness, we consider all possible proactive transfer policies with $T_{EDIP2} \in \{1, 2, 3, 4, 5, 6\}$, where $T_{EDIP2} = 6$ is the case where no proactive transfers are done. For each policy, we compare the average ICU congestion level, the average hospital LOS, the readmission and mortality rate, as well as the frequency of demand-driven ICU discharges.

Figure 8 shows the in-hospital mortality rate and average ICU occupancy level under the various proactive transfer policies for a 15 bed ICU. We note that the qualitative insights from all of our simulations are robust to variations in the ICU readmission rates between 1-12%, so we've elected only to present the results for the parameters given in Tables 7 and 8. Because proactive transfers reduce the likelihood of death, we see that proactively admitting more patients to the ICU could reduce mortality rates. The reduction could be as much as 22% from a nominal mortality rate of 4.4% when there are no proactive transfers to 3.4% when all patients are proactively transferred. However, this reduction in mortality rate comes with an increase in ICU occupancy. We see that proactively admitting *all* patients could result in an average ICU occupancy level of 97%, which may have implications on other quality of care measures. That said, we see a very moderate increase in occupancy when admitting the two highest EDIP2 groups ($T_{EDIP2} = 4$), while

benefiting from an approximately 5% reduction in mortality risk. Thus, while there is merit to physicians' concerns about ICU congestion, the slight increases in ICU occupancy due to strategic proactive transfers for high risk patients may be worth the gains seen in reducing mortality rates.

Figure 8 In-hospital mortality rate under various proactive ICU transfer policies, with 95% confidence intervals, ICU size $N = 15$



Kc and Terwiesch (2012) demonstrates that high ICU occupancy increases the likelihood of demand-driven discharges and, subsequently, readmissions. We now examine how the higher ICU occupancy levels due to proactive transfers impact these measures. Figures 9 and 10 show the weekly demand-driven discharge rate and readmission rate. As the number of ICU beds decreases, the system becomes more congested, the number of demand-driven discharges and, consequently, the number of readmissions, increases. Interestingly, the differences between the demand-driven discharge (readmission) rates are not statistically significant when comparing no proactive transfers ($T_{EDIP2} = 6$) to proactively admitting the top two severity groups ($T_{EDIP2} = 4$). However, being very aggressive with proactive transfers could result in worse care and outcomes; we can see a dramatic increase in demand-driven discharges and readmissions when all patients are proactively transferred. Again, this suggests that some proactive transfers could help improve quality of care at the system level, but it must be done carefully. Indeed, there is evidence that being in the ICU may increase the risk of acquiring an infection from other sick patients (Ylipalosaari et al. 2006, Marshall and Marshall 2005), making it even more undesirable to consider proactively admitting *all* patients. Still, it seems that proactively admitting the most severe patients could save lives without needlessly clogging the ICU.

Figure 9 Frequency of demand-driven ICU discharge, with 95% confidence interval

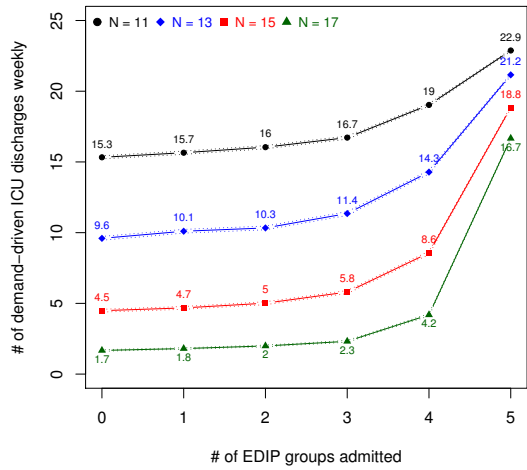
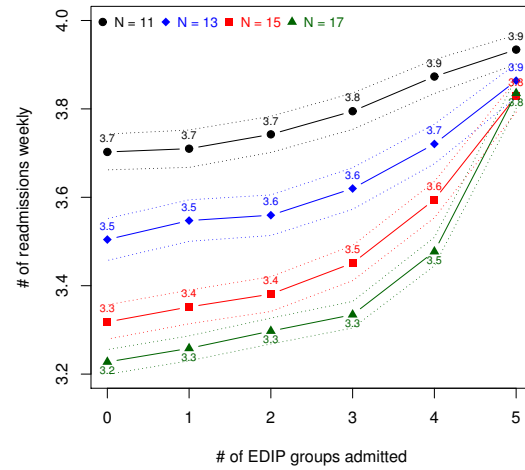


Figure 10 Frequency of readmissions to the ICU, with 95% confidence interval



6. Conclusion and Discussion

Patients who deteriorate and require unplanned transfers to the ICU have worse outcomes. In an effort to mitigate the number of unplanned transfers, the Early Detection of Impending Physiologic Deterioration (EDIP2) score was developed to predict the likelihood a patient will ‘crash’ and require ICU care. The sentiment behind this model was that physicians could act proactively based on this score and admit patients to the ICU before their risk of adverse outcomes increases. In this work, we use empirical and simulation models to estimate the impact of proactive admissions based on the EDIP2 score on individual patients, as well as on the collective patients within the hospital. We find that proactively transferring the most severe patients could reduce mortality rates without sacrificing other patient outcomes; however, proactively transferring too many patients could result in an ICU which is so highly congested that patients would be more likely to be demand-driven discharged and require ICU readmission.

Our empirical strategy relies on two study design decisions. First, we restrict our analysis to the nighttime EDIP2 decision epoch in order to strengthen the instrument and reduce the potential biases introduced by unobserved confounders. Second, we utilize a matching approach to reduce model dependency in order to enhance the robustness of our estimates for the effect of proactive transfers on patients of varying level of severity (as measured by the EDIP2 score). While these decisions can (and do) alter the sample size and characteristics, this is done in a careful manner in which to increase the reliability of our estimates. Such approaches may be beneficial in other healthcare settings where causal inference is challenging due to weak instruments.

This work presents a number of interesting directions for future research. First, we used simulation to compare different proactive transfer strategies. One could consider using a stochastic modeling and

dynamic optimization framework to examine whether alternative policies may be more effective. We note that our simulation model assumes that any patient with an EDIP2 score above a prespecified threshold will be admitted to the ICU; however, in practice, the EDIP2 provides guidance rather than a mandate for physicians making proactive transfer decisions. One could consider policies with possibly lower EDIP2 thresholds to use as an automated alarm to bring physicians to a patient's bedside for evaluation and information gathering, rather than simply as an ICU transfer alarm. Additionally, one could consider explicitly incorporating the future information provided by the EDIP2 score in determining an optimal transfer policy in a similar way that Xu and Chan (2015) uses future information to make ED admission decisions.

References

- Altman, E., T. Jiménez, G. Koole. 2001. On optimal call admission control in resource-sharing system. *IEEE Transactions on Communications* **49**(9) 1659–1668.
- Baiocchi, M., D. S. Small, S. Lorch, P. R. Rosenbaum. 2010. Building a stronger instrument in an observational study of perinatal care for premature infants. *Journal of the American Statistical Association* **105**(492) 1285–1296.
- Baker, D. R., P. J. Pronovost, L. L. Morlock, R. G. Geocadin, C. G. Holzmueller. 2009. Patient flow variability and unplanned readmissions to an intensive care unit. *Critical Care Medicine* **37**(11) 2882–2887.
- Barnett, M. J., P. J. Kaboli, C. A. Sirio, G. E. Rosenthal. 2002. Day of the week of intensive care admission and patient outcomes: A multisite regional evaluation. *Medical Care* **40**(6) 530–539.
- Bound, J., D. A. Jaeger, R. M. Baker. 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association* **90**(430) 443–450.
- Bountourelis, T., D. Eckman, L. Luangkesorn, A. Schaefer, S. Nabors, G. Clermont. 2012. Sensitivity analysis of an ICU simulation model. *Simulation Conference (WSC), Proceedings of the 2012 Winter*. 1–12.
- Bountourelis, T., L. Luangkesorn, A. Schaefer, L. Maillart, S. Nabors, G. Clermont. 2011. Development and validation of a large scale ICU simulation model with blocking. *Simulation Conference (WSC), Proceedings of the 2011 Winter*. 1143–1153.
- Brown, S. E. S., S. J. Ratcliffe, S. D. Halpern. 2013. An empirical derivation of the optimal time interval for defining ICU readmissions. *Medical Care* **51**(8) 706–714.
- Cavallazzi, R., P. E. Marik, A. Hirani, M. Pachinburavan, T. S. Vasu, B. E. Leiby. 2010. Association between time of admission to the ICU and mortality: A systematic review and metaanalysis. *Chest* **138**(1) 68–75.
- Chan, C. W., V. F. Farias, N. Bambos, G. J. Escobar. 2012. Optimizing intensive care unit discharge decisions with patient readmissions. *Operations Research* **60**(6) 1323–1341.
- Chen, L. M., C. M. Martin, S. P. Keenan, W. J. Sibbald. 1998. Patients readmitted to the intensive care unit during the same hospitalization: clinical features and outcomes. *Critical Care Medicine* **26**(11) 1834–1841.

- Chrusch, C. A., K. P. Olafson, P. M. McMillan, D. E. Roberts, P. R. Gray. 2009. High occupancy increases the risk of early death or readmission after transfer from intensive care. *Critical Care Medicine* **37**(10) 2753–2758.
- DeLong, E. R., D. M. DeLong, D. L. Clarke-Pearson. 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* **44**(3) 837–845.
- Duncan, K. D., C. McMullan, B. M. Mills. 2012. Early warning systems: The next level of rapid response. *Nursing* **42**(2) 38–44.
- Dunn, O. J. 1961. Multiple comparisons among means. *Journal of the American Statistical Association* **56**(293) pp. 52–64.
- Durbin, C. G. J., R. F. Kopel. 1993. A case-control study of patients readmitted to the intensive care unit. *Critical Care Medicine* **21**(10) 1547–1553.
- Ensminger, S. A., I. J. Morales, S. G. Peters, M. T. Keegan, J. D. Finkielman, J. F. Lymp, B. Afessa. 2004. The hospital mortality of patients admitted to the ICU on weekends. *Chest* **126**(4) 1292–1298.
- Escobar, G. J., M. N. Gardner, J. D. Greene, D. Draper, P. Kipnis. 2013. Risk-adjusting hospital mortality using a comprehensive electronic record in an integrated health care delivery system. *Medical Care* **51**(5) 446–453.
- Escobar, G. J., J. D. Greene, M. N. Gardner, G. P. Marelich, B. Quick, P. Kipnis. 2011. Intra-hospital transfers to a higher level of care: Contribution to total hospital and intensive care unit (ICU) mortality and length of stay (LOS). *Journal of Hospital Medicine* **6**(2) 74–80.
- Escobar, G. J., J. C. LaGuardia, B. J. Turk, A. Ragins, P. Kipnis, D. Draper. 2012. Early detection of impending physiologic deterioration among patients who are not in intensive care: Development of predictive models using data from an automated electronic medical record. *Journal of Hospital Medicine* **7**(5) 388395.
- Gans, N., H. Shen, Y. Zhou, N. Korolev, A. McCord, H. Ristock. 2015. Parametric forecasting and stochastic programming models for call-center workforce scheduling. *Manufacturing and Service Operations Management*, to appear .
- Gao, H., A. McDonnell, D. A. Harrison, T. Moore, S. Adam, K. Daly, L. Esmonde, D. R. Goldhill, G. J. Parry, A. Rashidian, C. P. Subbe, S. Harvey. 2007. Systematic review and evaluation of physiological track and trigger warning systems for identifying at-risk patients on the ward. *Intensive Care Medicine* **33**(4) 667–679.
- Green, L. V. 2002. How many hospital beds? *Inquiry* **39**(4) 400–412.
- Greene, W. H. 2011. *Econometric Analysis*. 7th ed. Prentice Hall.
- Imbens, G. W., P. R. Rosenbaum. 2005. Robust, accurate confidence intervals with a weak instrument: quarter of birth and education. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **168**(1) 109–126.
- Imbens, G. W., J. M. Wooldridge. 2009. Recent developments in the econometrics of program evaluation. *Journal of Economic Literature* **47**(1) 5–86.
- Iwashyna, T. J., A. A. Kramer, J. M. Kahn. 2009. Intensive care unit occupancy and patient outcomes. *Critical Care Medicine* **37**(5) 1545–1557.

- Joshua D. Angrist, D. B. R., Guido W. Imbens. 1996. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* **91**(434) 444–455.
- Kc, D. S., C. Terwiesch. 2012. An econometric analysis of patient flows in the cardiac intensive care unit. *Manufacturing and Service Operations Management* **14**(1) 50–65.
- Kim, S.-H., C. W. Chan, M. Olivares, G. Escobar. 2015. ICU admission control: An empirical study of capacity allocation and its implication for patient outcomes. *Management Science* **61**(1) 19–38.
- Kollef, M. H., Y. Chen, K. Heard, G. N. LaRossa, C. Lu, N. R. Martin, N. Martin, S. T. Micek, T. Bailey. 2014. A randomized trial of real-time automated clinical deterioration alerts sent to a rapid response team. *Journal of Hospital Medicine* **9**(7) 424–429.
- Kramer, A. A., T. L. Higgins, J. E. Zimmerman. 2013. The association between ICU readmission rate and patient outcomes. *Critical Care Medicine* **41**(1) 24–33.
- Lee, J. Y., S. K. Park, H. J. Kim, S.-B. Hong, C.-M. Lim, Y. Koh. 2009. Outcome of early intensive care unit patients readmitted in the same hospitalization. *Journal of Critical Care* **24** 267–272.
- Lu, B., R. Greevy, X. Xu, C. Beck. 2011. Optimal nonbipartite matching and its statistical applications. *The American Statistician* **65**(1) 21–30.
- Luyt, C.-E., A. Combes, P. Aegerter, B. Guidet, J.-L. Trouillet, C. Gibert, J. Chastre, for the CUB-Ra Network. 2007. Mortality among patients admitted to intensive care units during weekday day shifts compared with “off” hours. *Critical Care Medicine* **35**(1) 3–11.
- Marshall, J., K. Marshall. 2005. ICU-acquired infection: Mortality, morbidity, and costs. H. K. F. van Saene, M. A. D. L. Cal, L. Silvestri, eds., *Infection Control in the Intensive Care Unit*. Springer Milan, 605–620.
- Milbrandt, E. B., A. Kersten, M. T. Rahim, T. T. Dremsizov, G. Clermont, L. M. Cooper, D. C. Angus, W. T. Linde-Zwirble. 2008. Growth of intensive care unit resource use and its estimated cost in medicare. *Critical Care Medicine* **36**(9) 2504–2510.
- Mullins, P. M., M. Goyal, J. M. Pines. 2013. National growth in intensive care unit admissions from emergency departments in the united states from 2002 to 2009. *Academic Emergency Medicine* **20**(5) 479–486.
- Nelson, C. R., R. Startz. 1990. Some further results on the exact small sample properties of the instrumental variable estimator. *Econometrica* **58**(4) 967–976.
- Peck, J., J. Benneyan, D. Nightingale, S. Gaehde. 2012. Predicting emergency department inpatient admissions to improve same-day patient flow. *Academic Emergency Medicine* **19**(9) 1045–1054.
- Peck, J. S., S. A. Gaehde, D. J. Nightingale, D. Y. Gelman, D. S. Huckins, M. F. Lemons, E. W. Dickson, J. C. Benneyan. 2013. Generalizability of a simple approach for predicting hospital admission from an emergency department. *Academic Emergency Medicine* **20**(11) 1156–1163.
- Pepe, M. S. 2004. *The statistical evaluation of medical tests for classification and prediction*. 1st ed. Oxford University Press.

- Rosenbaum, P. R. 2010. *Design of Observational Studies*. Springer-Verlag New York.
- Rosenbaum, P. R. 2015. How to see more in observational studies: Some new quasi-experimental devices. *Annual Review of Statistics and Its Application* **2**(1) 21–48.
- Rosenberg, A. L., C. Watts. 2000. A systematic review of risk factors and outcomes. *Chest* **118**(2) 492–502.
- Royal College of Physicians. 2012. National early warning score (news): Standardising the assessment of acute-illness severity in the NHS.
- Rubin, D. B. 2008. For objective causal inference, design trumps analysis. *The Annals of Applied Statistics* **2**(3) 808–840.
- Sheu, C.-C., J.-R. Tsai, J.-Y. Hung, C.-J. Yang, H.-C. Hung, I.-W. Chong, M.-S. Huang, J.-J. Hwang. 2007. Admission time and outcomes of patients in a medical intensive care unit. *The Kaohsiung Journal of Medical Sciences* **23**(8) 395–404.
- Shmueli, A., M. Baras, C. L. Sprung. 2004. The effect of intensive care on in-hospital survival. *Health Services and Outcomes Research Methodology* **5**(3-4) 163–174.
- Shmueli, A., C. Sprung, E. Kaplan. 2003. Optimizing admissions to an intensive care unit. *Health Care Management Science* **6**(3) 131–136.
- Šidák, Z. 1967. Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association* **62**(318) 626–633.
- Simpson, H. K., M. Clancy, C. Goldfrad, K. Rowan. 2005. Admissions to intensive care units from emergency departments: a descriptive study. *Emergency Medicine Journal* **22**(6) 423–428.
- Small, D. S., P. R. Rosenbaum. 2008. War and wages: The strength of instrumental variables and their sensitivity to unobserved biases. *Journal of the American Statistical Association* **103**(483) 924–933.
- Staiger, D., J. H. Stock. 1997. Instrumental variables regression with weak instruments. *Econometrica* **65**(3) 557–586.
- Stenhouse, C., S. Coates, M. Tivey, P. Allsop, T. Parker. 2000. Prospective evaluation of a modified early warning score to aid earlier detection of patients developing critical illness on a general surgical ward. *British Journal of Anaesthesia* **84**(5) 663.
- Stock, J. H., J. H. Wright, M. Yogo. 2002. A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics* **20**(4) 518–529.
- Stuart, E. A. 2010. Matching methods for causal inference: A review and a look forward. *Statistical Science* **25**(1) 1–21.
- Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. MIT Press, Cambridge, MA.
- Xu, K., C. W. Chan. 2015. Using future information to reduce waiting times in the emergency department. *Working paper, Stanford Graduate School of Business* .
- Yang, F., J. R. Zubizarreta, D. S. Small, S. Lorch, P. R. Rosenbaum. 2014. Dissonant conclusions when testing the validity of an instrumental variable. *The American Statistician* **68**(4) 253–263.

Ylipalosaari, P., T. I. Ala-Kokko, J. Laurila, P. Ohtonen, H. Syrjälä. 2006. Intensive care acquired infection is an independent risk factor for hospital mortality: a prospective cohort study. *Critical Care* **10**(2) R66.

Zubizarreta, J. R. 2012. Using mixed integer programming for matching in an observational study of kidney failure after surgery. *Journal of the American Statistical Association* **107**(500) 1360–1371.

Zubizarreta, J. R., D. S. Small, N. K. Goyal, S. A. Lorch, P. R. Rosenbaum. 2013. Stronger instruments via integer programming in an observational study of late preterm birth outcomes. *Annals of Applied Statistics* **7** 25–50.

Zweig, M. H., G. Campbell. 1993. Receiver-operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry* **39**(4) 561–577.

Appendix A: Matching Formulation and Covariate Balance

A.1. Matching Formulation

Following the notation in Zubizarreta (2012), let $\mathcal{T} = \{t_1, \dots, t_T\}$ be the set of discouraged units, that is, the subjects that encountered high ICU congestion, and $\mathcal{C} = \{c_1, \dots, c_C\}$, the set of encouraged units that faced low ICU congestion, with $T \leq C$. Define $\mathcal{P} = \{p_1, \dots, p_P\}$ as the set of observed covariates. Each discouraged unit $t \in \mathcal{T}$ has a vector of observed covariates $\mathbf{x}_{t,\cdot} = \{\mathbf{x}_{t,p_1}, \dots, \mathbf{x}_{t,p_P}\}$, and each encouraged $c \in \mathcal{C}$ has a similar vector $\mathbf{x}_{c,\cdot} = \{\mathbf{x}_{c,p_1}, \dots, \mathbf{x}_{c,p_P}\}$. Based on these covariates, there is a distance $0 \leq \delta_{t,c} < \infty$ between each pair of discouraged and encouraged units.

We solve:

$$\begin{aligned}
 & \underset{\mathbf{a}}{\text{minimize}} && \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} \delta_{t,c} a_{t,c} - \lambda \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} a_{t,c} \\
 & \text{subject to} && \sum_{c \in \mathcal{C}} a_{t,c} \leq 5, \quad t \in \mathcal{T} \\
 & && \sum_{t \in \mathcal{T}} a_{t,c} \leq 1, \quad c \in \mathcal{C} \\
 & && -b_k \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} a_{t,c} \leq \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} a_{t,c} v_{k,t,c} \leq b_k \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} a_{t,c}, \quad k \in K_1 \\
 & && \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} a_{t,c} v_{k,t,c} \geq c_k \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} a_{t,c}, \quad k \in K_2 \\
 & && a_{t,c} \in \{0, 1\}, \quad t \in \mathcal{T}, \quad c \in \mathcal{C}
 \end{aligned} \tag{5}$$

In our study, $\delta_{t,c}$ is the absolute difference between the EDIP2 scores of discouraged unit t and encouraged unit c , and λ is a tuning parameter that regulates the trade-off between finding close matches in the covariates and matching as many pairs as possible (see Zubizarreta et al. (2013) for a discussion on the choice of this parameter; here, λ is set to the median of the $\delta_{t,c}$'s).

The first two sets of constraints require each discouraged unit to be matched to up to 5 different encouraged units (we determined this matching ratio in view of the large number of available discouraged units before matching and the low expected efficiency gains in going from a 1:5 to a 1:6 matching ratio under an additive treatment effect model). The third constraint only allows each encouraged unit to be matched at most once.

The third set of constraints is the covariate balance constraint, where $b_k \geq 0$ is a scalar tolerance that defines the maximum level of imbalance allowed for the k^{th} constraint and $v_{k,t,c} = f(x_{t,p}) - f(x_{c,p})$ for some suitable function $f(\cdot)$ of the observed covariates (see Zubizarreta (2012) for specific examples of $f(\cdot)$). The fourth set of constraints is the imbalance constraint, where $c_k \geq 0$ is a scalar that defines the minimum level of separation required for the k^{th} constraint.

A.2. Covariate Balance

By means of the integer program (5) above (specifically, by imposing the balancing constraints described above), we balanced the means and in some cases the marginal and joint distributions of the covariates. Tables 9-11, show the balance in means for the five risk covariates, the seven indicators for day of the week, and the twelve indicators for calendar month after matching. Figure 11 shows that the number of observations for each hospital in the encouraged and discouraged groups is highly similar (with maximum difference of 0.3%). Since every hospital is almost equally represented in the encouraged and discouraged group after matching, unobserved confounders at the hospital level are very unlikely to bias our estimates. The number of males and females in the two groups are similarly balanced as well. Finally, we matched exactly for the 38 indicators of disease categories, therefore balancing the joint distribution of the disease categories and hospitals, and disease categories and sex (we actually imposed this constraint by matching separately for each disease category). In summary, we find that our matched sample is well-balanced, thereby reducing model dependence and allowing for a more robust estimate of effect modification.

Table 9 Balance table for risk covariates in means

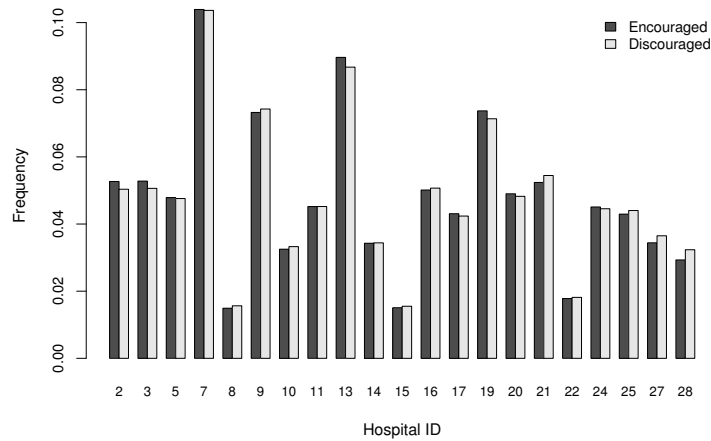
Covariate	Encouraged	Discouraged	Standardized Difference
Age	67.74	67.69	0.00
COPS2	44.89	45.07	0.00
LAPS2	72.28	73.02	-0.02
CHMR	0.04	0.04	-0.02
EDIP2	0.01	0.01	-0.04

Table 10 Balance table for day-of-week

Covariate	Encouraged	Discouraged	Std diff
Sunday	0.15	0.14	0.04
Monday	0.14	0.18	-0.10
Tuesday	0.14	0.17	-0.09
Wednesday	0.14	0.15	-0.03
Thursday	0.14	0.14	0.01
Friday	0.14	0.11	0.10
Saturday	0.14	0.11	0.10

Table 11 Balance table for calendar month

Covariate	Encouraged	Discouraged	Std diff
January	0.07	0.12	-0.17
February	0.08	0.12	-0.14
March	0.09	0.14	-0.15
April	0.09	0.10	-0.04
May	0.09	0.10	-0.02
June	0.10	0.07	0.09
July	0.09	0.06	0.13
August	0.09	0.05	0.14
September	0.08	0.04	0.16
October	0.09	0.06	0.13
November	0.09	0.09	0.02
December	0.05	0.06	-0.07

Figure 11 Balance table for hospital ID after match

Appendix B: Simulation Parameters and Results

Patients staying in the ward are modeled by a discrete time Markov Chain with the transition probability matrix \mathbf{T} below. From the leftmost to the rightmost column of \mathbf{T} , the corresponding states are patient classes 1-5, “crash”, fully-recover and die-in-ward. The top to bottom rows represent patient classes 1-5. Each $\mathbf{T}_{i,j}$ represents the probability of a patient transitioning from EDIP2 group i to the state corresponding to the j^{th} column in \mathbf{T} within each period.

$$\mathbf{T} = \begin{bmatrix} 0.9187 & 0.0051 & 0.0007 & 0.0002 & 0.0002 & 0.0023 & 0.0716 & 0.0012 \\ 0.5906 & 0.2862 & 0.0510 & 0.0118 & 0.0096 & 0.0256 & 0.0113 & 0.0139 \\ 0.2592 & 0.3681 & 0.1961 & 0.0503 & 0.0388 & 0.0451 & 0.0061 & 0.0362 \\ 0.1371 & 0.2658 & 0.2422 & 0.1205 & 0.1064 & 0.0687 & 0.0054 & 0.0538 \\ 0.0742 & 0.1111 & 0.1487 & 0.1261 & 0.2952 & 0.0932 & 0.0079 & 0.1437 \end{bmatrix}$$