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) can occur due to unexpected patient deterioration. Such patients tend to have higher mortality rates and longer lengths-of-stay than direct admissions to the ICU. As such, the medical community has invested substantial efforts in the development of patient risk scores with the intent to identify patients at risk of deterioration. In this work, we consider how one such risk score could be used to trigger proactive transfers to the ICU. We utilize a retrospective dataset from 21 Kaiser Permanente Northern California hospitals to estimate the potential benefit of transferring patients to the ICU at various levels of patient risk of deterioration. In order to reduce the sensitivity of our findings to key identification and modeling assumptions, we use a combination of multivariate matching and instrumental variable approaches. Using our empirical results to calibrate a simulation model, we find that proactively transferring the most severe patients could reduce mortality rates and lengths-of-stay without increasing other adverse events; however, proactive transfers should be used judiciously as being too aggressive could increase ICU congestion and degrade quality of care.

Key words: Intensive Care Units, Empirical Models, Matching

1. Introduction

Intensive Care Units (ICUs) provide care for critically ill patients and often operate near full capacity (Green 2002). ICU admissions in the US 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 ICU admission decisions for patients in general medical-surgical wards and Transitional Care Units (TCUs), because unplanned transfers to the ICU from these units are associated with worse patient outcomes than direct admissions (e.g., Barnett et al. (2002), Luyt et al. (2007)). We use a physiologic risk score (Escobar et al. 2012) that is dynamically updated 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.

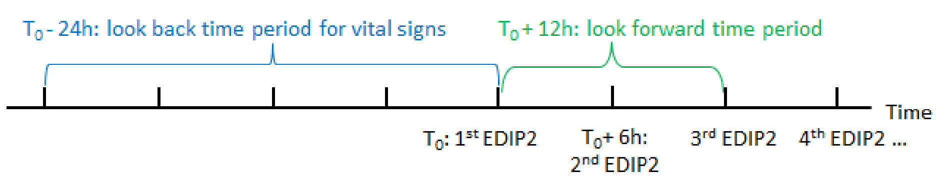
Recognizing the risks associated with unplanned transfers, the US Institute for Healthcare Improvement advocates for the development of early warning systems to support the work of rapid response teams (RRTs) with the hope that this would reduce catastrophic medical events that can lead to unplanned 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).

Our study setting 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. Similar to some university hospitals (e.g., Kollef et al. (2014)), KPNC is starting to embed predictive models into the electronic medical record (EMR). KPNC has developed an early warning system 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), a comorbidity score (COmorbidity Point Score, COPS2), as well as a dynamic in-hospital deterioration risk estimate (Early Detection of Impending Physiologic Deterioration score, version 2, EDIP2) (Escobar et al. 2012, 2013) which is updated throughout a patient's stay in the ward/TCU. The score is updated every 6 hours and has recently been deployed to provide dynamic risk scores to alert a RRT at two pilot hospitals (Kipnis et al. 2016).

The EDIP2 score predicts the probability of death or unplanned transfer from the ward or the TCU to the ICU 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). When using the c-statistic as a measure of model sensitivity and specificity, the EDIP2 out-performs the updated NEWS score and a machine-learning based eCART model with c-statistic of 0.82 versus 0.79 and 0.76, respectively (Kipnis et al. 2016).

Figure 1 **Timeline for the EDIP2 score**



The main premise of the EDIP2 score is to alert the RRT of a patient's risk of deterioration so that they may consider discrete interventions. "Some interventions performed by the response team are simple (administration of oxygen, intravenous fluids, diuretics, and bronchodilators and performance of diagnostic tests)," but often do not correspond to admitting a patient to the ICU (Jones et al. 2011). This is in contrast to what we propose, which is to *proactively admit* patients to the ICU based on their EDIP2 scores *before* the patient crashes. We will refer to this as a *'proactive ICU transfer'* throughout this paper.

Despite the improved predictive power of the EDIP2 score, there are concerns that, if every alert led to proactive transfer, ICU congestion would substantially increase. As such, the current use of the EDIP2 at KPNC is only to call the RRT, not necessarily initiate an admission to the ICU. Our goal is to develop an understanding as to whether such a fear is well-founded. Specifically, if proactive transfers can reduce LOS and mortality for individual patients, then it is possible that proactive ICU transfers will reduce ICU congestion. However, the actual benefit depends on the precise magnitude of the reductions in LOS. This is because by proactively transferring a patient, there is a guarantee that the patient will consume limited ICU resources. However, some proactively admitted patients may never have needed ICU care, so we have needlessly increased ICU congestion, possibly preventing other patients from getting needed care. As such, the relationship between the *ICU load* for proactive transfers may be higher or lower than for traditional, reactive transfers. Whether it is higher or lower is an empirical question, and at the heart of what we are trying to answer. Moreover, due to the externalities one patient can impose on other patients, it is also important to examine how proactively transferring some patients impacts the ability to treat other patients.

We estimate the effect of ICU transfers for patients of varying severity, as measured by the EDIP2 score. Because it is not feasible to conduct randomized controlled trials which explore the benefit of ICU admissions, we utilize a comprehensive retrospective dataset of nearly 300,000 hospitalizations. A common challenge with using such datasets is there are often unobserved confounders which can increase the likelihood of both ICU admission and adverse patient outcomes (i.e., endogeneity is present). To address this problem, we utilize an instrumental variable approach and make a number of design choices to improve the reliability of our estimates. Specifically, we utilize a new near-far matching methodology (Baiocchi et al. 2010, Zubizarreta et al. 2013) that, to the best of our knowledge, has not been used in the Operations Management (OM) literature. Indeed, empirical OM works which utilize instrumental variables typically assume the strength of an instrument is given. In contrast, we make a number of design choices to strengthen our instrument and reduce the potential biases due to unobserved confounders. 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 dynamically updated severity score. 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 ICU transfers on patient mortality risk and length of stay for patients of varying levels of severity, as measured by the EDIP2 score. Our dataset is very comprehensive and includes a dynamically updated severity score (EDIP2), longitudinal patient trajectories (bed histories), as well as patient demographics; these allow us to better model the complex setting for ICU transfers.

- Our empirical approach is guided by design choices to make the study more robust to unobserved confounders and model misspecification. Specifically, we focus our analysis to the night-time period, where we find that the effect of the instrument (ICU congestion) on the treatment (ICU admission) is stronger (and thus the estimates are less sensitive to violations to the exclusion restriction) and use recent developments in multivariate matching to reduce model dependence in the outcome analyses (and in this way 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 to explore the impact of different proactive ICU transfer policies. To the best of our knowledge, this is the first study to examine proactive admission based on a dynamic model of risk. We find that proactively transferring patients to the ICU may reduce mortality rates and lengths-of-stay, but, if done too aggressively, may increase ICU readmissions as well as the likelihood of discharging a patient from the ICU before he/she has completed his/her nominal length-of-stay due to the need to accommodate a new, more severe patient.

1.1. Related Literature

Our work is related to three broad areas of research: 1) healthcare operations management, 2) the use of predictive modeling to guide operational decisions, and 3) empirical methodologies.

In both the medical and operations management literatures, a number of works have examined the flow of critical patients through the ICU. One area of focus has been on 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.

A number of works have also considered the ICU admission decision (e.g. Shmueli et al. (2004), Kim et al. (2015)). Our work differs from this body of literature in a number of important ways. First, the question we are considering is fundamentally different, as we focus on the combined role of a Rapid Response Team, a new dynamic model of patient severity (the EDIP2 score), with proactive ICU transfers from the ward or TCU. In our study, patients are transferred from the ward/TCU to the ICU due to unexpected rapid deterioration, which can happen any time during their stay in the ward. This means that the ICU transfer decision in our study is made *continuously* throughout a patient’s stay in the ward/TCU. In contrast, the ICU admission decision considered in prior works is a *one-time* decision which must be made once the patient is admitted to the hospital. As such, the nature of the ICU admission decision is different both in terms of frequency and timing; moreover, the patient populations considered are quite different which could result in differences in the impact of the decision on outcomes. Another differentiating factor is that we utilize recent empirical approaches, which reduce potential biases introduced by unobserved covariates.

The use of RRT in hospitals has been increasing as a number of studies have documented that timely access to critical care can substantially improve patients outcomes (e.g. Evans et al. (2015)). The role of the RRT is to bring a medical team trained in critical care to the bedside of a patient who exhibits signs of physiologic deterioration. While the RRT may end up recommending ICU admission, it is most common for the RRT to perform simple interventions (e.g. administration of oxygen or intravenous fluids) to stabilize the patient (Jones et al. 2011). There are also benefits of using RRTs in a proactive manner (e.g. Danesh et al. (2012), Butcher et al. (2013), Guirgis et al. (2013)); however, the proactive aspect does not relate to the ICU admission decision, as we examine. Rather, the focus of these works is to proactively round on high risk patients (e.g. those recently discharged from the ICU) in order to appear at the bedside of these patients prior to the summoning of a RRT, as is traditionally done. To the best of our knowledge, our work is the first to study *proactive admission decisions*. Moreover, we consider how to make this decision based on a more accurate, dynamic severity measure, the EDIP2 score.

There have been substantial efforts by the medical community to develop predictive models for patient outcomes (e.g readmissions, death, admissions, etc.). A primary motivation behind this work has been to utilize such models to guide operational decisions and allow clinicians and administrators to better utilize limited healthcare resources. This approach has been considered in the emergency department setting (e.g. Peck et al. (2012), Xu and Chan (2016)) and call centers (Gans et al. 2015). In contrast to these prior works, we do not directly use the predicted probability of deterioration or death in the ward/TCU provided by the EDIP2 model. Rather, we use the dynamically updated EDIP2 score as an important covariate to estimate the effect of ICU admission on patient outcomes for different values of the EDIP2 score. Then, using simulation, we assess the impact of proactive transfer policies for different severity groups classified by their EDIP2 scores. There have been a number of simulation studies examining the impact of ICU congestion on patient delays and diversions (e.g. Lowery (1992), Bountourelis et al. (2012) among others). To the best of our knowledge, we are the first to rely on causal models to estimate impact of patient transfers from the ward/TCU at different levels of patient severity and, in turn, utilize these estimates to develop an understanding of the potential benefits of proactively transferring patients into the ICU.

More broadly, the tension we examine is a short-term increase in resource utilization with the intent of preventing longer-term problems which may arise in the future and consume even more resources. An analogous question arises in the manufacturing literature because failures during factory operations can be more costly than replacing a machine before failure, while being too proactive can also become very costly (see McCall (1965), Pierskalla and Voelker (1976), Barlow and Proschan (1996) and related literature). In the preventative health screening setting, early detection (Özekici and Pliska 1991) and early interventions (Ormeci et al. 2016) can increase the likelihood of positive outcomes for cancer patients. Our work is differentiated in that we consider a very different problem setting (proactive ICU transfers) and we also utilize state-of-the-art empirical approaches to rigorously estimate the causal effect of transferring patients at different severity levels, as measured by the EDIP2 score, in order to calibrate our simulation model.

A major challenge in estimating the causal effect of ICU transfer on health outcomes is that it is unethical to conduct a randomized experiment, so we must rely on observational data, which can be subject to biases introduced by unobservable covariates. To address this challenge, we utilize an instrumental variable (IV) approach. In the empirical OM literature, the strength of an IV is typically taken as given and instrumental variable analysis tends to rely on strong parametric assumptions implied by regression models. Unfortunately, it is common for IVs to be weak in healthcare settings, including the setting we study here, and this can lead to inference problems. Another problem when doing routine regression analysis is that, with pure model-based adjustments, a few observations can unduly influence the results of a study (see Imbens (2015) and Rosenbaum (2016)). To address both the problems of weak instruments as well as model dependence, we draw upon the literature on design of observational studies (Rosenbaum 2010) and use recent advancements in the methodology of near-far matching (Baiocchi et al. 2010, Zubizarreta et al. 2013, Yang et al. 2014).

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 patients' hospital stay.

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 which are assigned at the time of hospital admission. The *COmorbidity Point Score 2 (COPS2)* score is a measure of chronic disease burden and a score greater than 65 could be someone with 3–4 significant comorbidities (e.g., diabetes, Chronic Heart Failure, and cancer). The *Laboratory Acute Physiology Score 2 (LAPS2)* score is based on laboratory tests and measures a patient's acute instability over the 24–72 hours preceding hospital admission. A patient with a LAPS2 score greater than 110 is considered very sick, potentially in shock. 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).

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). Figure 2 depicts a few hypothetical patient pathways.

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). The EDIP2 score utilizes vital signs (e.g. temperature and oxygen saturation), vital signs trends, and laboratory tests from the past 24–72 hours (e.g., glucose levels), the COPS2, LAPS2 and CHMR severity scores, as well as patient diagnoses and demographics to determine a patient's EDIP2 score. More details can be found in Escobar et al. (2012) and Kipnis et al. (2016).

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. While there is some differentiation across ICUs (e.g. Medical versus Surgical ICU), the general practice in the study hospitals is that the boundaries between these units are relatively fluid. For instance, if the medical ICU is very full, a patient may be admitted to the surgical ICU instead. 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. We focus our study on patients who are admitted to a Medical service via the ED as this comprises the largest proportion of admitted patients ($> 60\%$). Additionally, there are limited standards for the care pathways for these types of patients, so that they can be highly varied, as compared to elective admissions and surgical cases. As such, these patients are more likely to experience variation in transfer decisions due to operational factors, such as the availability of resources, which we can leverage in our empirical approach to identify the impact of ICU transfer decisions on patients of varying severity. Specifically, because there are no established standards for which patients should be admitted to the ICU (Task Force of the American College of Critical Care Medicine, Society of Critical Care Medicine 1999), patients of similar severity may receive different care (e.g. ICU transfer versus no ICU transfer) due to random variation in ICU bed availability, which will allow us to estimate the causal effect of ICU transfer for these patients. We first eliminate 39 hospitalizations with unknown patient gender or missing inpatient unit code. Next, we eliminate 5,426 hospitalizations because there are inconsistent records for the inpatient unit entry/exit times (e.g. discharge took place prior to admission). 5,998 patients are missing unit admission and discharge times during their hospital stay. We drop 5,781 hospitalizations for patients who experience hospital transfers. 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. Out of all hospitalizations, 14.2% are admitted to the ICU at least once and 4.4% 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

	Min	Max	Mean	Median	Std. Dev.
First EDIP2	0.000	0.990	.012	.006	.022
Female (%)			53.80		
CHMR (%)	0.00	97.58	4.04	1.55	7.39
COPS2	0.00	306.00	45.00	29.00	43.03
LAPS2	0.00	274.00	73.24	69.00	36.51
Age	18.00	109.00	67.34	70.00	17.71

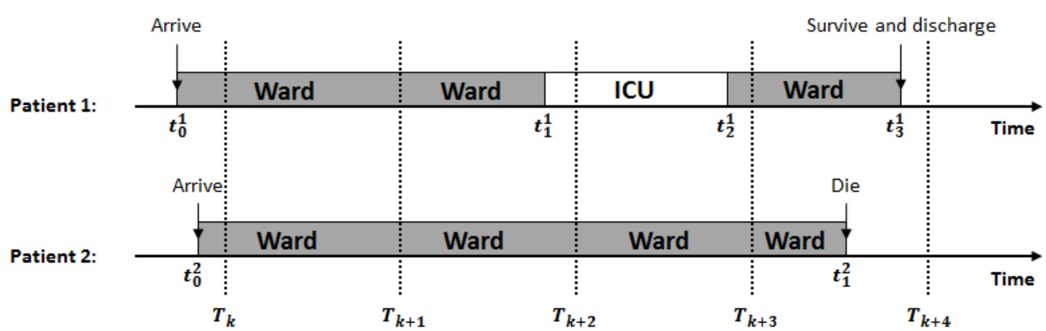


Figure 2 **Examples of patient pathways. Each T_i denotes a time when an updated EDIP2 score will be assigned to a patient if he/she is in the Ward or TCU. Note that there are exactly 6 hours between each EDIP2 assignment: $T_{i+1} - T_i = 6$.**

2.2. Actions

We define an EDIP2 decision epoch as the time comprised between an EDIP2 score measurement (at 4am, 10am, 4pm and 10pm) and the following 6 hours before the next EDIP2 score measurement takes place. For this, we require the patient to be in the ward or TCU because otherwise an EDIP2 score would not be recorded and this would not be an EDIP2 decision epoch. Each patient may have multiple EDIP2 decision epochs during his/her hospital stay. For example, in Figure 2 for Patient 1, there are three decision epochs: $[T_k, T_{k+1}), [T_{k+1}, T_{k+2})$ and $[T_{k+3}, T_{k+4})$. For Patient 2, there are four EDIP2 decision epochs: $[T_k, T_{k+1}), [T_{k+1}, T_{k+2}), [T_{k+2}, T_{k+3})$ and $[T_{k+3}, T_{k+4})$.

At the beginning of each of these epochs, we record whether the patient was transferred to the ICU in the following 6 hours (i.e., during the decision epoch) and call this an *action*. If, instead, the patient remains in the ward or TCU until the next EDIP2 measurement, we refer to this as *no action*. Thus, for Patient 1, if we consider the first EDIP2 decision epoch $[T_k, T_{k+1})$, there is no action. On the other hand, if we consider the second EDIP2 decision epoch $[T_{k+1}, T_{k+2})$, then there is an action. For Patient 2, there are 4 decision epochs and for each of them there is no 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 beginning of the EDIP2 decision epoch. In Figure 2, for Patient 1 the LOS for the first decision epoch would be $\tau = t_3^1 - T_k$; for the second decision epoch, it would be $\tau = t_3^1 - T_{k+1}$; and for the third decision epoch it would be $\tau = t_3^1 - T_{k+3}$. Table 2 summarizes the statistics for in-hospital mortality and hospital remaining length-of-stay considering the first EDIP2 decision epoch.

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

	Mean mortality	Mean LOS since first EDIP2 (hours)	Std. Dev.
All	3.2%	90.5	135.2
Transferred to ICU	9.5%	149.1	270.2
Never transferred to ICU	2.2%	81.0	93.4

3. Empirical Models and Approach

Our goal is to estimate the benefit of ICU admission for patients of different severity. 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: Physicians consider many factors when deciding whether to admit a patient to the ICU. While we will utilize our rich set of data 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 ICU admission results in *worse* patient outcomes. To address this concern, we utilize an instrumental variable approach.

Weak instruments: While instrumental variables can be effective at removing endogeneity biases, problems can arise if the instrument is not strongly correlated with the endogenous variable. 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). Additionally, the IV estimates based on weak instruments are highly sensitive to small violations of the exclusion restriction (Small and Rosenbaum 2008). To address this problem, we restrict the analysis to a cohort where our instrument exerts a much stronger influence on the endogenous variable, ICU admission.

Effect modification: Our goal is to estimate the causal effect of admissions to the ICU at different levels of the EDIP2 score. In other words, we need to assess how the effect of ICU admissions is modified by the severity of the patients as measured by the EDIP2 score. We use parametric statistical models for this purpose. It is important to make sure that there is sufficient overlap in the covariate distributions across levels of the instrumental variable, so that the predictions of the models are an interpolation and not an extrapolation; in doing so, the results will be less dependent on specific parametric assumptions (Rosenbaum 2010). Without this balancing of covariates, it is possible that a few, unrepresentative observations, could impart a large influence over the effect estimates (Imbens 2015, Rosenbaum 2016).

3.2. Design Choices to Strengthen the Instrument and Reduce Model Dependence

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 we find the instrument has a stronger effect on ICU admissions so that violations to the exclusion restriction are less likely. Second, we use recent advancements in multivariate matching to reduce model dependence in the outcome analyses. Naturally, these two choices will result in a smaller sample for analysis, but 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. Certainly, these gains come with the caveat that our findings will fundamentally apply to the matched sample in the night-time period.

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 (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 the KPNC hospitals included in our study, 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. We confirm that this is the case in our data (see Appendix A). In contrast to most studies in the empirical OM literature which tend to take the strength of an IV as given by the available data, we leverage the differential impact of ICU occupancy due to operational changes (i.e. staffing levels) on ICU admission by time of day to strengthen the IV. This allows us to obtain more robust effect estimates on the outcomes.

3.2.2. Multivariate Matching. In observational studies, matching methods are often used to adjust for covariates (Stuart 2010). 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). 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.1 for details.

3.3. Parametric Models

We now introduce the parametric models we use to estimate the potential benefits of ICU transfers for patients of varying severity.

In all of our models, we use 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 ICU admission, and 2) it must affect the outcome only through the treatment (the so-called "exclusion restriction" (Angrist et al. 1996)). To examine the first assumption, we use logistic regression to see how ICU occupancy impacts the ICU transfer decision when adjusting for several 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. Assuming that observed patient risk factors are reasonable proxies for unobservable risk measures, ICU occupancy is unlikely to be related to unobservable risk measures.

We utilize the IV framework in Angrist et al. (1996) where an IV is conceptualized as an "encouragement" to receive treatment that affects the outcome only through the treatment. In this framework, the IV takes two levels—encouragement and discouragement—which correspond to non-busy and busy ICUs in our setting. Formally, we define an ICU to be "busy" when the ICU occupancy is above the 90^{th} percentile of its occupancy distribution. An ICU is "not-busy" when the ICU occupancy is below 70^{th} percentile of its occupancy distribution. Following Yang et al. (2014), we do not use observations with ICU occupancy between the 70^{th} and 90^{th} percentiles. 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 $(70^{\text{th}}, 90^{\text{th}})$ percentiles will be dropped. Comparing with other potential cutoffs, the $\{70^{\text{th}}, 90^{\text{th}}\}$ 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 A.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 discussion about Figure 2 in Section 2.3). 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 admission decision, Z_i be the instrument of ICU busyness, and X_i be patient, hospital and seasonality controls that include patient demographics (age, gender), severity scores (EDIP2, CHMR, COPS2, LAPS2), 38 disease categories, and other indicators for hospital, day of the week, and month (see Table 12 in the Appendix for more details). Additionally, we define T_i^* as the corresponding latent variable capturing the likelihood of ICU transfer. We have that

$$T_i = \mathbb{1}\{T_i^* > 0\} \qquad \text{where} \qquad T_i^* = X_i^T\beta_1 + \beta_2 Z_i + \epsilon_i \tag{1}$$

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

where ϵ_i and η_i are assumed to be correlated normal random variables. We take a natural logarithmic transformation for the hospital length-of-stay because its distribution is skewed (see Table 2). Our estimates include 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 two-stage-least-squares (Wooldridge 2010). The treatment equation is the same as before in equation (1). For the binary outcome *Mortality*, the second equation is

$$Y_i = \mathbb{1}\{Y_i^* > 0\} \qquad \text{where} \qquad Y_i^* = X_i^T \beta_5 + \beta_6 T_i + \nu_i$$

and (ϵ_i, ν_i) follows a bivariate normal distribution with correlation coefficient ρ. 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, similar to Kim et al. (2015), we include a covariate that measures the average occupancy of every unit a patient visits during his hospital stay. This is because there is evidence (e.g. Kuntz et al. (2014)) that occupancy levels can impact a patient's outcome, which could potentially invalidate our instrument. We find that our instrumental variable, ICU occupancy during the EDIP2 epoch, has a low correlation with the average occupancy experienced by a patients with a correlation coefficient of -0.168.

4. Empirical Results

In this section, we present and discuss our main empirical results. First, we examine the impact of our study design choices in terms of strengthening the instrument and reducing model dependence. Second, we present our effect estimates. Next, we compare the results to those obtained under other common study designs. Finally, in order to provide a better understanding of the population of patients to which the results in principle generalize, we describe our matched sample and compare it to the full patient sample.

4.1. Design Choices

In our study, we make two basic design choices to make the instrument stronger and reduce model dependence. One choice 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). The other choice involves confining the study to the night-time period, when the instrument is considerably stronger. In our study, we solved the near-far matching problem using integer programming as in Zubizarreta et al. (2013). We found matched groups of patients with similar or balanced covariate distributions for important prognostic factors such as age and the EDIP2 score, and dissimilar levels of encouragement to receive the treatment (ICU admission). More specifically, we matched patients that faced non-busy ICUs (encouraged patients) to patients that faced busy ICUs (discouraged patients) with a 1:5 matching ratio, matching in total 85,208 observations (15,149 discouraged patients; 88% of all the available discouraged patients before matching in the data set). See Appendix A.1 for further details on the near-far matching implementation using integer programming. Tables 13–15 in Appendix A.2 summarize covariate balance after matching for patient- and hospital-level covariates as well as for other important seasonality covariates. The tables show that after matching the covariates are well balanced as per common standards in the causal inference literature. As a result, the effect estimates reported below are less sensitive to model misspecification (Imbens 2015).

To evaluate the strength of the instrument after matching night-time decision epochs (instead of using the full sample), 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 night-time matched sample has only 40% of the number of observations in the first whole-day EDIP2 sample, we see the coefficient estimate for the ICU occupancy (*IV*) is much larger and the p-value is lower. Additionally, 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 IV in the whole-day full sample and the first night-time IV after match in probit regression models

	Sample Size	*IV*	(Std. Err.)	P-val.	Pct. Incr. in Prob (*Admit*)
Whole-day full sample	168,351	0.098	(0.039)	0.012	**34%**
Night-time matched sample	84,870	0.201	(0.072)	0.005	**95%**

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

Table 4 summarizes the estimation results for the mortality and remaining LOS models after night-time matching. Moreover, we present a number of robustness checks which considers alternative IV definitions and additional covariates in Appendix A.3. We find our empirical results robust to these alternative specifications. Note that because we are using full MLE to estimate these models, the coefficients in the first-stage are slightly different than those of Table 3.

For both outcomes, the instrument is highly significant at the 1% level. Being encouraged for ICU transfer (when the ICU is not busy) increases the probability of transfer by 97% on average. We estimate that ICU transfer is associated with a reduction in the average LOS by 34 hours (95% CI: [-40, -31] hrs). We also find that ICU transfer has a highly significant impact in reducing mortality risk: ICU transfer reduces the average estimated in-hospital mortality from 2.62% to 0.06% (95% CI: [-2.59%, -2.53%]). Note that our estimates are for the average effect. While 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. In practice, it would rarely be the case that very low severity patients are transferred to the ICU. In fact, most medical literature on rapid response teams involves only checking on the patients and not necessarily admitting them, and therefore, the average effect documented in this literature is typically smaller. That said, the estimated benefits seem quite large. This may be in part due to Do-Not-Resuscitate (DNR) orders, so that those who are transferred to the ICU and who conform to our instrument are the ones who can actually benefit from ICU care. We cannot estimate the impact of ICU transfer for patients who would never be admitted to the ICU (either being too sick or too well), regardless of ICU congestion.

Table 4 Estimation results using the night-time IV after matching

Y	*IV* (SE)	Pct. Incr. in Prob (*Admit*)	*Admit* (SE)	$\Delta\bar{Y}$	95% CI
Mortality	0.203** (0.067)	97%	-1.665*** (0.162)	-2.56%	[-2.59%, -2.53%]
Remaining LOS	0.203** (0.073)	97%	-0.841** (0.281)	-33.81 hrs	[-39.55, -30.89]

, * Significance at the 1%, 0.1% levels respectively

Our results suggest that ICU transfers can improve patient outcomes on average. We will utilize these results to obtain the estimated mortality and remaining length-of-stay (LOS) when transferred or not transferred to the ICU for patients of varying EDIP2 severity to calibrate a simulation model in Section 5.

4.3. Comparison to Other Study Designs

In the current analyses, we made a number of study design choices to increase the reliability and robustness of our empirical analysis. These choices included focusing the analysis to the night-time period and using optimal multivariate matching with an IV. In an effort to understand better the implications of such design choices, we compare our approach to two common approaches: (i) using an ordinary least squares approach without using an IV nor night-time matching, and (ii) using an IV approach but without night-time matching. These results are summarized in Table 5.

As we can see, under (i), ICU admission is estimated to result in worse patient outcomes. This effect is likely to be biased due to endogeneity, since sicker patients are more likely to be admitted to the ICU and at the same time suffer worse health outcomes. Under (ii), we see that the estimated effect of ICU admission on LOS is not statistically significant, but it is under (iii). We believe the lack of significance in (ii) may be due to weak instruments. Specifically, the magnitude of the estimate and the p-value of the IV is less

Table 5 Estimated regression coefficients (i) without IV nor night-time matching, (ii) with IV but no night-time matching, (iii) (our approach) with IV and night-time matching.

Model	Outcome Measure	Estimated Coefficients (s.e.) IV: ICU Occupancy	ICU Admission
i	Mortality		0.592***(0.062)
	LOS		0.490***(0.028)
ii	Mortality	0.095*(0.039)	-0.814***(0.256)
	LOS	0.097*(0.040)	0.061 (0.112)
iii	Mortality	0.203**(0.067)	-1.665***(0.162)
	LOS	0.203**(0.073)	-0.841**(0.281)

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

than that of (iii). Additionally, the partial F-statistic is 8.638, which is below the rule-of-thumb of 10, while under (iii), using both the IV and night-time matching, the IV is significant at the 1% level with a partial F-statistic of 11.029. As such, we believe that our estimation results are more robust to unobservable biases due to our design choices.

4.4. Description of the Night-Time Matched Sample

In order to design a study that is less sensitive to model misspecification and violations to the exclusion restriction (Angrist et al. 1996), we confined our study to the night time and used multivariate matching (Zubizarreta et al. 2013). Naturally, this implies that without further, untestable, modeling assumptions the results will fundamentally apply to the night time. Here, we follow the work of Imbens (2010) and Rosenbaum (2010) and emphasize internal validity over external validity in order to provide more reliable evidence of the causal effect of ICU admission at different levels of the EDIP2 score. As such, it is not immediately obvious if/how our empirical findings will extend to other times during the day.

The night-time analysis is important in two ways. First, even if our results only apply to the night time, using these rigorously estimated results to calibrate a simulation model would allow us to develop an understanding of the potential benefits of proactively admitting patients to the ICU *during the night*. This is valuable from a managerial standpoint, because of the fact that night-time physician staffing tends to be much lower than during the rest of the day, which makes having an automated early warning system to inform proactive ICU admissions especially useful. Second, as discussed next, we believe that it is possible that our results may generalize to admission of patients during non-night time decision epochs.

Table 6 summarizes the means of the risk covariates for the full sample and night-time matched sample. We quantify the differences in means using standardized differences (Std. Dif.), which are simply the difference in means between the two samples standardized by the average standard deviation of the two samples. We can see that for all risk covariates, except for the EDIP2 score, the absolute value of the standardized differences between the full and matched sample are well less than 0.1, suggesting that these samples are quite similar (Rosenbaum and Rubin 1985). The difference in EDIP2 scores lends more evidence to our argument that patients are more likely to be admitted to the ICU at night, thereby increasing the strength of our instrument. We found a similar pattern for 30 other comorbidity and seasonality covariates.

We believe the difference in the strength of the instrumental variable is likely due to the differences in operational practices between the night time and the entire day, rather than the difference between patient populations as the samples appear very similar on all other dimensions. That is, physician staffing levels are lower during the night time, making lack of ICU congestion more likely to act as an encouragement for ICU transfer, thereby increasing the strength of the IV. Because of this, it is possible the results from the night time cohort may generalize to the entire day. Of course, this assumes the populations are similar based on unobservables as well. Since we cannot completely rule out the possibility that there are differences between the patient populations during the night time and the entire day, it is possible the empirical findings will not extend to other times during the day.

Table 6 Comparison of patient characteristics in full sample versus matched sample via standardized difference

	Full	Matched	Std. Dif.
Age	67.35	67.73	−0.02
Female (%)	53.81	54.57	−0.01
EDIP2	0.012	0.007	0.30
CHMR	0.040	0.037	0.05
COPS2	45.07	44.92	0.00
LAPS2	73.25	72.41	0.02

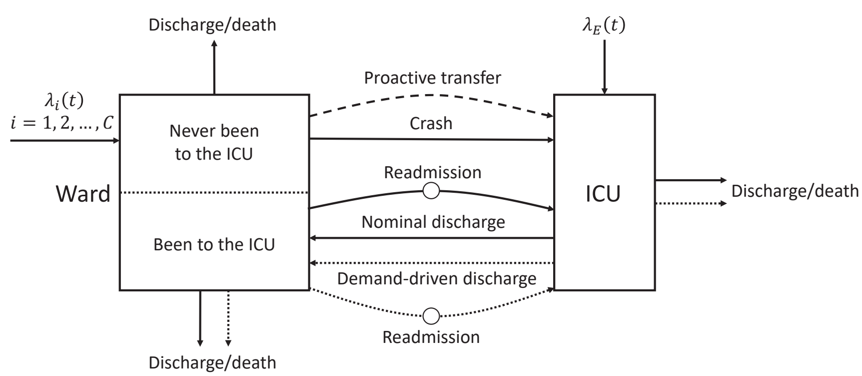
5. System Level Effect of Proactive Admissions

Thus far, we have focused on the impact of ICU transfer on *individual* patients of varying EDIP2 risk levels. Our empirical findings provide evidence that such transfers could improve patient outcomes (reducing mortality risk and LOS) and the magnitude of the impact varies depending on a patient's severity. Given these improvements in patient outcomes, it is conceivable that proactively admitting patients may reduce ICU congestion. However, given the limited ICU resources, physicians naturally have concerns about needlessly creating ICU demand. Specifically, by proactively transferring patients 'before they *really* need it', the near-term ICU congestion will increase, which could create access issues for other, more critical patients who may arrive in the near future. However, if this patient will ultimately need ICU care later and will require increased resources, the short-term increase in congestion could have long-term benefits. It remains to understand which scenario is more likely to occur. To do this, 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. Our simulation model is depicted in Figure 3. 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 $\mathbf{N}$.

Figure 3 **Simulation Model**



Patients can arrive at the ICU as transfers from the ward or via an external arrival stream of **Direct Admits** (e.g. directly from the ED). Recall that our analysis focuses on patients admitted to a medical service (rather than surgical service which can be impacted by elective surgical schedules), so we model the arrivals of the direct admits as a non-homogenous Poisson process with rate $\lambda_E(t)$, which has been shown to be a good model for patient arrivals (Kim and Whitt 2014). We assume these patients have a hospital LOS which is lognormally distributed with mean $1/\mu_E$ and standard deviation σ_E. Moreover, a proportion $p_E \sim f_{p_E}(p)$ of the patient's hospital LOS is spent in the ICU, where $f_{p_E}(p)$ is a known probability mass function (pmf) with finite support on $[0, 1]$. The remaining portion of their hospital LOS is spent in the ward. These patients survive to hospital discharge with known probability $1 - d_E$.

The second way patients can be admitted to the ICU is via transfer from the wards. We refer to these patients as **Ward Patients**. We consider two types of ward patients: (a) those who have been to the ICU and (b) those who have not. We first describe the dynamics of ward patients who have not been to the ICU. To capture the varying level of severity for these patients, we consider C patient classes. Patients of type i

arrive at the ward according to a non-homogeneous Poisson process with rate $\lambda_i(t), i = 1, 2, \ldots, C$. Every 6 hours, a patient's EDIP2 score is updated, so patient i's class will now be $j \in \{1, 2, \ldots, C\}$. Alternatively, three other possible events may occur: the patient may 1) '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 requires immediate ICU transfer due to *crashing* on the ward, he will have a hospital LOS which is lognormally distributed with mean $1/\mu_C$ and standard deviation σ_C. We assume that a proportion $p_W \sim f_{p_W}(p)$ of the patient's hospital LOS is spent in the ICU, where $f_{p_W}(p)$ is a known pmf with finite support. The remaining $1 - p_W$ proportion is spent in the ward, as a patient (a) who has been to the ICU. Crashed patients survive to hospital discharge with probability $1 - d_C$.

Direct admits and patients who crash on the ward receive the highest priority for ICU admission. If there are no available ICU beds at the time of arrival (or crash), the current ICU patient with the shortest remaining service time will be "demand-driven discharged", i.e., he/she will be discharged in order to create space to accommodate the incoming, more severe patient (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)). Note that, one could also consider incorporating rerouting direct-admits or crashed patients to other hospitals if all ICU beds are occupied, rather than initiating a demand-driven discharge. However, such inter-hospital transfers are incredibly rare—especially for critically ill patients—at KPNC. Still, we will examine the state of patients who are demand-driven discharged to make sure we are not too aggressive in discharging critical patients.

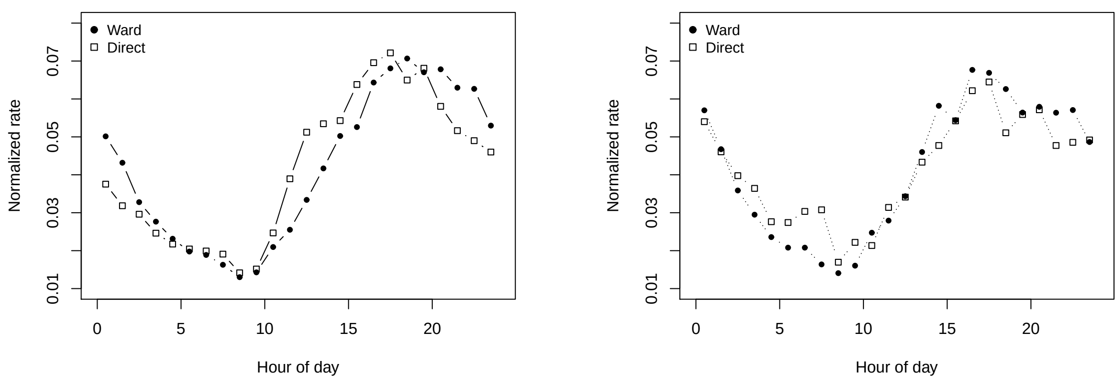
In principle, any patient in the ward can be *proactively transferred* to the ICU at each EDIP2 decision epoch. Such proactive transfers can only occur if there is an available ICU bed for the transferred patient. If there are not enough available beds in the ICU to accommodate all proactive ICU transfer requests, the most at risk patients (those with the highest EDIP2 score) will be given priority. If a patient from EDIP2 group i is proactively transferred to the ICU, his hospital LOS is lognormally distributed with mean $1/\mu_{A,i}$ and standard deviation $\sigma_{A,i}$. Similar to the crashed ward patients, we assume that a proportion $p_W \sim f_{p_W}(p)$ 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 $r_{A,i}$. Otherwise, it is r_D. This proactively admitted patient will be a type (a) patient who has been to the ICU for the proportion $1 - p_W$ of his/her LOS not spent in the ICU.

Note that for type (a) patients (those in the ward who have been to the ICU), their mortality risk, readmission risk, and LOS are dictated by how they got to the ICU—i.e., as a direct admit, a crashed patient, or a proactively admitted patient. We do not allow these patients to be proactively admitted to the ICU.

5.2. Model Calibration

We now calibrate our simulation model using the data described in Section 2 and our empirical results from Section 4. Figure 4 depicts the normalized empirical arrival rates of all patients to the ward and directly admitted to the ICU in weekends versus weekdays. The empirical hourly arrival rates are determined using 12 months of data from all 21 hospitals and are normalized via a multiplicative factor so that the average number of arrivals per day is equal to 1. We will scale these normalized arrival rates to vary the load on the system, which allows us to maintain the same relative hourly demand from the ward and direct admits.

Figure 4 **Normalized arrival rates of Direct Admits and Ward Patients. Normalized so that the average number of arrivals (direct admits + ward patients) per day is equal to 1.**



(a) Weekday arrivals (b) Weekend arrivals

5.2.1. Direct Admits. We start by considering external arrivals. We use our full dataset from KPNC to calibrate the average hospital LOS, standard deviation of the hospital LOS, the mortality rate, ICU readmission rate, and the proportion of hospital LOS spent in the ICU. We use sample averages to determine these parameters which are summarized in Table 7. Note that we use the empirical distribution for p_E (see Figure 11 in the Appendix). Because patients who are demand-driven discharged exhibit higher readmission rate than those naturally discharged, we set r_D to be 15% larger than r_E (Kc and Terwiesch 2012, Chan et al. 2012). If a demand-driven discharged patient is readmitted to the ICU, we set his hospital LOS to be 15% longer than the nominal LOS_E as suggested by Kc and Terwiesch (2012). Note that the parameters μ_E and σ_E are determined by accounting for the expected number of readmissions, so that $1/\mu_E = E[LOS_E](1-r_E)$ and $\sigma_E = \frac{\text{std. dev. } LOS_E}{E[LOS_E]\mu_E}$.

Table 7 Direct Admit parameters. Note that the readmission rate for demand-driven discharged patients is calibrated to be 15% greater than the nominal readmission rate.

	Mean	95% CI
d_E (%)	9.41	[9.12, 9.69]
r_E (%)	15.76	[15.43, 16.10]
$E[p_E]$ (%)	50.79	[50.49, 51.09]
$E[LOS_E]$ (days)	6.52	[6.45, 6.58]
(std. dev. LOS_E)	(6.78)	
r_D	$1.15 \times r_E$	

5.2.2. Ward Patients. We now turn our attention to the ward patients who may be admitted to the ICU after crashing or via a proactive transfer. In choosing the number of EDIP2 groups and the size of each group, we must balance having more groups to enable more flexibility in various transfer policies versus having enough samples within each group to reasonably estimate transition probabilities between each EDIP2 group and the absorbing states (crash, death in the ward, discharge alive). With that in mind, we elect to have 10 EDIP2 groups ($C = 10$) for illustrative purposes. Additionally, we divide the top 10% of patients into 5 groups and the bottom 90% into 5 groups, in order to enable more flexibility for proactive transfers of the most severe patients. Table 8 summarizes the partitioning of these 10 groups.

Table 8 Summary statistics of ten EDIP2 groups .

Group	Range of EDIP2	Mean	Number of observations	Proportion
1	[0.000, 0.002]	0.002	28,051	17.6%
2	[0.003, 0.004]	0.003	32,358	20.3%
3	[0.005, 0.007]	0.006	31,903	20.0%
4	[0.008, 0.011]	0.009	23,819	14.9%
5	[0.012, 0.023]	0.016	27,002	16.9%
6	[0.024, 0.027]	0.025	3,584	2.2%
7	[0.028, 0.032]	0.030	3,130	2.0%
8	[0.033, 0.040]	0.036	3,138	2.0%
9	[0.041, 0.057]	0.048	3,189	2.0%
10	[0.058, 1.000]	0.107	3,221	2.0%

We use our full dataset from KPNC to calibrate the Markovian transition matrix $\mathbf{T} \in \mathbb{R}^{10\times 13}_{[0,1]}$ (see Appendix B.1). We can then determine the nominal probability of crashing, dying in the ward, and surviving to hospital discharge when no proactive transfers are done as predicted by our Markov Chain based simulation model. We find that the mortality rate on the ward is 1.93%, which is comparable to the empirical rate of 2.2% reported in Table 2. We also conduct a sensitivity analysis over 1,000 different Markovian transition matrices selected uniformly at random over the 95% confidence intervals of the estimated transition matrix. The expected mortality rates for these transition matrices range from 1.04–3.17%.

We leverage our empirical findings from Section 4 to calibrate the mortality risk and hospital LOS of a ward patient depending on whether he/she is proactively admitted to the ICU or admitted after crashing. For each patient in EDIP2 group i, we can utilize our empirical results to predict the probability of death and remaining hospital LOS if the patient is admitted to the ICU at their given EDIP2 score (i.e., an *action* is taken in the current EDIP2 decision epoch). We use the average predicted probability and LOS for each EDIP2 group i to calibrate the probability of death and LOS for patients who are proactively admitted to the ICU. The remaining parameters to calibrate are the probability of death and mean remaining hospital LOS if a patient crashes. For patients who are not proactively admitted, they will stay in the ward for a random amount of time. These patients will eventually leave the ward either by 1) dying in the ward, 2) being discharged alive from the ward, or 3) crashing. The three possible absorbing states and the parameters for crashed patients will determine the expected LOS and probability of death if not proactively admitted as given by our Markov Chain based simulation model. We solve an optimization problem (described in Appendix B.2) to determine the crashed parameters with an objective of minimizing the relative squared error between the predicted probability of death (LOS) from our empirical model when there is *no action* taken at that EDIP2 score versus the probability of death (LOS) indicated by our Markov Chain model.

Similar to the direct admits, we use the empirical distribution for the proportion of hospital LOS which is spent in the ICU (p_W) (see Figure 11 in the Appendix). We use the proportion of all patients who are transferred to the ICU from the ward who visit the ICU more than once during the same hospitalization as the ICU readmission rates of crashed patients. Finally, we set the ICU readmission rates for proactive patients, $r_{A,i} = \beta \times r_C, \forall i$. For our main simulations, we set $\beta = 1$, but we also run robustness checks for $\beta \in (0, 1)$. Similar to direct admits, we appropriately scale the LOS by the readmission rates. To calibrate the standard deviations for each LOS parameter, we use the same coefficient of variation (0.81) as determined by the LOS across all ward patients. Tables 9 and 10 summarize the parameters for ward patients.

5.3. Proactive ICU Transfer Policies

We consider a number of different ICU transfer policies. To start, we assume that proactive transfers can only happen during the night-time decision epoch. This is because our empirical results fundamentally apply to the night time sample as described in Section 4.4. In Appendix B.3, we relax this constraint and consider the potential benefits of proactive transfer if proactive transfers can occur at any decision epoch and under the assumption that our empirical results generalize to other parts of the day.

Table 9 Common Ward Patient parameters. Note that the readmission rate for demand-driven discharged patients is the same for ward patients as direct admits. Parameters which are induced from other parameters do not include confidence intervals.

	Mean	95% CI
r_C (%)	16.88	[16.12, 17.63]
$E[p_W]$ (%)	46.92	[46.18, 47.65]
d_C (%)	57.28	
$E[LOS_C]$ (days)	15.09	
(std. dev. LOS_C)	(12.19)	
r_D	$1.15 \times r_E$	
r_A	$\beta \times r_C (\beta = 1)$	

Table 10 Expected mortality and LOS under proactive ICU transfers for 10 EDIP2 groups

EDIP2 Group	Mortality (%) $d_{A,i}$	$LOS_{A,i}$ (days) Mean	std. dev.
1	0.01	0.85	0.68
2	0.02	0.91	0.74
3	0.04	0.97	0.78
4	0.05	1.04	0.84
5	0.11	1.17	0.95
6	0.18	1.36	1.10
7	0.28	1.45	1.17
8	0.39	1.57	1.27
9	0.70	1.85	1.50
10	6.84	3.77	3.04

We define a **Static Threshold Policy** by threshold T_{EDIP2}. Any patient in EDIP2 group $i \geq T_{EDIP2}$ will be proactively transferred if there are available ICU beds. 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 $T_{EDIP2} \in \{1, \ldots, 11\}$, where $T_{EDIP2} = 11$ is the case where no proactive transfers are done. For comparison, we also consider a **Random Policy**, where, for every available ICU bed, we select a patient uniformly at random in the ward to proactively admit into the ICU (regardless of EDIP2 score). We will also consider **State-dependent Threshold Policies** in Section 5.5.

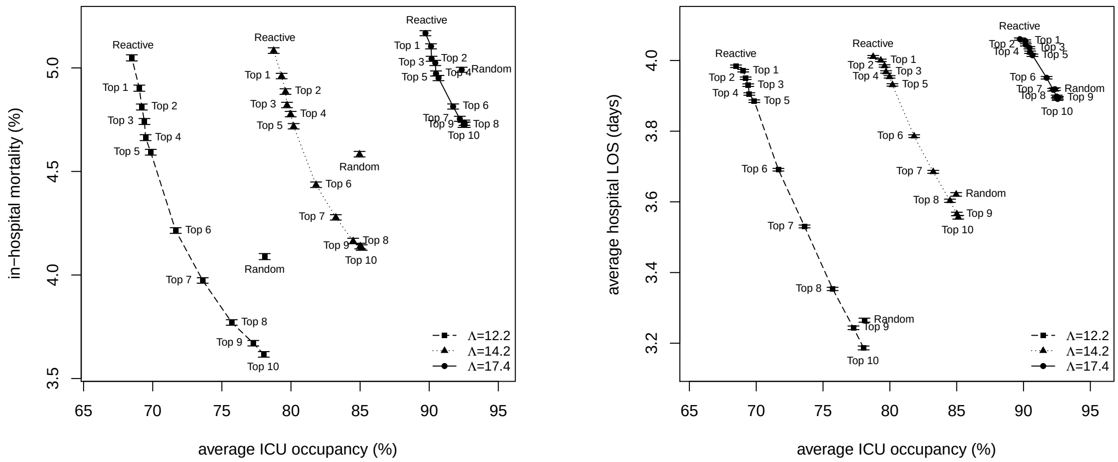
5.4. Results

Our baseline simulation considers an ICU with $\mathbf{N} = 15$ beds and an aggregate (ward patients and direct admits) arrival rate of 12.2, 14.2, and 17.4 patients/day. Patients can only be proactively transferred to the ICU during the night-time period. We simulate 1 year with 1 month of warm-up, over 2,000 iterations.

Figure 5 shows the in-hospital mortality rate and average hospital LOS versus ICU occupancy level under the various proactive transfer policies. Because proactive transfers can reduce the likelihood of death and average LOS, we see that more aggressive proactive ICU transfers can simultaneously reduce mortality rates and average LOS. However, these reductions come with an increase in ICU occupancy. For instance, with a daily arrival rate of 14.2 patients/day, the nominal ICU occupancy without any proactive transfers (labeled 'Reactive') is 78.75%. This increases to 80.19% when proactively admitting the top 5 EDIP2 groups and all the way to 85.12% when proactively admitting all 10 EDIP2 groups. Thus, there is merit to physicians' concerns about ICU congestion, but it also comes with the benefit of reduced mortality and LOS.

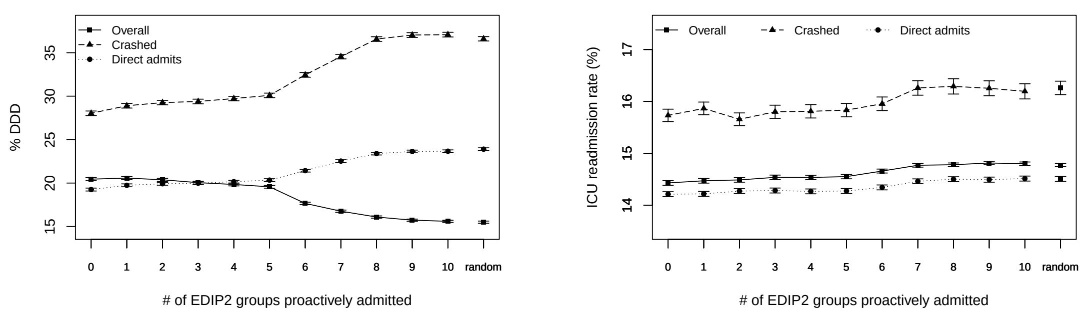
As seen in Figure 6, the impact of increased congestion also translates to other adverse events—demand-driven discharges and readmissions. We calculate the demand-driven discharge rate as the fraction of all ICU admissions which are discharged due to incoming demand. Similarly, we calculate the readmission rate

Figure 5 **In-hospital mortality rate and mean hospital LOS under various proactive ICU transfer policies, with 95% confidence intervals. ICU size $\mathrm{N} = 15$. Λ = daily arrival rate. Proactive transfers can only take place at night.**



(a) In-hospital Mortality Rates (b) Hospital LOS

Figure 6 **Adverse event rates under various proactive ICU transfer policies, with 95% confidence intervals. ICU size $\mathrm{N} = 15$. $\Lambda = 14.2$ patients/day. Proactive transfers can only take place at night.**



(a) Demand-driven discharge rates (b) Readmission rates

as the fraction of all ICU admissions that are followed by another ICU admission prior to hospital discharge (i.e. leaving the system). Interestingly, the differences between the demand-driven discharge (readmission) rates are not statistically significant when comparing no proactive transfers (Reactive) to proactively admitting the top five severity groups ($T_{EDIP2} = 6$). Moreover, we find that across all policies, patients who are demand-driven discharged stay in the ICU for 80-85% of their ICU LOS, which suggests that these patients may be sufficiently stable for such transfers (e.g. Lowery (1992)). Still, being very aggressive with proactive transfers could result in worse care and outcomes. While the aggregate demand-driven discharge rate goes down with more aggressive proactive transfers (because there are simply many more ICU admissions), the rate for the most critical patients—direct admits and crashed patients—increases. A similar (but smaller in magnitude) effect exists for ICU readmissions.

Our results suggest that some proactive transfers could help improve quality of care at the system level, but it must be done carefully. We see that proactively admitting up to 10% of ward patients ($T_{EDIP2} = 6$) can improve mortality and LOS without substantially increasing ICU congestion, demand-driven discharges, or ICU readmissions. However, proactively admitting 26% or more of ward patients ($T_{EDIP2} < 6$) can increase adverse events. Unsurprisingly, the impact of proactive admissions (and the resulting increased ICU congestion) on readmissions and demand-driven discharges depends highly on the system load. Figure 7(a) is an analog to Figure 6(a) and depicts the impact of being more aggressive with proactive transfers on demand-driven discharge rates for different arrival rates, which impacts the average ICU occupancy. We denote this as $\hat{\rho}$ when there are no proactive transfers. We can see that when the system is very lightly loaded (e.g. $\hat{\rho} \leq 0.3$), proactively admitting all 10 EDIP2 groups does not increase demand-driven discharge rates. However, as the system load increases, more aggressive proactive transfers results in an increase of adverse outcomes. Figure 7(b) summarizes when this increase begins and we find that proactively admitting more than the top 5 EDIP2 groups consistently comes with the cost of more demand-driven discharges, thereby supporting our initial observation that proactively admitting the most severe patients could save lives without needlessly clogging the ICU.

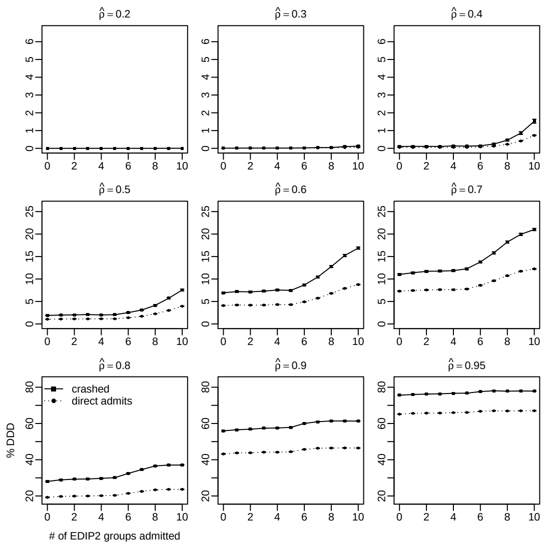
Note that in all of our experiments, the random policy is Pareto dominated by the static threshold policies. This is true even when we consider other random policies which aim to proactively transfer a similar number of patients as under the static threshold policies. Appendix B.3 provides additional simulation results which demonstrate the robustness of our main insights.

5.5. State-Dependent Policies

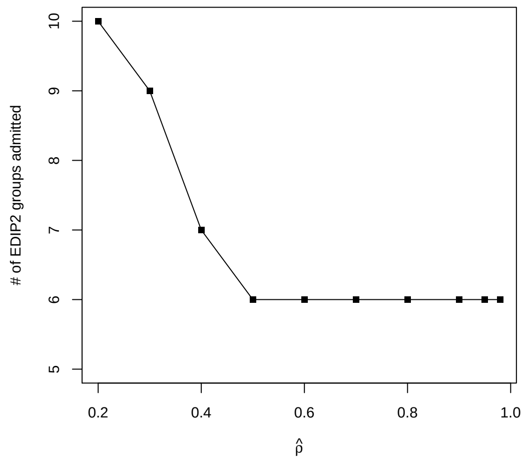
We also consider a modification of the Static Threshold policy, where instead we consider state-dependent thresholds (e.g. Altman et al. (2001)). For these experiments, we focus on the baseline scenario of $\mathbf{N} = 15$ beds and $\Lambda = 14.2$ patients/day.

As our model incorporates many features (e.g., demand-driven discharges, readmissions, etc.), solving a dynamic program for the optimal thresholds is computationally prohibitive. As such, we consider a set of state-dependent thresholds and select the best one via simulation. The family of policies we consider are parameterized by EDIP2 thresholds, $T_1 \geq T_2$, and a bed threshold, B. Suppose there are b available ICU beds. Then, 1) If $b < B$, *Proactively Admit* patients in EDIP2 group $i \geq T_1$. 2) If $b \geq B$, *Proactively Admit* patients in EDIP2 group $i \geq T_2$. 3) Otherwise, the patient will remain on the ward. If $T_1 = T_2$, we recover the static threshold policy. We can also generalize this to more than 2 thresholds.

Figure 7 **ICU size $N = 15$. $\hat{\rho}$ indicates the average ICU occupancy induced by arrival rates $\Lambda \in \{3.5, 4.8, 6.6, 8.7, 10.4, 12.2, 14.2, 17.4, 22.0\}$ patients/day.**



(a) Demand-Driven Discharge rates depending on how many EDIP2 groups can be proactively transferred to the ICU for various arrival rates.



(b) Point where more proactive transfers lead to increases in demand-driven discharge rates as a function of average ICU occupancy.

We use simulation and an exhaustive search over all possible state-dependent threshold policies with two thresholds which can proactively admit 1, 2, . . . , up to 6 EDIP2 groups. Because proactive transfers reduce mortality and LOS for all patients, aggressive proactive transfers will improve both of these measures and we find that no state-dependent policy outperforms the static threshold policy in mortality and LOS. However, we do find that the demand-driven discharge and readmission rates for crashed and direct admits can improve by allowing state-dependent policies. Table 11 summarizes the relative difference in outcomes where we report the 'best' state-dependent policy as the one that improves upon the static threshold policy in demand-driven discharges and readmissions, but also has the lowest mortality rate and mean LOS. Table 18 in the appendix provide the results when allowing 3 and 4 thresholds. We find that in some cases (proactively admitting the top 2 EDIP2 groups) the state-dependent policy can have statistically significant improvements in readmissions and demand-driven discharges, while the mortality rate and LOS are statistically equal. In some instances (e.g. Top 6), improvements in demand-driven discharge and readmission rates comes at the expense of increases in mortality rates and LOS. That said, these differences are all less than 5.31% from the static threshold policy, with an average of less than 1.01%. In a 15-bed ICU, this amounts to approximately reducing by 7 demand-driven discharges and 2 readmissions per year. Thus, we find that while state-dependent policies may be able to improve patient outcomes, the improvement is very small. As static threshold policies are easier to convey to clinicians and implement in practice, we find that the slight gains achieved with state-dependent policies may not be worth the added complexity.

Table 11 **Percentage differences between the best 2-threshold state-dependent policy and static policy**

# of groups	Mortality	LOS	$\text{DDD}_{\text{crashed}}$	$\text{DDD}_{\text{direct admit}}$	r_{crashed}	$r_{\text{direct admit}}$
Top 1	0.99*	0.07	-1.62*	-1.65*	-0.34	-0.06
Top 2	0.05	-0.03	-1.83*	-1.82*	0.19	-0.35
Top 3	1.23*	0.17*	-2.75*	-2.59*	-0.52	-0.26
Top 4	0.00	0.00	0.00	0.00	0.00	0.00
Top 5	0.53*	0.04	-1.21	-1.40*	-0.41	-0.19
Top 6	1.22*	0.84*	-5.31*	-5.22*	-0.48	-0.39

*: $p < 0.05$ difference in means based on t-tests

5.6. Estimated Transfer Policies Used in Practice

We estimate the current ICU transfer policies used in practice at two representative hospitals whose 99th percentile of the ICU occupancy distribution is 15 beds. As in Section 5.5, we consider state-dependent threshold policies. We consider the following probit regression model to estimate the thresholds from the data. For each patient i at EDIP2 alarm time t, let occ_{it} be the number of ICU beds occupied, $\kappa_{occ_{it}}$ be the threshold of admission as a function of the current ICU occupancy, X_{it} be the same control variables used in the empirical analysis excluding the EDIP2 score, and $\xi_{it} \sim N(0,1)$.

$$Admit_{it} = 1\{\beta_{EDIP2} EDIP2_{it} + X_{it}\theta + \xi_{it} \geq \kappa_{occ_{it}}\} \tag{3}$$

Because the admission threshold $\kappa_{occ_{it}}$ can change at any ICU occupancy level, we enumerate over all possible combinations of the number and location of occupancy level thresholds, and choose the model with the lowest Bayesian information criterion (BIC) to be the estimated empirical admission policy to obtain a parsimonious model that fits the data. We find the best fit model for both hospitals to be a static policy. The thresholds of ICU transfer (as measured by the EDIP2 scores) are 0.543 and 0.327 for the two hospitals, regardless of the ICU occupancy. Note that both thresholds fall in the top EDIP2 group (Table 8). Therefore, the estimated admission policies at both hospitals correspond to admitting only the most severe patients. As we have seen in Figure 5, proactively transferring the top 5 EDIP2 groups (instead of just the top) helps to reduce both the in-hospital mortality rates and the average LOS in hospital without significantly effecting demand-driven discharge rates and ICU readmission rates. Thus, there are potential benefits to extending the current ICU transfer practice to be more aggressive.

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 EDIP2 score was developed to predict the likelihood a patient will 'crash' and require ICU care. In this work, we empirically estimate the impact of ICU admissions on patient outcomes for patients with varying severity, as measured by the EDIP2. Using a high fidelity simulation model, 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 high ICU congestion so that patients are more likely to be demand-driven discharged and/or require ICU readmission. While some gains can be achieved by allowing for more complex transfer policies, such as those where the severity of patients to proactively transfer depends on the number of ICU beds available, we find the difference in outcomes to be minimal. Thus, it may be more reasonable to focus on using simple threshold policies which are desirable for practical implementation.

Our simulation model has been calibrated from our empirical findings and our extensive dataset. Certainly, the insights generated from the simulation study are highly dependent on the reliability of our empirical results and the fidelity of the data. As we are using a very large data set from multiple hospitals and because we make a number of important design choices to increase the reliability and robustness of our empirical analysis, we believe the risks of misspecification are small. While we have run a number of sensitivity analyses to test the robustness of our results, we must acknowledge that if there are other first order dynamics that we fail to account for, this could raise questions as to the validity of our simulation results.

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. While these decisions can alter the study sample, 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. While our design choices have improved the reliability of our estimation results, this is fundamentally true only for the final study cohort. While we believe that the qualitative results likely generalize to the full population, more work is necessary to confirm whether this is indeed the case.

One limitation of our dataset is the lack of patient code status. The estimated effect of ICU transfer on patient outcomes may be overestimated for patients who are not full code as they will not be transferred from the ward/TCU should their condition deteriorate.

Despite the limitations of our study, our results have been invaluable to our partner hospitals. They recently deployed a pilot program where the EDIP2 score is made available to clinicians on an hourly basis at two hospitals. It is currently being used to trigger warnings to a Rapid Response Team (Escobar et al. 2016), but the intent is to have it inform proactive ICU transfers. Our study lends support to this goal. Moreover, the results have been communicated to the remaining 19 hospitals in the hospital system in considering further deployment of the dynamic EDIP2 warning system.

While our findings are specific to the EDIP2, we expect that qualitatively, the benefits of proactive ICU transfer based on the MEWS score (or other scores) would be similar to our findings. Of course, because the EDIP2 is more efficient (Kipnis et al. 2016), the magnitude of the benefits will likely be higher in our study as the EDIP2 is better able to capture the severity of patients who may need ICU care.

The EDIP2 score has high specificity and sensitivity for all 21 hospitals in our study setting, including those with specialized ICUs (Kipnis et al. 2016). As such, we believe that qualitative insights are likely to exist in hospitals with varying ICU resources. Of course, the exact magnitude of the benefits of proactively admitting up to the top 5 EDIP2 groups will vary depending on case mix and size of ICUs.

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 (2016) use predictions of future patient arrivals 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.

Angrist, J. D., G. W. Imbens, D. B. Rubin. 1996. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* **91**(434) 444–455.

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.

Barlow, R., F. Proschan. 1996. *Mathematical Theory of Reliability*. Society for Industrial and Applied Mathematics.

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. G. Nabors, G. Clermont. 2012. Sensitivity analysis of an ICU simulation model. *Proceedings of the 2012 Winter Simulation Conference (WSC)*. 1–12.

Butcher, B. W., E. Vittinghoff, J. Maselli, A. D. Auerbach. 2013. Impact of proactive rounding by a rapid response team on patient outcomes at an academic medical center. *Journal of hospital medicine* **8**(1) 7–12.

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.

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.

Danesh, V., L. Guerrier, E. Jimenez. 2012. 3: Proactive vs. reactive rapid response systems: Decreasing unplanned ICU transfers. *Critical Care Medicine* **40**(12) 1–328.

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.

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. 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) 388–395.

Escobar, G. J., V. Liu, Y. S. Kim, B. J. Turk, A. Ragins, B. A. Dummett, C. L. Adams, E. A. Scruth, P. Kipnis. 2016. Early detection of impending deterioration outside the icu: a difference-in-differences (did) study. Am Thoracic Soc, A7614–A7614.

Evans, R. S., K. G. Kuttler, K. J. Simpson, S. Howe1, P. F. Crossno, K. V. Johnson, M. N. Schreiner, J. F. Lloyd, W. H. Tettelbach, R. K. Keddington, A. Tanner, C. Wilde, T. P. Clemmer. 2015. Automated detection of physiologic deterioration in hospitalized patients. *Journal of the American Medical Informatics Association* **22**(2) 350–360.

Gans, N., H. Shen, Y.-P. Zhou, N. Korolev, A. McCord, H. Ristock. 2015. Parametric forecasting and stochastic programming models for call-center workforce scheduling. *Manufacturing & Service Operations Management* **17**(4) 571–588.

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.

Guirgis, F. W., C. Gerdik, R. L. Wears, D. J. Williams, C. J. Kalynych, J. Sabato, S. A. Godwin. 2013. Proactive rounding by the rapid response team reduces inpatient cardiac arrests. *Resuscitation* **84**(12) 1668–1673.

Imbens, G. W. 2010. Better late than nothing: Some comments on deaton (2009) and heckman and urzua (2009). *Journal of Economic literature* **48**(2) 399–423.

Imbens, G. W. 2015. Matching methods in practice: Three examples. *Journal of Human Resources* **50**(2) 373–419.

Iwashyna, T. J., A. A. Kramer, J. M. Kahn. 2009. Intensive care unit occupancy and patient outcomes. *Critical Care Medicine* **37**(5) 1545–1557.

Jones, D. A., M. A. DeVita, R. Bellomo. 2011. Rapid-response teams. *New England Journal of Medicine* **365**(2) 139–146.

Kc, D. S., C. Terwiesch. 2012. An econometric analysis of patient flows in the cardiac intensive care unit. *Manufacturing & 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.

Kim, S.-H., W. Whitt. 2014. Are call center and hospital arrivals well modeled by nonhomogeneous poisson processes? *Manufacturing & Service Operations Management* **16**(3) 464–480.

Kipnis, P., B. J. Turk, D. A. Wulf, J. C. LaGuardia, V. Liu, M. M. Churpek, S. Romero-Brufau, G. J. Escobar. 2016. Development and validation of an electronic medical record-based alert score for detection of inpatient deterioration outside the ICU. *Journal of Biomedical Informatics* **64** 10 – 19.

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.

Kuntz, L., R. Mennicken, S. Scholtes. 2014. Stress on the ward: Evidence of safety tipping points in hospitals. *Management Science* **61**(4) 754–771.

Lowery, J. C. 1992. Simulation of a hospital's surgical suite and critical care area. *Proceedings of the 24th Conference on Winter Simulation*. WSC '92, ACM, New York, NY, USA, 1071–1078.

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.

McCall, J. J. 1965. Maintenance policies for stochastically failing equipment: A survey. *Management Science* **11**(5) 493–524.

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.

Ormeci, E. L., E. D. Gunes, D. Kunduzcu. 2016. A modeling framework for control of preventive services. *Manufacturing & Service Operations Management* **18**(2) 227–244.

Özekici, S., S. R. Pliska. 1991. Optimal scheduling of inspections: A delayed markov model with false positives and negatives. *Operations Research* **39**(2) 261–273.

Peck, J. S., J. C. Benneyan, D. J. Nightingale, S. A. Gaehde. 2012. Predicting emergency department inpatient admissions to improve same-day patient flow. *Academic Emergency Medicine* **19**(9) 1045–1054.

Pierskalla, W. P., J. A. Voelker. 1976. A survey of maintenance models: The control and surveillance of deteriorating systems. *Naval Research Logistics Quarterly* **23**(3) 353–388.

Rosenbaum, P. R. 2010. *Design of Observational Studies*. Springer-Verlag New York.

Rosenbaum, P. R. 2016. *Observation and Experiment*. Springer.

Rosenbaum, P. R., D. B. Rubin. 1985. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician* **39**(1) 33–38.

Royal College of Physicians. 2012. National early warning score (news): Standardising the assessment of acute-illness severity in the NHS.

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.

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.

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.

Stuart, E. A. 2010. Matching methods for causal inference: A review and a look forward. *Statistical Science* **25**(1) 1–21.

Task Force of the American College of Critical Care Medicine, Society of Critical Care Medicine. 1999. Guidelines for intensive care unit admission, discharge, and triage. *Critical Care Medicine* **27** 633–638.

Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. MIT Press, Cambridge, MA.

Xu, K., C. W. Chan. 2016. Using future information to reduce waiting times in the emergency department via diversion. *Manufacturing & Service Operations Management* **18**(3) 314–331.

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.

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: Supplemental Information on Empirical Analysis

Figure 8 depicts variation in the percentage of ICU transfers by 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. $\geq 90^{\text{th}}$ percentile) and low occupancy ($\leq 50^{\text{th}}$ percentile) is much greater when restricting to the night-time EDIP2 decision epoch versus considering all four. This suggests the instrument is stronger when only considering the night-time decision epoch. We refer to the ICU occupancy for all four EDIP2 time points as the "whole-day instrument" and the ICU occupancy at 10pm as the "night-time instrument". We do not include the 4am-9:59am decision epoch into the night-time instrument, because nearly half of the decision epoch is staffed by day-time physician levels. Finally, we find that the night-time effect is strongest during the first four EDIP2 scores.

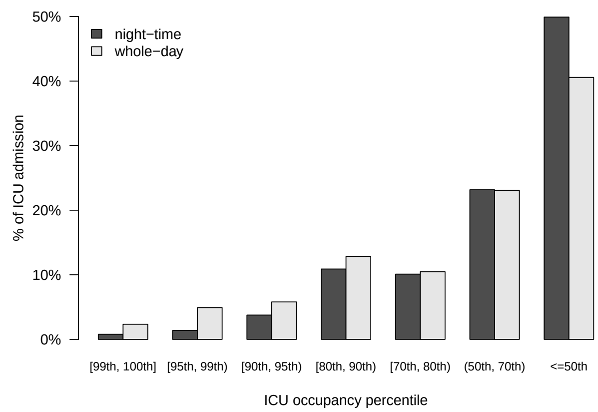


Figure 8 Percentage of ICU transfer by ICU occupancy during night-time and whole-day

Table 12 Control Variables used in Empirical Analysis

Variables	Description
Age	Patient age at time of hospital admission, in years
Gender	Males were coded 0 and females 1
EDIP2	Predicted probability of unplanned transfer from the medical-surgical ward or the TCU to the ICU or death on the ward within the next 12 hours (Escobar et al. 2012); updated every 6 hours at 4am, 10am, 4pm, 10pm, range in [0, 1]; based on vital signs, laboratory test results, COPS2, LAPS2, transpired hospital LOS and care directives;
CHMR	Predicted in-hospital mortality risk, range in [0, 1] (Escobar et al. 2012); based on primary condition-specific models that employed age, gender, admission type, LAPS2 and COPS2;
COPS2	Comorbidity Point Score 2 (Escobar et al. 2013); measures chronic disease burden during the 12 months prior to hospital admission; integer values range in [0, 306];
LAPS2	Laboratory-based Acute Physiology Score 2 (Escobar et al. 2013); measures a patient's acute instability based on lab tests and vital signs 72 hours preceding hospital admission; integer values range in [0, 274];
Diagnosis	Primary diagnosis, grouped into 38 broad disease categories (e.g. pneumonia); categorical variables
Hospital ID	21 hospital IDs; categorical variables
Month/Day	Month/Day of week of hospital admission; categorical variables

A.1. Matching Formulation

Let $\mathcal{T} = \{t_\mathbf{1}, ..., t_\mathbf{T}\}$ be the set of discouraged units, i.e., the subjects that encountered high ICU congestion, and $\mathcal{C} = \{c_\mathbf{1}, ..., c_\mathbf{C}\}$, the set of encouraged units that faced low ICU congestion, with $T \leq C$. Define $\mathcal{P} = \{p_\mathbf{1}, ..., p_\mathbf{P}\}$ as the set of observed covariates. Each discouraged unit $t \in \mathcal{T}$ has a vector of observed covariates $\boldsymbol{x}_{t,\cdot} = \{\mathbf{x}_{t,p_\mathbf{1}}, ..., \mathbf{x}_{t,p_\mathbf{P}}\}$, and each encouraged $c \in \mathcal{C}$ has a similar vector $\boldsymbol{x}_{c,\cdot} = \{\mathbf{x}_{c,p_\mathbf{1}}, ..., \mathbf{x}_{c,p_\mathbf{P}}\}$. Let $0 \leq \delta_{t,c} < \infty$ denote the distance between each pair of discouraged and encouraged units. We solve:

$$\begin{aligned}
\underset{\boldsymbol{a}}{\text{minimize}} \quad & \sum_{t\in\mathcal{T}}\sum_{c\in\mathcal{C}} \delta_{t,c} a_{t,c} - \lambda \sum_{t\in\mathcal{T}}\sum_{c\in\mathcal{C}} a_{t,c} \\
\text{subject to} \quad & \sum_{c\in\mathcal{C}} a_{t,c} \leq 5, \; t \in \mathcal{T} \\
& \sum_{t\in\mathcal{T}} a_{t,c} \leq 1, \; c \in \mathcal{C} \\
& -b_k \sum_{t\in\mathcal{T}}\sum_{c\in\mathcal{C}} a_{t,c} \leq \sum_{t\in\mathcal{T}}\sum_{c\in\mathcal{C}} a_{t,c} v_{k,t,c} \leq b_k \sum_{t\in\mathcal{T}}\sum_{c\in\mathcal{C}} a_{t,c}, k \in K_1 \\
& \sum_{t\in\mathcal{T}}\sum_{c\in\mathcal{C}} a_{t,c} v_{k,t,c} \geq c_k \sum_{t\in\mathcal{T}}\sum_{c\in\mathcal{C}} a_{t,c}, k \in K_2 \\
& a_{t,c} \in \{0,1\}, \; t \in \mathcal{T}, c \in \mathcal{C}
\end{aligned} \tag{4}$$

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 (set to the median of the $\delta_{t,c}$'s) that regulates the trade-off between finding close matches in the covariates and matching as many pairs as possible (see Zubizarreta et al. (2013)).

The first constraint requires 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 encouraged 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 second constraint only allows each encouraged unit to be matched at most once. The third set of constraints are the covariate balance constraints, 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 et al. (2013)). The fourth set of constraints are the imbalance constraints, 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 (4) 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 13–15, 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. In the tables, the standardized difference in means for covariate p is defined as $\frac{\bar{x}_{t,p} - \bar{x}_{c,p}}{\sqrt{(s^2_{t,p} + s^2_{c,p})/2}}$, where $\bar{x}_{t,p}$ and $\bar{x}_{c,p}$ are the sample means for the discouraged and encouraged units after matching, and $s^2_{t,p}$ and $s^2_{c,p}$ are the corresponding sample variances before matching (Rosenbaum and Rubin 1985). Figure 9 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 13 Balance table for risk covariates in means

Covariate	Encouraged	Discouraged	Std diff
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 14 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 15 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

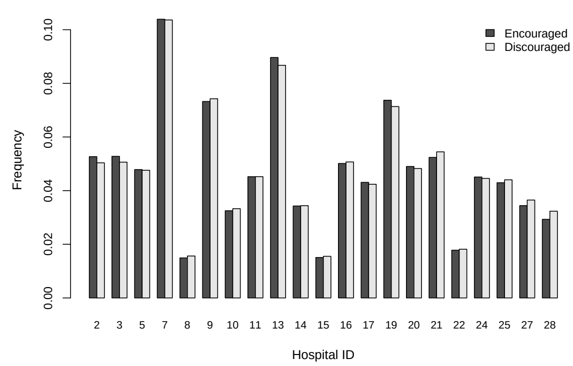


Figure 9 Balance table for hospital ID after match

A.3. Robustness Checks

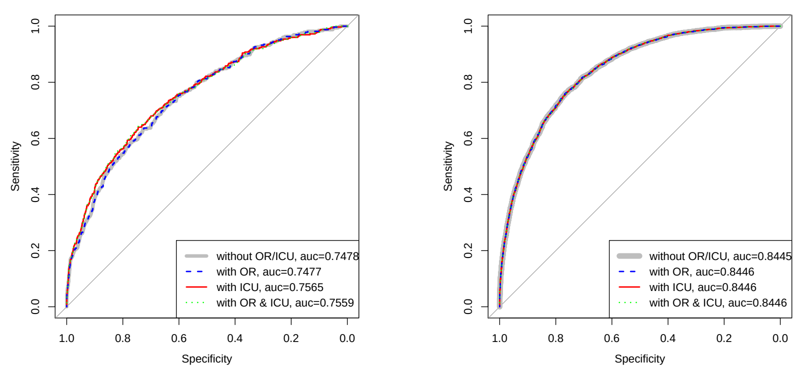
We now consider the robustness of our initial empirical results under alternative specifications.

A.3.1. Alternative IV Definition. In defining the binary instrumental variable from the continuous ICU occupancy levels, we use the 90^{th} percentile and 70^{th} percentile of the ICU occupancy distribution for each hospital as the threshold for "busy" and "not-busy". We also tried different thresholds, including the 65^{th}, 67.5^{th}, 72.5^{th} and 75^{th} percentiles as the "not-busy" threshold, and 92.5^{th} and 87.5^{th} percentiles as the "busy" threshold. The estimation results are similar with only slight changes in the coefficient estimates.

A.3.2. Additional Covariates. In our econometric models, we have included both patient severity factors and seasonality controls. We also considered including indicators of whether a patient had been admitted to the ICU or OR before being admitted to an inpatient unit. We fit a logistic regression of the ICU transfer decisions on all patient severity risk factors and seasonality controls, including the two additional indicators and constructed 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). 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 10 shows the ROC curves for the ICU transfer model and mortality model with and without the two additional risk factors. 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 or mortality. To avoid over-fitting, we opted not to include the two additional covariates as controls.

Figure 10 ROC Curves



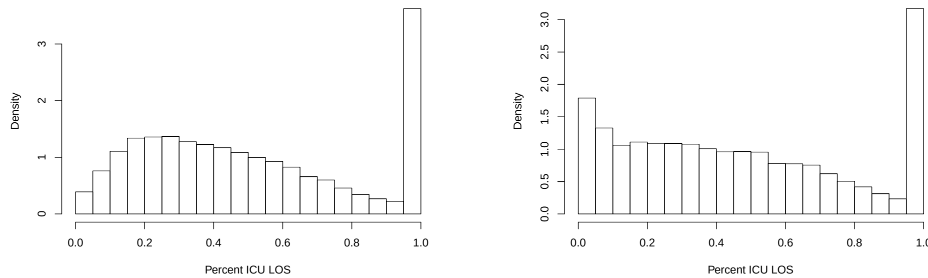
(a) Night-time ICU transfers (b) Mortality of night-time patients

Appendix B: Supplemental information for Simulation

B.1. Transition Matrix for Ward Patients

Patients in the ward are modeled by a discrete time Markov Chain with the transition probability matrix $\mathbf{T}$. There are 10 transient states: $i \in \{1, 2, \ldots 10\}$, where state i denotes the patient is currently in EDIP2 group

Figure 11 Empirical probability mass function for proportion of hospital LOS spent in the ICU.



(a) Direct Admits $f_{p_E}(p)$

(b) Ward patients who are admitted to ICU $f_{p_W}(p)$

i. There are 3 absorbing states: $i = 11$ corresponds to a patient crashing; $i = 12$ corresponds to a patients being discharged alive; and $i = 13$ corresponds to a patient dying in the ward. $\mathbf{T}_{i,j}, i = 1, 2, \ldots, 10, j = 1, 2, \ldots, 13$ represents the probability of a patient transitioning from EDIP2 group i to state j within each period. We calibrate $\mathbf{T}_{i,j}$ from our data using the proportion of transitions to each state:

$$\mathbf{T}_{i,j} = \begin{cases} \frac{\sum_k \sum_t 1\{EDIP2_k(t)=i\} \times 1\{EDIP2_k(t+1)=j\}}{\sum_k \sum_t 1\{EDIP2_k(t)=i\}}, & i, j = 1, 2, \ldots 10; \\ \frac{\sum_k \sum_t 1\{EDIP2_k(t)=i \times 1\{crash_k(t+1)\}\}}{\sum_k \sum_t 1\{EDIP2_k(t)=i\}}, & i = 1, 2, \ldots 10, j = 11; \\ \frac{\sum_k \sum_t 1\{EDIP2_k(t)=i \times 1\{discharge_k(t+1)\}\}}{\sum_k \sum_t 1\{EDIP2_k(t)=i\}}, & i = 1, 2, \ldots 10, j = 12; \\ \frac{\sum_k \sum_t 1\{EDIP2_k(t)=i \times 1\{death_k(t+1)\}\}}{\sum_k \sum_t 1\{EDIP2_k(t)=i\}}, & i = 1, 2, \ldots 10, j = 13. \end{cases}$$

where $1\{x\}$ is an indictor variable equal to 1 if x is true; $EDIP2_k(t)$ is the EDIP2 group for patient k during epoch t; $crash_k(t)$ denotes whether patient k crashed during EDIP2 epoch t; $discharge_k(t)$ denotes whether patient k is discharged from the ward alive during EDIP2 epoch t; and, $death_k(t)$ denotes whether patient k died in the ward during EDIP2 epoch t. We sum over all patients, k, and all EDIP2 epochs, t. The estimated transition matrix is:

$$\mathbf{T} = \begin{bmatrix} 0.8134 & 0.0728 & 0.0108 & 0.0020 & 0.0009 & 0.0001 & 0.0001 & 0.0001 & 0.0000 & 0.0001 & 0.0012 & 0.0982 & 0.0003 \\ 0.3216 & 0.4445 & 0.1223 & 0.0226 & 0.0072 & 0.0004 & 0.0003 & 0.0002 & 0.0003 & 0.0003 & 0.0021 & 0.0774 & 0.0008 \\ 0.0742 & 0.3491 & 0.3638 & 0.1075 & 0.0345 & 0.0017 & 0.0011 & 0.0009 & 0.0007 & 0.0008 & 0.0031 & 0.0609 & 0.0015 \\ 0.0229 & 0.1351 & 0.3608 & 0.2844 & 0.1259 & 0.0060 & 0.0040 & 0.0025 & 0.0024 & 0.0021 & 0.0048 & 0.0468 & 0.0023 \\ 0.0058 & 0.0488 & 0.1682 & 0.2893 & 0.3608 & 0.0287 & 0.0194 & 0.0146 & 0.0105 & 0.0079 & 0.0086 & 0.0330 & 0.0045 \\ 0.0019 & 0.0147 & 0.0695 & 0.1604 & 0.4567 & 0.0838 & 0.0599 & 0.0475 & 0.0366 & 0.0223 & 0.0140 & 0.0246 & 0.0082 \\ 0.0013 & 0.0105 & 0.0483 & 0.1249 & 0.4235 & 0.1020 & 0.0829 & 0.0670 & 0.0521 & 0.0364 & 0.0190 & 0.0228 & 0.0090 \\ 0.0010 & 0.0066 & 0.0320 & 0.0931 & 0.3625 & 0.1068 & 0.1007 & 0.1038 & 0.0817 & 0.0555 & 0.0233 & 0.0214 & 0.0117 \\ 0.0008 & 0.0043 & 0.0185 & 0.0577 & 0.2678 & 0.0917 & 0.1031 & 0.1292 & 0.1490 & 0.1116 & 0.0320 & 0.0171 & 0.0171 \\ 0.0007 & 0.0015 & 0.0074 & 0.0212 & 0.1119 & 0.0444 & 0.0611 & 0.0872 & 0.1557 & 0.3749 & 0.0616 & 0.0143 & 0.0581 \end{bmatrix}$$

B.2. Optimization Problem to Calibrate Crashed Parameters

We use our empirical results in Section 4 to determine the predicted mortality rate and LOS for patients in each of the 10 EDIP2 groups based on whether they are admitted at that EDIP2 severity level (before crashing) versus not. The average predicted values are summarized in Table 16. To emphasize the translation of our empirical findings to the simulation model where proactive ICU admissions are possible, we label the predictive values when an action is taken (i.e. ICU admission within the 6 hour EDIP2 decision epoch) at a specific EDIP2 severity score as *Proactive*. In contrast, we label no action within the epoch as *Reactive*.

Patients not proactively transferred to the ICU stay in the ward until they crash or are discharged (alive or dead) from the ward. Thus, our Markov Chain model, with 6 hour time slots, gives for patients in EDIP2 group i a probability of death, MD_i, and an expected LOS, $MLOS_i$, when not proactively transferred as:

$$MD_i \triangleq P_i(\text{death}|\text{not proactively transferred}) = P_i(\text{death in ward}) + P_i(\text{crash}) \times d_C$$
$$MLOS_i \triangleq \mathbb{E}[LOS_i|\text{not proactively transferred}] = 6 \cdot \mathbb{E}[\text{\# of periods in ward}|\text{patient group } i] + P_i(\text{crash}) \times LOS_C$$

Our objective is to determine d_C in order to minimize the sum of squared *percentage* errors between the reactive predicted probability of death summarized in Table 16, which we denote by PD_i^R, and MD_i. As our empirical results suggest patients proactively transferred to the ICU have lower mortality risk than if they crash, we add a constraint that $d_C \geq PD_i^A, \forall i$, where PD_i^A is the predicted probability of death with proactive transfer summarized in Table 16. The optimization problem is formulated as:

$$\min_{d_C} \sum_{i=1}^{10} \left(\frac{PD_i^R}{MD_i} - 1 \right)^2$$
$$\text{s.t. } d_C \geq PD_i^A, \forall i$$

We formulate and solve a similar optimization problem for LOS_C:

$$\min_{LOS_C} \sum_{i=1}^{10} \left(\frac{PLOS_i^R}{MLOS_i} - 1 \right)^2$$
$$\text{s.t. } LOS_C \geq PLOS_i^A, \forall i$$

We chose to minimize the sum of squared percentage errors due to the large variation in the magnitude of mortality across the 10 EDIP2 groups; e.g., PD_{10}^R is 25 times that of PD_1^R. Thus, an optimization problem whose objective is to minimize the sum of squared errors would result in a d_C which are dominated by the top EDIP2 group at the cost of not fitting the lower EDIP2 groups well. This is less of an issue for the LOS optimization model and we find the results for LOS are similar under both optimization objective functions.

Solving the optimization problems result in $d_C = 57.3\%$ and $LOS_C = 15.1$ days. Table 17 summarizes the mortality rates and mean LOS for each EDIP2 group based on our Markov Chain model using these crashed parameters. Recall that we use the predicted probability of death and predicted LOS under ICU transfers at the given EDIP2 scores for the proactive parameters in our model.

Table 16 Summary of mean predicted mortality risk and LOS for 10 EDIP2 groups when admitted to the ICU (Proactive) or not admitted (Reactive) in a given EDIP2 decision epoch

	Mortality (%)		LOS (day)	
EDIP2 Group	Proactive PD_i^A	Reactive PD_i^R	Proactive $PLOS_i^A$	Reactive $PLOS_i^R$
1	0.01	1.26	0.85	1.96
2	0.02	1.83	0.91	2.12
3	0.04	2.42	0.97	2.25
4	0.05	3.20	1.04	2.42
5	0.11	4.84	1.17	2.72
6	0.18	6.90	1.36	3.15
7	0.28	8.49	1.45	3.36
8	0.39	10.63	1.57	3.64
9	0.70	15.46	1.85	4.29
10	6.84	33.19	3.77	7.47

Table 17 Markov Chain model: Expected mortality and LOS under proactive and reactive ICU transfers for 10 EDIP2 groups

	Mortality (%)		LOS (day)	
EDIP2 Group	Proactive PD_i^A	Reactive MD_i^R	Proactive $PLOS_i^A$	Reactive $MLOS_i^R$
1	0.01	1.73	0.85	2.54
2	0.02	2.30	0.91	2.74
3	0.04	3.04	0.97	2.94
4	0.05	3.96	1.04	3.13
5	0.11	5.56	1.17	3.34
6	0.18	7.62	1.36	3.51
7	0.28	8.63	1.45	3.58
8	0.39	9.97	1.57	3.64
9	0.70	12.53	1.85	3.73
10	6.84	22.02	3.77	3.79

Table 18 Percentage differences between the best 3- and 4-threshold state-dependent policy static policy

3-threshold						
# of groups	Mortality	LOS	$DDD_{crashed}$	$DDD_{directadmit}$	$r_{crashed}$	$r_{directadmit}$
Top 2	0.19	-0.04	-1.94*	-1.02	0.18	-0.50*
Top 3	1.01*	0.27*	-2.39*	-2.21*	-0.50	-0.13
Top 4	0.74*	0.23*	-2.01*	-1.39*	-0.55	-0.17
Top 5	1.09*	0.26*	-3.25*	-3.17*	-0.25	-0.32
Top 6	2.83*	1.73*	-7.25*	-7.10*	-0.80	-0.76*

4-threshold						
# of groups	Mortality	LOS	$DDD_{crashed}$	$DDD_{directadmit}$	$r_{crashed}$	$r_{directadmit}$
Top 3	2.04*	0.44*	-2.90*	-3.04*	-0.54	-0.31
Top 4	1.32*	0.31*	-2.45*	-2.58*	-0.54	-0.22
Top 5	1.31*	0.37*	-4.23*	-3.93*	-0.77	-0.33
Top 6	4.46*	2.52*	-8.69*	-8.14*	-1.35*	-0.42

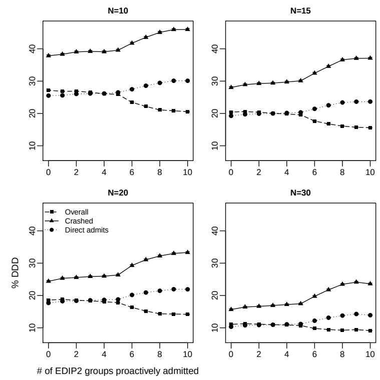
*: $p < 0.05$ difference in means based on t-tests

B.3. Simulation Robustness Checks

ICU size: We consider 4 different ICU sizes $\mathbf{N} = 10, 15, 20, 30$ operated at approximately 70%, 80% and 90% average ICU occupancy under reactive transfer. The trends for in-hospital mortality rates and LOS are highly similar across the 4 ICU sizes. We find that capacity pooling results in higher demand-driven discharge and readmission rates for small ICUs with the same ICU occupancy level (e.g. Figure 12). Despite the slight changes in the magnitude of the effect of proactive admissions in ICUs of different sizes, we see the qualitative insights (e.g. proactively admitting up to 5 EDIP2 groups can be beneficial) are robust.

Parameter calibration: We also vary the calibration of some of our model primitives. Specifically, we vary $\beta = [0.1, 0.2, \ldots, 0.9]$, which impacts the ICU readmission rates for proactive transfers, as well as the mortality and readmission rates for external arrivals (d_E and r_E) and the readmission rate for crashed patients (r_C) over the 95% confidence intervals for these parameters. Similar to our results for different ICU

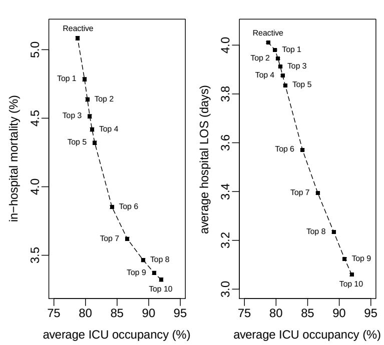
Figure 12 **Demand-driven discharge under 4 ICU sizes at daily arrival rates $\Lambda = 9.7, 14.2, 18.7, 27.8$, which correspond to approximately 80% ICU congestion for each of the ICUs.**



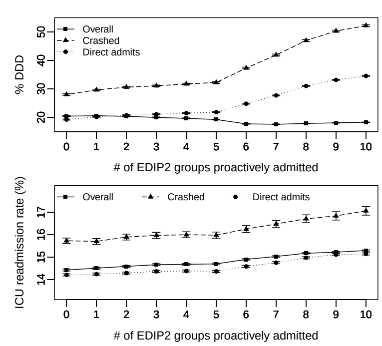
sizes, we find that qualitative insights are robust to these variations in parameter calibration. In fact, we find that the differences in most outcomes (LOS, mortality rates, demand-driven discharge) are on average 1.2% and no more than 3.2%. Because β directly impacts the readmission rates for proactive transfers, varying β by an order of magnitude (from 1 to 0.1) can have a substantial impact on overall readmission rates. Specifically, across all of the various parameter combinations, we find that the mean relative change in ICU readmission is 5.3% with a maximum of 39.8%, which occurs when $\beta = 0.1$.

Proactive transfer during the whole day: We next consider the case where proactive transfers can occur during any EDIP2 decision epoch (instead of just the night-time one). Here we assume that our empirical estimates can be generalized to the whole day. These results are summarized in Figure 13. While the main insights of this scenarios are consistent with our initial findings which restrict to night-time proactive transfers, we find that with more frequent proactive ICU transfer decisions, the effects on outcomes are more drastic because proactive ICU transfers are done more aggressively.

Figure 13 **Proactive transfers can occur throughout the day.** $\mathrm{N} = 15$ **ICU beds.** $\Lambda = 14.2$ **patients/day.**



(a) Mortality and LOS versus ICU occupancy.



(b) Demand-driven discharge and readmission rates