Managing Hospital Care: Data-driven decisions and comparisons

Wenqi Hu

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

under the Executive Committee

of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2018

ABSTRACT

Managing Hospital Care: Data-driven decisions and comparisons

Wenqi Hu

This dissertation focuses on utilizing data-driven approaches to objectively measure variation in the quality of care across different hospitals, understand how physicians make dynamic admission and routing decisions for patients, and propose potential changes in practice to improve the quality of care and patient flow management. This analysis was performed in the context of Intensive Care Units (ICUs) and the Emergency Department (ED).

In the first part, we assess variation in the overall quality of care provided by both urban and rural hospitals under the same integrated healthcare delivery system when augmenting administrative data with detailed patient severity scores from the electronic medical records (EMRs). Using a new template matching methodology for more objective comparison, we found that the use of granular EMR data significantly reduces the variation across hospitals in common patient severity-of-illness levels. Further, we found that hospital rankings on 30-day mortality and estimates of length-of-stay (LOS) are statistically different from rankings based on administrative data.

In the second part, we study ICU admission decision-making dynamically throughout a patient's stay in the general ward/the Transitional Care Unit (TCU). We first used an instrumental variable approach and modern multivariate matching methods to rigorously estimate the potential benefits and costs of transferring patients to the ICU based on a real-time risk score for deterioration. We then used the quantified impact to calibrate a comprehensive simulation model to evaluate system performances under various new ICU transfer policies. We show that proactively transferring the most severe patients to the ICU could reduce mortality rates and LOS without increasing ICU congestion and causing other adverse effects.

In the third part, we focus on understanding how physicians make ICU admission decisions for patients in the ED. We first used two sets of reduced-form regressions to understand 1) what and how patient risk factors and system controls impact the admission decision from the ED; and 2) what are the potential benefits of admitting patients from the ED to the ICU. We then proposed a dynamic discrete choice structural model to estimate to what extent physicians account for the intertemporal externalities when deciding to admit a specific patient to the ICU, to the ward or let him/her wait in the ED. Note that the structural model estimation is still an ongoing process and more investigation is required to fine tune the details. Therefore, we will not discuss the structural model estimation results in this chapter, but only present the modeling framework and key estimation strategy.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my deepest gratitude to my advisors, Professors Carri W. Chan and José R. Zubizarreta, for their dedicated support, guidance and connections throughout my studies. It has been great honor to be able to work with and learn from them. Besides their input on the academic work, their encouragement and sharing on personal life and career perspectives have been priceless and highly appreciated.

I thank Professor Fanyin Zheng for guiding me through the work in Chapter 4 of this dissertation; this chapter would not have existed without her generous coaching and patience. I also thank my medical collaborator, Dr. Gabriel J. Escobar at Kaiser Permanente Northern California, for his professional knowledge and guidance throughout my research work.

Professor Song-Hee Kim has always been my academic sister and mentor throughout the studies. I am very grateful to her for all the encouragement and advice both professionally and personally.

I thank Professor Linda Green for kindly serving as my committee member and for her valuable feedback on my proposal and dissertation. I also thank Professor Costis Maglaras for his kind sharing and support on career issues.

Fellow students in the Decision, Risk, and Operations Division, the IEOR department, and my friends have been an indispensable part of the past five years. I could not have gone thus far without their continuous support, encouragement and accompany during the hard time. Special thanks to Anran, Lijian and Zhe for brainstorming on research work and sharing my happiness and anxiety, and to Letian for the tremendous technical support on the work in Chapter 4 of this dissertation.

Lastly, I would like to thank my family for their unconditional understanding and support, which allows me to focus on my studies without worrying about any family duty.

Chapter 1

Introduction

Among all the high-income countries in the world, the U.S. has the highest healthcare spending. In 2015, U.S. healthcare expenditures increased by 5.8% to $3.2 trillion in total or $9,990 per person, accounting for 17.8% of the nation's Gross Domestic Product. Hospital care—the largest contributor to total U.S. healthcare expenditures—experienced a faster growth of 5.6% to $1 trillion in 2015 compared with 2014. With the tremendous and growing amount of spending on healthcare, especially hospital care, it has become increasingly important to objectively measure hospital performances, identify areas for quality improvement, and propose potential new policies. This dissertation focuses on utilizing data-driven approaches to (1) objectively measure variation in overall quality of care across different hospitals and (2) understand and potentially improve dynamic admission decision-making in the Intensive Care Unit (ICU) and the Emergency Department (ED).

Chapter 2 of this dissertation focuses on assessing variation in overall quality of care across different hospitals when incorporating detailed patient severity scores from the electronic medical records (EMRs) into the comparison. This work is published on Medical Care (Hu *et al.*, 2018b). Chapters 3 and 4 of this dissertation focus on understanding the ICU admission decisions in the general ward/Transitional Care Unit (TCU) or the ED for patients with varying disease conditions and severity levels. The work in Chapter 3 is published on MSOM (Hu *et al.*, 2018a) and Chapter 4 is an ongoing work.

1.1. Quality of Care Comparisons

This stream of research focuses on finding an objective way to compare hospital performance when hospitals may vary in their location and patient mix.

It has long been recognized that widespread variation in medical practices undermines the consistency and quality of healthcare. Over the past 40 years, healthcare providers, researchers, and policy makers have increasingly described and investigated the substantial variation in medical practices and the healthcare quality across industrialized countries (Australian Commission on Safety and Quality in Health Care, 2015; Eskander *et al.*, 2015; Institute of Medicine, 2001; McGlynn *et al.*, 2003; Department of Health, 2015). To better measure the large variation in quality of care delivered by U.S. hospitals, the Centers for Medicare and Medicaid Services (CMS) launched a Hospital Inpatient Quality Reporting program in 2004 (Jha *et al.*, 2005; Lindenauer *et al.*, 2007). More recently, Medicare payment rates to more than 3,000 hospitals have been adjusted based on their performance across some of the metrics reported to CMS, notably, readmission rates. For example, in 2016, a weight of 40% was assigned to quality of care outcome measures, such as 30-day mortality rates.

Significant effort has been put into comparing hospital quality within and across hospitals, or even across countries, using mostly claims data. Note that claims data typically only contains selected patient demographics and treatment procedures that are necessary for reimbursement, and therefore is less detailed than the EMRs. Concerns have been raised as to the credibility of comparing hospital performances based on claims data given the limited information about patients available in these datasets. In Chapter 2, we exploited the richness of our EMR data to include patients' composite severity of illness status at hospital admission in addition to their demographics. We leveraged a new template matching methodology to better balance patient demographics and severity status, so as to objectively audit hospital performances based on a common patient mix that is representative of the population seen at all hospitals, rather than on each hospital's individual patient mix. We show that using only patient demographics and failing to control for a good measure of patients' severity of illness status could unfairly penalize hospitals that see more severe

patients.

1.2. ICU Admission Decisions

This stream of research aims at using rigorous econometric models and modern matching methodologies to understand physician decision-making dynamically throughout a patient's hospitalization and the impact on patient outcomes. With a better understanding of the current practice, we further suggest potential policy changes to help improve hospital performances. This analysis was performed in the context of the ICU and the ED.

ICUs are specialized inpatient units that provide intensive care and continuous monitoring to the most critically ill patients. ICUs are expensive units to operate, where the average daily cost is 1.73–2.55 times that of an average ward due to intensive staffing, invasive therapies, and expensive equipment (Milbrandt *et al.*, 2008). In addition, ICUs often run near full capacity (Green, 2002). Morever, ICU admissions have increased by 48.8% from 2002 through 2009 (Mullins *et al.*, 2013), and the usage of ICUs will likely continue to rise with the population aging (Milbrandt *et al.*, 2008). The high cost of ICU care and rising need means developing a better understanding of the ICU admission decision is of increasing importance.

However, there is a lack of established standards on which patients should be admitted to the ICU. One pathway to the ICU is from the general medical-surgical ward or the TCU, where patients experience unexpected rapid deterioration in their physiological conditions while staying in the ward and would require ICU transfer immediately. Such unplanned transfers of patients are typically associated with higher mortality and longer length-of-stay (LOS) than expected. Chapter 3 studies the potential benefits and costs of proactively transferring patients to the ICU based on a real-time risk score for deterioration. Another pathway to the ICU is from the ED, and Chapter 4 studies the decision to admit a patient from the ED to the ICU, the ward, or to continue waiting in the ED for a bed to become available in the future. Both pathways involve dynamic decision-making during a patient's stay rather than at a single point-in-time.

Chapter 3 is devoted to understanding the potential benefits and costs of using a real-time

physiologic risk score to trigger ICU transfers from the general ward/TCU before patients actually experience rapid deterioration in their physiological conditions and have to be transferred to the ICU immediately. In studying complex ICU admission and discharge decisions from observational data, researchers are typically concerned about missing unobservable severity factors from the data that could bias inferences, even with our rich set of data. To address this endogeneity bias, we utilize an instrumental variable (IV) approach by controlling for the ICU congestion level. Although traditional empirical healthcare studies using an IV would simply rely on the strength of the IV as given by the data, previous work has shown that if the instrument is not strongly correlated with the endogenous variable, the IV estimates can be biased and the confidence intervals may be misleading. Moreover, when using regression-based methods, if there is limited overlap in the covariate distributions across levels of the IV, it would rely heavily on the selection of a correct functional form to extrapolate the difference in the covariate distributions. To strengthen the IV and reduce model dependence, we restricted the analysis to the night-time period, when the IV has a stronger effect on ICU admission decisions, and use recent advancements in multivariate matching to balance the covariate distributions and separate the IV. Based on our rigorous estimates and the real hospital data, we calibrated a comprehensive simulation model to evaluate system performances under various new ICU admission policies. We found that proactively admitting the most severe patients could reduce mortality rates and the LOS without increasing ICU congestion and other adverse effects. Our results have helped in launching a pilot study in two hospitals since 2015 to alert the Rapid Response Teams (RRTs) for deterioration in patient conditions that potentially would need ICU care in the next 12 hours.

Chapter 4 aims to understand how physicians make ICU admission decisions for patients in the ED. We first used two sets of reduced-form regressions to understand 1) what and how patient risk factors and system controls impact the admission decision from the ED; and 2) what are the potential benefits of admitting patients from the ED to the ICU. We found that the sicker the patient, the less congested the ICU, the fewer severe patients in the ICU, or the more recent ICU discharges, all contribute to increasing the likelihood of ICU admission from the ED. We showed that being admitted to the ICU from the ED could significantly reduce patients' in-hospital mortality,

LOS, and readmission rates within one month. We then proposed a structural model to estimate to what extent physicians account for the intertemporal externalities when deciding to admit a specific patient to the ICU, to the ward or let him/her wait in the ED. Since admitting a patient can prevent future and potentially more severe patients from being admitted, it is conceivable that physicians would need to consider both the condition of the patient in question as well as the congestion levels and the workload in the ICU and non-ICU units. It is not clear whether or how physicians have been internalizing this trade-off in their decision-making. Thus, we used a dynamic discrete choice structural model to estimate the cost parameters physicians place on their decision choices and the intertemporal discount factor. Note that the structural model estimation is still an ongoing process and more investigation is required to fine tune the details. Therefore, we will not discuss the structural model estimation results in this chapter, but only present the modeling framework and key estimation strategy. For future work, it would be helpful to conduct counterfactual analyses to investigate whether patient flow from the ED to inpatient units can be improved by adjusting the way physicians make decisions.

Part I

Hospital Quality of Care Comparisons

Chapter 2

Incorporating Longitudinal Comorbidity and Acute Physiology Data in Template Matching for Assessing Hospital Quality: an Exploratory Study in an Integrated Health Care Delivery System

2.1. Introduction

Over the past 40 years, healthcare providers, researchers and policy makers have described and investigated the substantial variation in medical practice and health care quality across industrialized countries (Australian Commission on Safety and Quality in Health Care, 2015; Eskander *et al.*, 2015; Institute of Medicine, 2001; McGlynn *et al.*, 2003; Department of Health, 2015). To better measure the large variations in quality of care delivered by U.S. hospitals, the Centers for Medicare and Medicaid Services (CMS) launched a Hospital Inpatient Quality Reporting program in 2004 (Jha *et al.*, 2005; Lindenauer *et al.*, 2007). More recently, Medicare payment rates to hospitals have

been adjusted based on their performance across some of the metrics reported to CMS, notably, readmission rates.

Comparing hospitals objectively requires finding a good way to standardize patient populations in different centers. This boils down to determining which patient characteristics to consider and which statistical approach to use for risk-adjusting the patient population under comparison. At the time of hospital admission, patients will likely vary extensively across age, sex, socioeconomic status, disease codes, comorbid conditions, and physiological derangement. Moreover, they may undergo medical or surgical treatments of varying resource intensity. Because patients are non-randomly self-referred to different hospitals, the case-mix can vary substantially, and it has long been acknowledged that making comparisons based on crude mortality rates may penalize hospitals that have high mortality rates because they treat the most severely ill patients (Fleiss *et al.*, 2013; Park *et al.*, 1990; Moscucci *et al.*, 2005). It is, therefore, important to standardize the patient population to avoid potential confounding variables when conducting hospital quality comparisons. Unfortunately, it is challenging to do standardization accurately (lezzoni, 1997).

In order to standardize the patient population, policy makers must consider whether to utilize indirect or direct standardization methods in adjusting for patient populations across hospitals. The most common statistical approach used in indirect standardization is to employ regression-based risk adjustment modeling (Shahian and Normand, 2008). Since 2008, CMS has reported risk-adjusted 30-day mortality rates across all hospitals for Medicare fee-for-service patients with congestive heart failure (CHF), myocardial infarction and pneumonia. Indirect standardization compares a hospital's performance for its own specific case-mix (observed outcome) with the performance of a hypothetical average hospital that treats similar patients as the case-mix of the specific hospital (expected outcome), the ratio between the two outcomes typically abbreviated as the O/E ratio. There are several limitations regarding the method. First, indirect standardization allows the patient mix under comparison to vary across hospitals, and relies on model adjustment to make the quality outcomes comparable. If there is limited overlap between case-mix distributions across hospitals, risk adjustment models would heavily rely on specific parametric assumptions to extrapolate the counterfactual outcome from the reference hospitals, and such risk adjustment may

not be adequate (Shahian and Normand, 2008; Rosenbaum, 2010; Jones *et al.*, 2014). Second, risk-adjusted mortality has been shown to be an inconsistent measure of hospital quality when patient risk factors are correlated with hospital specific characteristics (Jones *et al.*, 2014; Glance *et al.*, 2006a). Third, risk-adjusted outcomes can be sensitive to the amount of patient characteristics adjusted for, i.e., estimated outcome measures based on different severity measures are of varying accuracy and can disagree with each other substantially and often, especially when using medical records versus administrative data (lezzoni, 1997; Glance *et al.*, 2006a; Iezzoni *et al.*, 1996). Systems may also employ different reference populations and lack stability over time, etc. (Scott *et al.*, 2011). Finally, confidence intervals for the O/E ratios may be underestimated when the risk-adjustment model used for performance comparison uses data from the hospitals being compared (Faris *et al.*, 2003; Goldstein and Spiegelhalter, 1996).

Recently, Silber *et al.* (2014a) proposed a new approach to assess hospital quality by direct standardization called template matching. Template matching uses optimal multivariate matching to find comparable matched samples of patients across hospitals. With template matching, we first identify a template sample that is representative of a patient target population of interest. We then use modern multivariate matching techniques (Zubizarreta, 2012; Zubizarreta *et al.*, 2016) to find matched samples of patients across hospitals that are highly comparable to the template sample. Unlike propensity score matching methods, the matching techniques in Zubizarreta (2012) and Zubizarreta *et al.* (2016) directly and flexibly balance covariates. In this manner, template matching ensures that the hospital comparisons are fair in terms of observed patient characteristics. In contrast to model-based approaches to risk adjustment and hospital comparisons, template matching confines hospital comparisons to comparable data and avoids the perils of model misspecification and extrapolation when there is limited overlap in covariate distributions across hospitals. In the future, template matching may become more common due to its flexibility in selecting the target patient population for hospital comparisons. It can also enhance transparency in understanding who is included or not in the comparisons. In this report, we describe how we build on the seminal work of Silber *et al.* (2014a) by incorporating additional data from comprehensive electronic medical records (EMRs).

In addition to the method of standardization, policy makers must also consider the tradeoff between using purely administrative data versus incorporating data available from EMRs. Administrative data are relatively inexpensive to collect and generally widely available. However, the amount of detailed patient characteristics included in such datasets can be quite limited. In contrast, data from modern comprehensive EMRs tend to have more detailed information regarding patient characteristics, although many centers cannot access such data. Traditional hospital comparison methods are typically based on administrative data, such as patient demographics, diagnoses and procedures that are derived from hospital reimbursement claims (Jha *et al.*, 2005; Lindenauer *et al.*, 2010; Krumholz *et al.*, 2006; Fonarow *et al.*, 2011)—databases that were not designed for the purposes of quality of care assessment. Comorbidities and chronic diseases are likely to be underreported, which may result in inaccurate quality assessment (Green and Wintfeld, 1993; Quan *et al.*, 2002). The lack of more detailed data, including present-on-admission (POA, present at the time the order for inpatient admission occurs) comorbidities and information on acute physiology, limits the ability to fully adjust for patient acute physiological conditions (Jollis *et al.*, 1993; Pine *et al.*, 1997, 2007; Glance *et al.*, 2006b). Both issues raise questions about the accuracy and objectiveness of the comparison based solely on administrative data.

Increasing availability and lower cost of EMRs has generated growing interest in augmenting administrative records with present-on-admission (POA) comorbidity conditions, secondary diagnoses and physiological data to better quantify patient severity (Pine *et al.*, 2007; Berner *et al.*, 2005; Fry *et al.*, 2007). Escobar *et al.* (2013) used logistic regression and found that enhancing traditional risk-adjustment models with physiological data enhanced the ability to predict hospital mortality but found that incorporating longitudinally captured comorbidity data did not. In this report, we examine how hospital rankings change when further adjusting for patients' comorbidity history and/or acute physiology compared to using only administrative data. However, we take a different methodological approach—template matching—to compare hospital performance. Specifically, we describe how hospital rankings may change after incorporating longitudinal comorbidities and acute physiology.

To achieve this goal, we defined a representative reference population. We then augmented

administrative data with information on longitudinal comorbidity and acute physiology and selected patients with similar characteristics to those of the reference population. After this, we ranked hospitals' 30-day mortality rates and obtained Hodges-Lehmann point estimates of length-of-stay (LOS) (Hodges and Lehmann, 2012). We concluded by comparing rankings obtained with varying degrees of information.

2.2. Methods

2.2.1. Study Setting

This project was approved by the KPNC Institutional Review Board, which has jurisdiction over all the hospitals included in this report.

Our setting consisted of 18 of the 21 hospitals in Kaiser Permanente Northern California (KPNC), an integrated health care delivery system. We focused on patients admitted for 4 specific diagnosis groups: (1) sepsis and pneumonia, (2) congestive heart failure (CHF), (3) hip fracture and (4) cancer[1]. We chose these groups because they were among the largest disease categories in our entire study population, which ensured enough hospitalizations from each hospital for performance comparison. We dropped three of the smallest hospitals because they had insufficient numbers of patients, as we aimed to ensure that all included had at least 50 patients in each of the 4 disease categories.

2.2.2. Patient Characteristics

For each hospitalization, we had patient level admission data which included the patient's age, sex, admitting hospital, admitting type (medical or surgical), admission venue (emergency department [ED] or not), admission diagnosis, longitudinal comorbidity burden and acute physiology. The co-

[1]The broad cancer category is a class of low severity cancers grouped by KPNC based on biologic plausibility (i.e., relative similarity from a disease standpoint) and on the observed mortality rate for modeling purposes (Escobar *et al.*, 2008). Such grouping of patients' primary conditions has been used in developing risk adjustment and predictive models in Escobar *et al.* (2008, 2013). After all, there is no easy way to come up with patient groupings, and KPNC has to make some pragmatic choices to strike a balance between similarity among patient conditions and number of observations.

morbidity burden was captured using the continuous COmorbidity Point Score 2 (COPS2), which measures a patient's chronic disease burden during the 12 months prior to hospital admission with integer values ranging in [0, 306]. Acute physiology was captured using the continuous Laboratory Acute Physiology Score 2 (LAPS2), which measures a patient's acute instability based on laboratory tests and vital signs 72 hours preceding hospital admission with integer values ranging in [0, 274] (see Escobar *et al.* (2013) and its appendix for more information on these two scores). The availability of these scores permitted us to analyze their effects separately. Note the COPS2 and LAPS2 scores were developed to predict in-hospital and 30-day mortality rates. In Escobar *et al.* (2013), they showed that the c-statistic for predicting in-hospital mortality increased from 0.798 to 0.883 when including COPS2 and LAPS2, compared with using only administrative data. A similar improvement was also shown for predicting 30-day mortality. Therefore, both scores capture additional information on patient severity at the time of hospital admission, and should be adjusted for when objectively comparing hospital quality.

2.2.3. Defining a Representative Patient Sample

We used the template matching methodology developed by Silber *et al.* (2014a). Figure 2.1 illustrates the template matching process we employed.

We first defined a common reference population (the template). As recommended by Silber *et al.* (2014a), we wanted the template to be representative of the 4 diagnosis groups typically treated at all 18 hospitals. Our definition of "representative" for each diagnosis group meant that, with respect to the pooled population from all 18 hospitals, the following were similar: (1) the proportions of the 4 diagnosis groups; (2) the proportions of ED/non-ED and medical/surgical patients; and (3) mean patient COPS2 and LAPS2 scores.

We chose a template size of 250. The size of the template was restricted by the number of patients treated in the smallest hospitals in each diagnosis group after factoring in the joint distribution of sex, admission type/venue, and the proportion of patients admitted during flu season. We excluded small hospitals with small numbers of patients in at least one disease category from the performance comparison, because the patient mix would be too different to allow comparisons

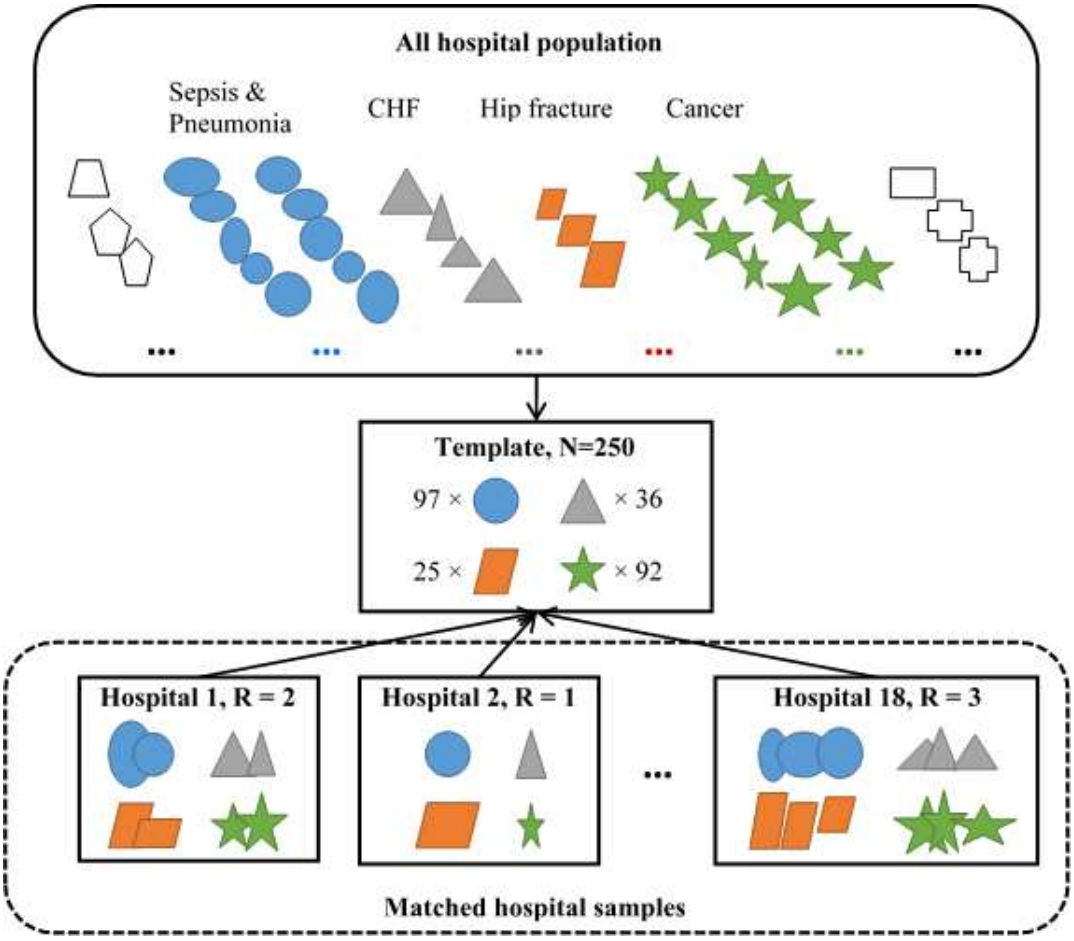


Figure 2.1: Each shape represents patients from one diagnosis group (circle: sepsis & pneumonia, triangle: CHF, diamond: hip fracture, star: cancer; trapezoid, pentagon, rectangle and cross represent other disease categories seen by the 18 hospitals but not included in this study). The proportion of each of the 4 diagnosis groups in the template is the same as that in the pooled hospital population. For conciseness of presentation, we use only one shape to represent all patients selected under each diagnosis group, and label the exact number of patients from each diagnosis group next to each shape. Each patient in the template is matched to R patients from each of the 18 hospitals, where R is the matching ratio varying between 1 and 5 depending on the number of cases in each hospital. In the figure, for example, $R = 2$ for hospital 1 and $R = 1$ for hospital 2. The number of sepsis & pneumonia patients matched in hospital 1 would therefore be $97 * 2 = 194$. The matched patients at each hospital are very similar to the template, and patients from all 18 hospital matched samples are also very similar to each other.

to other hospitals, and including them would come at the cost of significant reduction in template size and statistical power. With a template of 250 patients, we ensured that there were nearly double the number of cases to choose from even the smallest hospitals to achieve good matches. For the largest hospitals, each patient in the template could be matched to up to five patients to

take advantage of the larger case population.

To construct the template, similar to Silber *et al.* (2014a), we randomly generated 500 subsamples of size 250 from the pooled population. Each of these had the pre-determined proportion of diagnosis groups and admission type/venue as mentioned above. From these 500 random samples, we selected the random sample that was closest to the pooled population for that diagnosis group in terms of the Mahalanobis distance between the mean patient characteristics.

2.2.4. Matching Similar Hospital Samples

We examined how hospital rankings change when matching on 4 data type combinations: (1) administrative data only, (2) administrative data and COPS2, (3) administrative data and acute physiology (LAPS2), (4) administrative data, COPS2 and LAPS2. Note that current hospital comparison methods, such as that used by the CMS, lie between our matching design (1) and (2), because they take into account administrative data and a more simplified comorbidity index than COPS2.

For each data type combination and diagnosis group, we performed multivariate matching between the template and each hospital, minimizing the Mahalanobis distance between the matched pairs of patients while balancing the covariates using the methods in Zubizarreta (2012). These were implemented using the statistical software package 'designmatch' for R (Zubizarreta *et al.*, 2016). We calculated the Mahalanobis distance using all continuous covariates. To take advantage of the different number of cases in small and large hospitals and increase matching efficiency, we varied matching ratios from 1:1 up to 1:5. In summary, we required exact matches for the 4 diagnosis groups, matching on the proportions of each categorical variable (sex, flu season admission indicator[2], admitting type, and admission venue), and mean matching for the continuous variables (age, COPS2, LAPS2). For these latter variables, the difference between the mean patient ages in the template and the hospital matched samples was restricted to be within 0.1 standard deviation (Rosenbaum and Rubin, 1985). In addition to balancing the means between the template and

[2]Indicator on whether a patient is admitted during the flu season, so that after matching the fraction of total admissions that were admitted during the flu season is the same across all hospitals

the overall hospital matched samples, we wanted the matched pairs of patients to be as close on COPS2 and/or LAPS2 as possible to ensure each pair of patients is also comparable to each other. Therefore, for pairs of patients who were more than one standard deviation away from each other on COPS2 or LAPS2, we modified their pairwise distance with an added penalty equal to the mean pairwise Mahalanobis distance.

2.2.5. Assessing the Quality of Matching

We assessed quality of the matches across two dimensions: (1) we compared the covariate distributions across 18 hospitals; and (2) for each of the 4 data type combinations, we compared how balanced the 18 hospital matched samples were with respect to age, COPS2 and/or LAPS2.

To compare the covariate distributions across 18 hospitals, we used a formal Kruskal-Wallis test (Kruskal and Wallis, 1952) for continuous covariates and a Pearson's χ^2 test for categorical covariates to examine the univariate balance on each covariate across 18 hospitals. The Kruskal-Wallis test is a non-parametric analogy to the one-way ANOVA test based on ranks, and it tests if there is stochastic dominance on the covariate between some pairs of hospitals. We then inspected the differences in age, COPS2 and LAPS2 among pairs of patients from 18 hospitals, who were matched to the same template patient, which we referred to as a matched group to this template patient. If a matched patient from one hospital is very different from a matched patient from a different hospital, while both are being matched to the same template patient, then significant variation within this matched group may remain, and patients from the 18 hospitals in this matched group may not be comparable to each other.

2.2.6. Ranking Hospital Performances

We focused on two patient outcomes in measuring hospital performance: 30-day mortality rates and LOS in hospital. For the latter outcome, we derived Hodges-Lehmann point estimates (Hodges and Lehmann, 2012). The Hodges-Lehmann estimator is a robust nonparametric location estimator of the LOS distribution based on Wilcoxon's signed-rank statistic, and has been used in other studies to measure hospital LOS (Silber *et al.*, 2014a), time to discharge (Shuster *et al.*, 2008) and

intensive care unit free days (Young *et al.*, 2015), etc. Under each level of information matched on, leveraging the paired group nature in the 18 hospital matched samples and the varying matching ratios, we used a Cochran-Mantel-Haenszel test (Agresti, 2012) for testing the differences in 30-day mortality rates across 18 hospitals. The Cochran-Mantel-Haenszel test examines whether 30-day mortality rates are the same across all 18 hospitals stratified by the 250 matched patient groups. For the continuous outcome of hospital LOS, we used a Mark-Skillings test (Agresti, 2012) to check if the LOS distribution in all 18 hospitals was independent from the hospital assignment. Because patients who died in the hospital typically exhibited a different (shortened) pattern of in-hospital LOS, we set their LOS to be the 95th percentile of the pooled LOS distribution (12.8 days), similar to the modification methods used in Lin *et al.* (2017); Brock *et al.* (2011); Liu *et al.* (2010), and Silber *et al.* (2014b). Setting the LOS of deceased patients to the 95th percentile of the pooled LOS distribution helps to reflect that death is ultimately the worst outcome. Some studies drop the patients who died in hospital when analyzing LOS. We decided to retain such patients because we utilized the matching structure to perform stratified statistical tests on patient outcomes across hospitals, where each group of patients matched to the same template patient formed a natural stratification. Dropping patients who died in hospital after matching when analyzing LOS would impair the use of stratified statistical tests.

2.3. Results

Our final study cohort consisted of 41,620 patient hospitalizations that began at one of the 18 hospitals between Jan 2010 and Nov 2011. The proportions of hospitalizations for sepsis & pneumonia, CHF, hip fracture and cancer were 39.3%, 14.4%, 9.7% and 36.7%, respectively.

2.3.1. Defining a Representative Patient Sample

Table 2.1 summarizes the mean patient characteristics in the template compared with that in the pooled population for each diagnosis group. None of the differences in means between the template and the pooled population was statistically significant at the 5% level. Thus, we concluded that

the template was comparable to the patient population seen in all 18 hospitals.

Table 2.1: Mean patient characteristics in the Template ($N = 250$) and the Pooled Population ($N = 41,620$)

Diagnosis Group	Sepsis & Pneumonia		CHF		Hip Fracture		Cancer	
	Pop.	Temp.	Pop.	Temp.	Pop.	Temp.	Pop.	Temp.
Proportion (%)	39.3	38.8	14.4	14.4	9.7	10.0	36.7	36.8
Age	71.0	71.1	75.4	75.3	80.0	79.7	58.3	57.9
COPS2	53.7	54.7	66.6	68.7	33.4	31.6	24.1	25.2
LAPS2	98.7	98.9	84.2	85.1	58.5	56.0	22.7	23.1
Female (%)	50.4	50.5	48.6	50.0	70.1	68.0	63.3	64.1
Flu season (%)	32.8	39.2	33.8	38.9	29.3	36.0	28.5	32.6
Admission category (%)								
ED surgical	2.3	2.1	0.7	0.0	50.9	56.0	1.2	1.1
Non-ED surgical	0.2	0.0	0.6	0.0	3.8	4.0	75.2	73.9
ED medical	93.0	92.8	91.3	91.7	44.6	40.0	4.2	4.4
Non-ED medical	4.6	5.2	7.5	8.3	0.7	0.0	19.4	20.7

CHF indicates congestive heart failure; COPS2, COmorbidity Point Score 2; ED, emergency department; LAPS2, Laboratory Acute Physiology Score 2; Pop., population; Temp., template.

2.3.2. Comparing the Quality of Matching

Table 2.2 examines the maximum within-group differences on COPS2 and LAPS2. When only matching on the administrative data, considerable variation in COPS2 and LAPS2 remained within each matched patient group. Balancing both administrative data and COPS2 helped shrink the mean within-group variation in COPS2 from 144.3 to 15.3, but the variation in LAPS2 remained. A similar relationship appeared when balancing administrative data and LAPS2, but not COPS2. Only by actively balancing both the administrative data and the two severity scores could we meaningfully reduce the heterogeneity of patient severity of illness within each matched group.

Table 2.2: Summary statistics on the maximum differences within each matched patient group when matching on (1) administrative data only; (2) administrative data and COPS2; (3) administrative data and LAPS2; (4) administrative data, COPS2 and LAPS2

	Admin			Admin + COPS2			Admin + LAPS2			Admin+COPS2+LAPS2		
	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
Age	0.0	1.7	17.0	1.0	6.4	62.0	1.0	5.8	28.0	2.0	11.2	62.0
COPS2	30.0	144.3	269.0	0.0	15.3	216.0	27.0	139.2	254.0	0.0	25.5	259.0
LAPS2	27.0	127.7	245.0	24.0	127.1	278.0	0.0	14.7	75.0	0.0	28.0	108.0

Additional tables in Appendix A.1 present covariate distributions across the 18 hospital matched samples under 4 data type combinations. In examining covariate balances across the 18 hospital matched samples, we found that age was well balanced under all 4 levels of information. Because we matched exactly on hospital and fine balanced the distribution of sex, admission season and admission type, the proportions of female, flu season admissions, and the admission type/venue combinations were exactly the same in each hospital matched sample. In contrast, whenever COPS2 and LAPS2 were not included in the matching algorithms, substantial variation in these severity scores persisted.

2.3.3. Ranking Hospital Performance

After matching, all 18 hospital samples had significantly different 30-day mortality rates and LOS distributions at the 1% level and lower based on the Cochran-Mantel-Haenszel test and the Mark-Skillings test.

Figure 2.2 shows hospital rankings for both 30-day mortality and Hodges-Lehmann LOS point estimates under matching designs (1) and (4). Because the distribution of LOS was highly skewed to the right, we ranked hospitals based on the Hodges-Lehmann estimates of LOS in the hospital matched samples[3].

We also broke down the hospital rankings into the top, medium and bottom terciles and examined how hospitals moved up or down in their ranking terciles when adding COPS2 and/or LAPS2 to the comparisons. If COPS2 and/or LAPS2 provided no additional information, we would expect hospitals not to change their ranking terciles when adding them into performance comparisons, and we would reject the null hypothesis of independence between rankings. Tables 2.3 and 2.4 show that the p-values for change of hospital ranking terciles between administrative information and adding COPS2 or LAPS2 were significant at the 5% level for both 30-day mortality rates and Hodges-Lehmann estimates of LOS. This rejects the null hypothesis that the two rankings were

[3]The Hodges-Lehmann estimator estimates the pseudo-median when the underlying distribution is asymmetric, or it is a median-unbiased estimator if the underlying distribution is symmetric. Because the LOS distribution is typically highly skewed, using its mean tends to be biased. Hodges-Lehmann estimator tends to have greater efficiency than does the sample median. Rather than fitting a distribution to the LOS and estimating its mean or median, we decided to do a robust nonparametric comparison across hospitals.

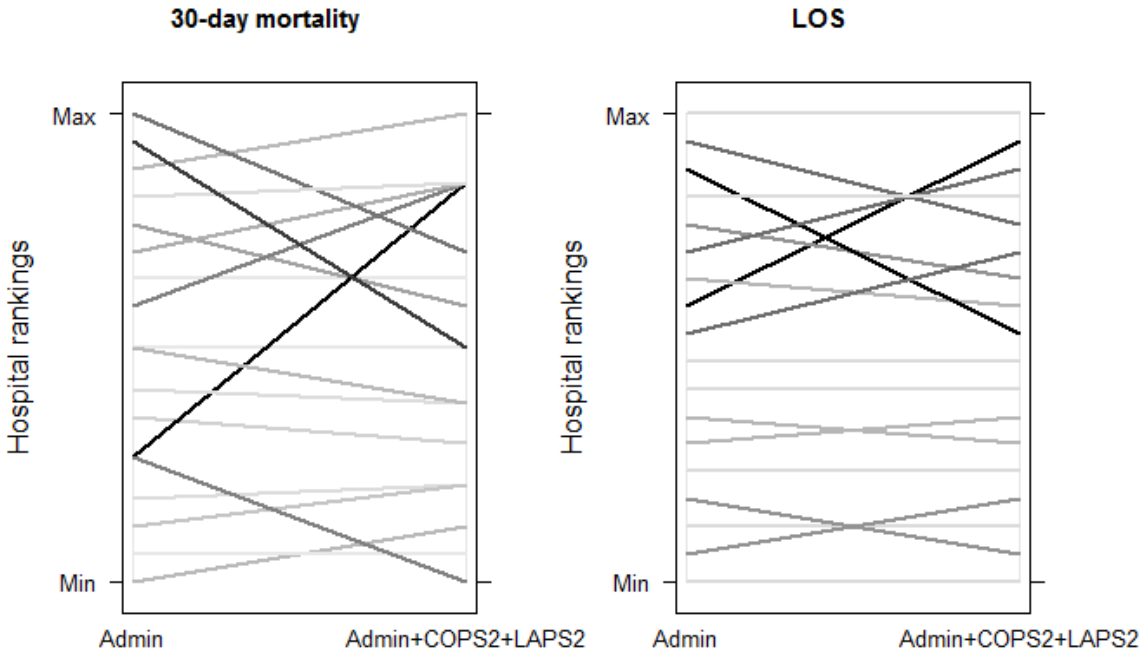


Figure 2.2: Parallel plot of 18 hospital rankings (min rank = 1, max rank = 18, y-axis) on 30-day mortality (left) and LOS (right) when matching on administrative data (Admin) or adding both COPS2 and LAPS2 (x-axis). Each line links a hospital's ranking against all other hospitals under the two levels of information. If a hospital's rankings on 30-day mortality rates or LOS are the same under the two data type combinations, the line should be horizontal. The darker the line color, the more changes in hospital rankings under the two levels of information matched on.

independent. Only when adding both COPS2 and LAPS2 were hospital ranking terciles independent from those using only administrative information. Thus, COPS2 and LAPS2 have significant effects on hospital performance rankings. Additional data on the mean and max absolute changes in hospital rankings are provided in Appendix A.2.

2.3.4. Sensitivity Analysis

In template matching, we have already controlled for observed patient characteristics, including disease group, sex, admission flu season, admitting type, admission venue, age, COPS2 and LAPS2. By controlling for observed patient characteristics, it is likely that we are also implicitly controlling for the distribution of unobserved patient risk factors, assuming observed patient characteristics are proxies to unobserved risk factors. Nevertheless, there may still exist some concern that the

Table 2.3: Number of hospitals in each combination of 30-day mortality rate ranking terciles between using administrative information only and when adding COPS2 or LAPS2 or adding both. P-value is calculated using the Fisher's exact test to examine whether the hospital ranking terciles are independent under any two levels of information.

	Admin + COPS2			Admin + LAPS2			Admin + COPS2 + LAPS2		
	Top	Medium	Bottom	Top	Medium	Bottom	Top	Medium	Bottom
Top	5	1	0	5	0	1	3	3	0
Medium	0	3	3	1	4	1	1	3	2
Bottom	1	2	3	0	2	4	2	0	4
P-value	0.04			0.01			0.08		
	Correlated			Correlated			Independent		

Table 2.4: Number of hospitals in each combination of truncated LOS ranking terciles between using administrative information only and when adding COPS2 or LAPS2 or adding both. P-value is calculated using the Fisher's exact test to examine whether the hospital ranking terciles are independent under any two levels of information.

	Admin + COPS2			Admin + LAPS2			Admin + COPS2 + LAPS2		
	Top	Medium	Bottom	Top	Medium	Bottom	Top	Medium	Bottom
Top	5	1	0	5	1	0	4	2	0
Medium	1	3	2	1	3	2	2	2	2
Bottom	0	2	4	0	2	4	0	2	4
P-value	0.02			0.02			0.10		
	Correlated			Correlated			Independent		

distribution of some unobserved patient risk factor is not similar across the 18 hospitals, which may impact how patients are assigned to hospitals (i.e., in a non-random way). A sensitivity analysis asks whether a departure of a specific size Γ (in odds ratio) from random hospital assignment would alter the conclusions about a specific hospital A of interest as compared with another hospital B or the rest 17 hospitals as a whole. We used a stylized sensitivity analysis model in Rosenbaum (1987) to assess how sensitive hospital rankings could be to a potential unobserved covariate U that impacts patients' hospital choice and outcome.

Specifically, let Z_i be a binary variable indicating whether a patient i chooses to go to hospital A, U_i be an unobserved covariate that is not controlled by matching, and X_i be the vector of observed covariates matched on. For the unobserved covariate to have the maximum impact on biasing the patient outcome, it is typically assumed that U_i is nearly perfectly predicting the patient outcome,

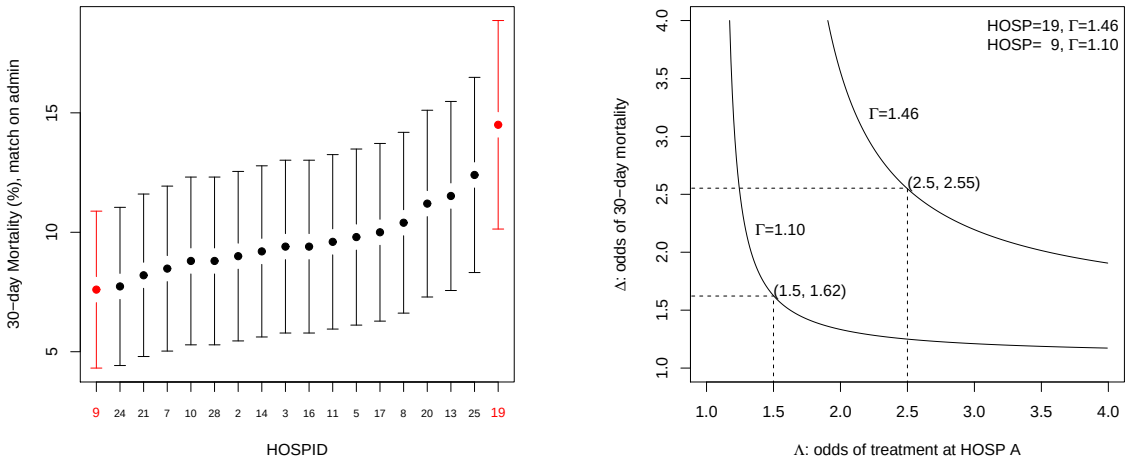
and is positively correlated with a patient choosing to go to hospital A versus the rest 17 hospitals. Following Rosenbaum (1987), we assume the distribution of Z_i given (X_i, U_i) follows a logit model

$$\log \frac{Pr(Z_i = 1|X_i = x_i, U_i = u_i)}{Pr(Z_i = 0|X_i = x_i, U_i = u_i)} = f(x_i) + \gamma u_i.$$

If $\gamma = 0$, then given the observed covariates X_i, the hospital choice Z_i is not affected by the unobserved covariate U_i, as would be the case in a conventional randomized experiment. For convenience of interpretation on the sensitivity parameter γ as in Rosenbaum (1987), we assume $U_i \in [0, 1]$ without loss of generality. Therefore, the least favorable values of U_i are binary, and two patients matched exactly on X have odds of choosing hospital A that differ by at most a factor of $\Gamma \triangleq e^{\gamma}$. As an aid to interpreting Γ, the one parameter model can be understood in terms of an equivalent two-dimensional model with parameters Λ and Δ through the formula $\Gamma = \frac{\Delta\Lambda+1}{\Delta+\Lambda}$, where Λ is the odds ratio of choosing hospital A when $Z = 1$ vs when $Z = 0$, and Δ is the odds ratio of death in 30 days when $Z = 1$ vs when $Z = 0$. See Rosenbaum (2002); Rosenbaum and Silber (2009) for a detailed discussion of Γ and its auxiliary parameters Λ and Δ.

For each hospital i of interest, we can calculate the sensitivity parameter Γ_i associated with the odds of choosing hospital i versus the rest 17 hospitals as a whole. Figure 2.3(b) shows the sensitivity of the two hospitals ranking the best and worst in terms of their 30-day mortality rates. Using hospital 19, which ranks the worst on 30-day mortality, as an example: in order to attribute the higher 30-day mortality rate among patients treated at hospital 19 to an imbalance in an unobserved covariate U, that unobserved covariate would need to more than double the odds of patients choosing to go to hospital 19 versus the rest 17 hospitals ($\Lambda = 2.5$), and *at the same time* more than double the odds of death in 30-days ($\Delta = 2.55$), corresponding with $\Gamma = 1.46$. Our physician collaborator indicated that in KPNC, such an unobserved covariate is not highly likely to exist as all hospitals fall within the same integrated healthcare delivery system. Therefore, the conclusion on hospital 19 ranking the worst in 30-day mortality rate is moderately insensitive to unobserved covariates.

Figure 2.3: Sensitivity of 30-day mortality rates rankings to unobserved covariates



(a) Hospital 30-day mortality rates with 95% confidence intervals

(b) Sensitivity of 30-day mortality rates rankings

2.4. Discussion

In the US, numerous studies have indicated that the quality of American health care is highly variable across and within medical systems (Institute of Medicine, 2001; McGlynn *et al.*, 2003). For example, the Dartmouth Atlas of Health Care has published over 60 reports since 1996 on the regional level variations in medical resources utilization and outcomes for the Medicare population. In 2012, the Organization for Economic Cooperation and Development undertook an international study that showed the wide health care variations for high-cost and high-volume procedures across and within 13 countries (OECD, 2014).

Recognizing the large variations in health care quality, hospitals have invested time and effort into finding ways to objectively measure their performance against other hospitals to identify potential areas for quality improvement. Following the release of national public data from CMS in 2004, numerous works have studied hospital performance on Medicare beneficiaries measured by the 30-day mortality and readmission rates for common medical conditions, including acute myocardial infarction, heart failure, pneumonia and stroke.

In this work, we have demonstrated how an integrated health care delivery system could use

the template-matching approach to take advantage of data from comprehensive EMRs to conduct objective internal hospital ranking on the same reference population. Further, our analyses also confirm that both patients' comorbidity history and acute physiology at the time of admission had substantial effects on hospital rankings. Neglecting either the comorbidity history or acute physiology would raise concerns about the comparability of the patient population and, consequently, the validity of the derived hospital rankings.

Our findings on the importance of incorporating physiologic conditions in hospital comparison are consistent with prior work using risk-adjustment models. The impact of incorporating the comorbidity history on hospital rankings is more prominent in our study as compared with that found in our previous work (Escobar *et al.*, 2013). As the patient population and indices are the same, the most likely reason is the risk adjustment process itself, although the work of others (Quail *et al.*, 2011) suggests that some of the effects we found could be due to varying effects of the indices themselves.

Compared with the common risk-adjustment models used in prior work, template matching gives a more objective way of auditing hospital quality by grading all hospitals on the same reference population that is typically seen by all 18 KPNC hospitals, thereby reducing the risk of model misspecification and extrapolation when there is limited overlap in covariate distributions across hospitals. In addition, template matching is more transparent because one knows exactly whether an individual patient is included or not in the comparison.

The template-matching methodology can be readily generalized to performance comparisons at other hospitals, with the flexibility to select the target patient population and risk covariates. In addition, Dr Escobar's team is working with Epic to embed LAPS2 and COPS2 into the Epic EMR core structure (currently used in more than 4,000 US hospitals), making it more widely available in the United States for hospital comparisons.

This study has several limitations. First, our study focused on four large and severe diagnosis groups typically treated at KPNC. It would be interesting to generalize to other large disease categories seen at other health care systems. Second, our data covered patient hospitalizations over the course of only two years, which limited the template size, and results from the smallest

hospitals might be less stable. Nevertheless, the severity of illness of the patient population seen at KPNC hospitals has been increasing over the past decade and KPNC has been continuously launching practice improvement initiatives. Therefore, grading hospitals on earlier years may not be comparable to their current practice and such comparison may not be objective. Finally, our inferences on hospital performance were conditional on the matching design. An interesting open methodological question for future study would be how to incorporate uncertainty in the hospital ranking estimates.

Given the increasing availability of data from comprehensive EMRs, which can be merged with administrative data even at the state level (Rosenthal *et al.*, 2010), our study suggests it might be worthwhile for hospital administrators and policy makers to consider regularly incorporating patients' comorbidity history, laboratory results, and vital signs for objective quality comparisons between hospitals. With more comprehensive adjustment for patient populations under comparison, hospital administrators can understand hospital performance better and identify best practices to improve quality of care.

Part II

ICU Admission Decisions from the Ward and the Emergency Department

Chapter 3

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

3.1. 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 3.1. 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).

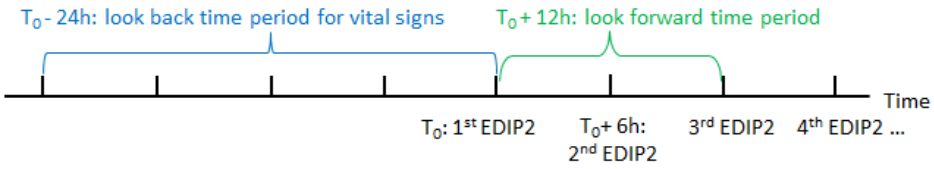


Figure 3.1: Timeline for the EDIP2 score

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. One specific challenge with using this dataset is that during the time period of the data, EDIP2 scores were not available to physicians who decided whom to transfer to the ICU on a patient-by-patient basis based on general diagnostics, and our data does not explicitly label ICU transfers as proactive or reactive. To the best of our knowledge, there are no datasets which explicitly indicate whether an admission was a proactive or reactive transfer. For each patient, we have data on his/her retrospective EDIP2 score provided at fixed six-hour increments whenever the patient is in the ward or TCU as well as whether he/she was transferred to the ICU during that period. Due to the natural variation in the ICU transfer decision (e.g. due to ICU congestion), some patients are transferred at much lower severity than other patients. Moreover, some patients of the same severity will be transferred to the ICU, while others will not. We leverage this variation to rigorously and robustly identify the causal effect of transferring patients at various levels of severity, as measured by the EDIP2 score. From electronic data, it is typically impossible to tell whether a patient was admitted proactively (before the patient really needs it); it is in the context of a simulated, structural model that we examine proactive decisions.

Another 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 the results of our empirical findings to calibrate a high-fidelity simulation model, which will allow us to examine how various

proactive ICU transfer policies that activate at different levels of the EDIP2 score—before a patient crashes and requires an immediate ICU transfer—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.

3.1.1. 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 health-care 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).

3.2. 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 3.2 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).

3.2.1. 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 3.1.

3.2.2. 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

Table 3.1: 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

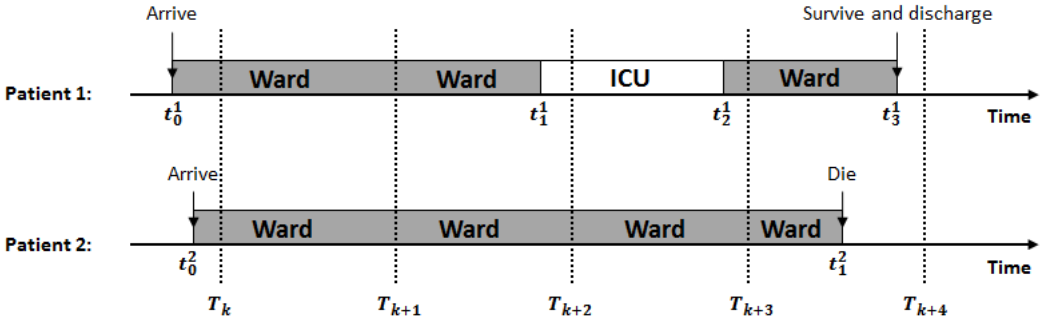


Figure 3.2: Examples of patient pathways. Each T_i denotes a time when an updated EDIP2 score will be assigned to a patient if he/she is in the Ward or TCU. Note that there are exactly 6 hours between each EDIP2 assignment: $T_{i+1} - T_i = 6$.

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 3.2 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.

3.2.3. 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 3.2, 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 3.2 summarizes the statistics for in-hospital mortality and hospital remaining length-of-stay considering the first EDIP2 decision epoch.

Table 3.2: 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

3.3. 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.

3.3.1. Empirical Challenges

In our study, we utilize the retrospective patient dataset described in Section 3.2. 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).

3.3.2. 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.

3.3.2.1. 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 B.1). 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.

3.3.2.2. 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 B.1.1 for details.

3.3.3. 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^{th}, 90^{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 B.1.3.1.

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 3.2 in Section 3.2.3). 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 non-busyness ($Z_i = 1$ if ICU is *not* busy, 0 if busy), 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 B.1 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

$$T_i = \mathbb{1}\{T_i^* > 0\} \qquad \text{where} \qquad T_i^* = X_i^T \beta_1 + \beta_2 Z_i + \epsilon_i \tag{3.1}$$

$$\log Y_i = X_i^T \beta_3 + \beta_4 T_i + \eta_i \tag{3.2}$$

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 3.2). 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 (3.1). 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 patient with a correlation coefficient of -0.168.

3.4. 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.

3.4.1. 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 B.1.1 for further details on the near-far matching implementation using integer programming. Tables B.2–B.4 in Appendix B.1.2 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 3.3.3. The results are summarized in Table 3.3. 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 non-busyness (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 3.3: Strength of the IV in the whole-day full sample and the first night-time IV after match in probit regression models, $IV = 1$ if ICU is *not* busy

	Sample Size	*IV*	(Std. Err.)	P-val.	% 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%**

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

Table 3.4 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 B.1.3. 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 3.3.

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 3.4: Estimation results using the night-time IV after matching, $IV = 1$ if ICU is *not* busy

Y	*IV* (SE)	% 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 3.5.

A limitation to our data set lies in that from electronic data, it is typically impossible to tell whether a patient was admitted proactively (before the patient really needs it) or because of crashing. To the best of our knowledge, there are no datasets which explicitly indicate whether an admission was an "early" or "crashing" transfer. Our objective is to evaluate the impact of transferring patients at differing severity levels. Due to variation in the ICU transfer decision, some patients are admitted at much lower severity than other patients. Moreover, some patients of the same severity will be transferred to the ICU, while others will not. As such, we aim to determine the causal effect of transferring a patient to the ICU at varying levels of severity. If the impact

on mortality and LOS on transferring crashing patients is higher than that on "early" transfers, then it is possible that our empirical estimates may be over-estimating the impact of transfers on mortality and LOS.

3.4.3. 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 3.5.

Table 3.5: 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.

		Estimated Coefficients (s.e.)	
Model	Outcome Measure	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)

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

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 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.

3.4.4. 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 3.6 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 3.6: 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

3.5. 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.

3.5.1. 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 3.3. 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 **N**.

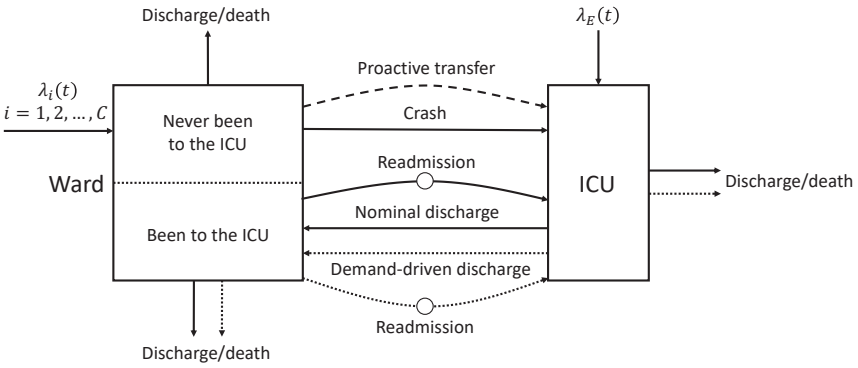


Figure 3.3: Simulation Model

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-homogeneous 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.

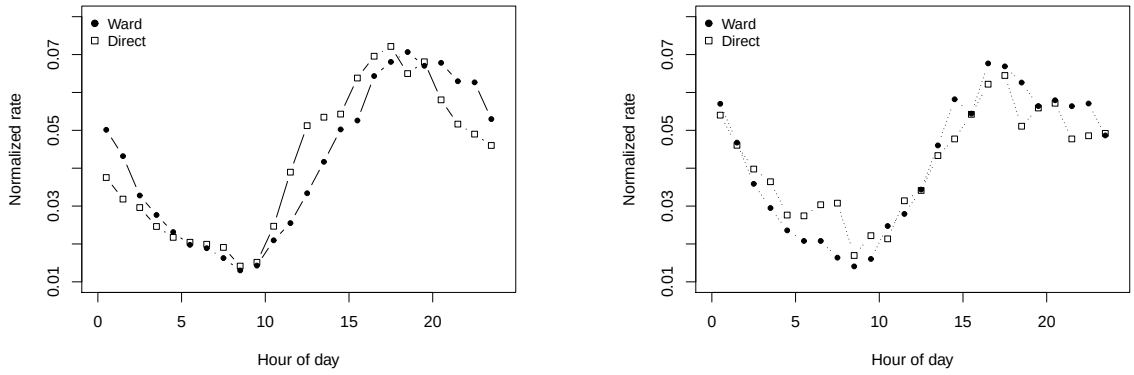
In principle, any patient in the ward can be *proactively transferred* to the ICU at each EDIP2 decision epoch. Such proactive transfers can only occur if there is an available ICU bed for the transferred patient. If there are not enough available beds in the ICU to accommodate all proactive ICU transfer 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.

3.5.2. Model Calibration

We now calibrate our simulation model using the data described in Section 3.2 and our empirical results from Section 3.4. Figure 3.4 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 3.4: 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

3.5.2.1. 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 3.7. Note that we use the empirical distribution for p_E (see Figure B.4 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 3.7: 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	$1.15 \times r_E$	

3.5.2.2. 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 3.8 summarizes the partitioning of these 10 groups.

We use our full dataset from KPNC to calibrate the Markovian transition matrix $\mathbf{T} \in \mathbb{R}_{[0,1]}^{10\times 13}$ (see Appendix B.2.1). 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 3.2. 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

Table 3.8: 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%

transition matrices range from 1.04–3.17%.

We leverage our empirical findings from Section 3.4 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.2.2) 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 B.4 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 3.9 and 3.10 summarize the parameters for ward patients.

Table 3.9: 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 3.10: 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

3.5.3. 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 3.4.4. In Appendix B.2.3, 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.

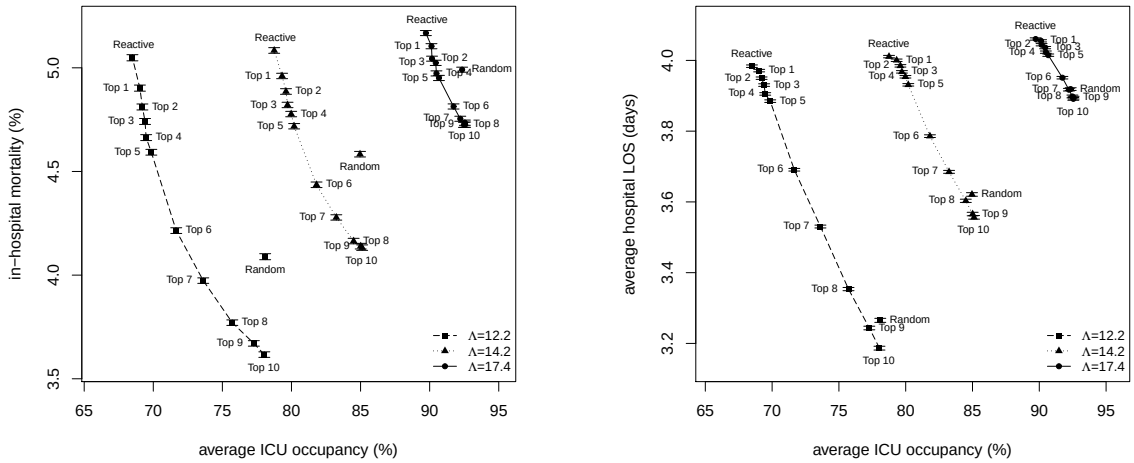
We define a **Static Threshold Policy** by threshold T_{EDIP2}. Any patient in EDIP2 group $i \geq T_{EDIP2}$ will be proactively transferred if there are available ICU beds. If the EDIP2 score is

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 3.5.5.

3.5.4. 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 3.5: In-hospital mortality rate and mean hospital LOS under various proactive ICU transfer policies, with 95% confidence intervals. ICU size $\mathbf{N} = 15$. Λ = daily arrival rate. Proactive transfers can only take place at night.



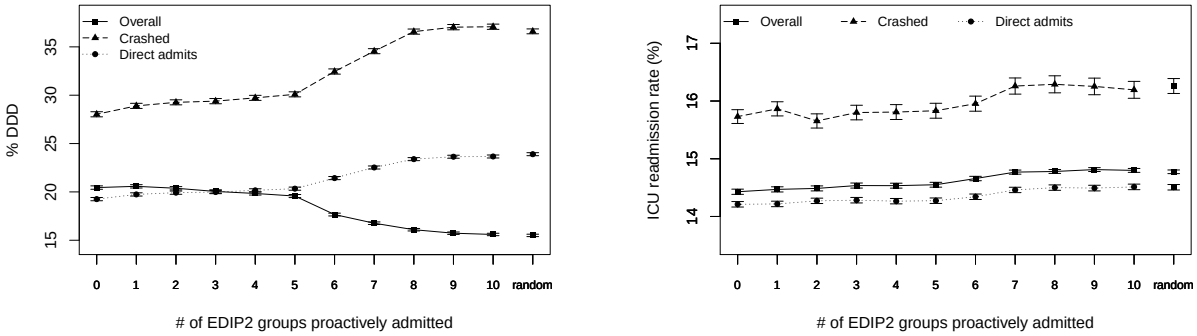
(a) In-hospital Mortality Rates

(b) Hospital LOS

Figure 3.5 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.

Figure 3.6: Adverse event rates under various proactive ICU transfer policies, with 95% confidence intervals. ICU size $\mathbf{N} = 15$. $\Lambda = 14.2$ patients/day. Proactive transfers can only take place at night.



(a) Demand-driven discharge rates

(b) Readmission rates

As seen in Figure 3.6, 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 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 3.7(a) is an analog to Figure 3.6(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 3.7(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.

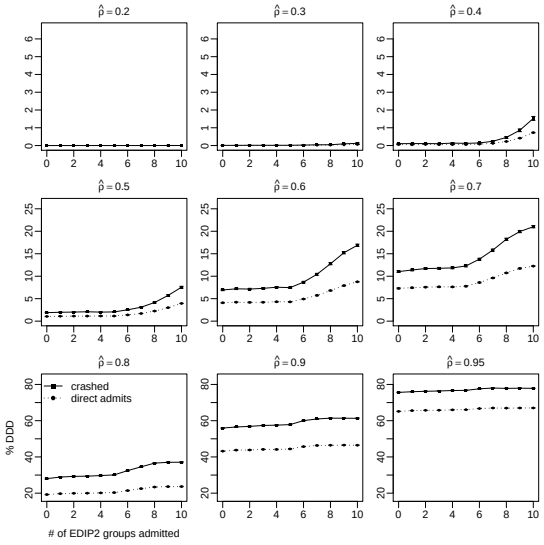
Note that in all of our experiments, the random policy is Pareto dominated by the static threshold policies. This is true even when we consider other random policies which aim to proactively transfer a similar number of patients as under the static threshold policies. Appendix B.2.3 provides additional simulation results which demonstrate the robustness of our main insights.

3.5.5. State-Dependent Policies

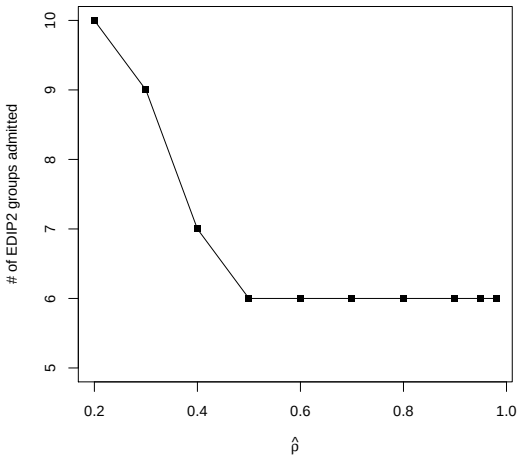
We also consider a modification of the Static Threshold policy, where instead we consider state-dependent thresholds (e.g. Altman *et al.* (2001)). For these experiments, we focus on the baseline scenario of $\mathbf{N} = 15$ beds and $\Lambda = 14.2$ patients/day.

As our model incorporates many features (e.g., demand-driven discharges, readmissions, etc.),

Figure 3.7: ICU size $\mathbf{N} = 15$. $\hat{\rho}$ 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.

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.

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 3.11 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 B.7 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 3.11: 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

3.5.6. 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 3.5.5, 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

$\xi_{it} \sim N(0,1)$.

$$Admit_{it} = 1\{\beta_{EDIP2} EDIP2_{it} + X_{it}\theta + \xi_{it} \geq \kappa_{occ_{it}}\} \tag{3.3}$$

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 3.8). Therefore, the estimated admission policies at both hospitals correspond to admitting only the most severe patients. As we have seen in Figure 3.5, 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.

3.6. 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 night-time 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. Another limitation of our dataset lies in that it does not identify whether patients are transferred to the ICU proactively or reactively. Therefore, the estimated impact of ICU transfer at each EDIP2 score could be the average impact on a mix of proactively transferred and "crashed" patients. Because crashed patients are likely to have worse outcomes than proactively transferred patients, the estimated effect of ICU transfer on patient outcomes may be underestimated for crashed patients as their outcomes could

have been better if proactively transferred before crashing.

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.

Chapter 4

A Structural Model of Admitting Inpatients from the Emergency Department

4.1. Introduction

Hospital Emergency Departments (EDs) provide services to individuals with emergency clinical needs or after-hour care. With the aging U.S. population, overall ED usage has been increasing over recent decades (National Center for Health Statistics, 2017). After the ED triage and initial treatment, emergency providers determine whether a patient can be discharged home safely or requires hospitalization, and if so, which inpatient unit the patient should be admitted to. Such patient disposition decisions were made 130 million times in 2013 or on average 42 times per 100 patients, resulting in over 12 million annual admissions to hospitals from the ED (Rui *et al.*, 2013). Hence, the ED is a primary source for hospitalizations in the United States.

Inpatient flow management from the ED is often faced with high variability in demand and high utilization in bed capacity, particularly for the Intensive Care Unit (ICU) which often operates under high congestion levels. When making inpatient admission decisions, emergency providers often trade off between admitting a patient immediately to the ICU, placing the patient in the

general ward, or letting the patient continue to wait in the ED if there is downstream congestion in the inpatient units (referred to as *boarding* in the ED). Patients boarding in the ED or rerouted to the ward due to ICU congestion have been shown to be associated with increased mortality and extended length-of-stay (LOS) (Robert *et al.*, 2012; Shmueli *et al.*, 2003; Kim *et al.*, 2015; Chan *et al.*, 2017). Alternatively, admitting the new and more critical patient from the ED when the ICU is full by discharging a current ICU occupant may increase the likelihood of readmission for the 'demand-driven' discharged patient (Kc and Terwiesch, 2012). A typical ED admission process is that the ED physician would page the admitting physician and discuss the patient's condition over the phone, or the admitting physician would come down to the ED to assess the patient and issue admission orders. Despite several general guidelines on patient management in the ED, the admission decision remains a matter of individual clinical judgment on the patient's need for intensive care, and competing demands on ICU and non-ICU resources (Mandell *et al.*, 2007). Studies have demonstrated that the ED admission decision exhibits a wide range of practice across emergency providers and EDs, partially affected by local standards of care and the availability of primary care (Pines *et al.*, 2013).

This work aims to understand how ED physicians trade off between admitting a patient to the ICU, the general ward, or boarding in the ED, and the extent to which physicians are forward looking when making the admission decision. Specifically, we propose a model to estimate how much ED physicians account for the intertemporal externalities in making the admission decision because admitting a patient in the current period takes up the inpatient resources (especially the limited ICU beds), which can prevent admitting a future patient with potentially more critical conditions. Our study setting is Kaiser Permanente Northern California (KPNC), an integrated health care delivery system with 21 hospitals.

We first use two sets of reduced-form regressions to understand 1) what and how patient risk factors and system controls impact the admission decision from the ED; and 2) what are the potential benefits of admitting patients from the ED to the ICU. To evaluate which factors affect the admission decision, we use logistic and multinomial logit models on a per patient per decision epoch basis. To estimate the potential benefits of ICU admission on patient outcomes, we use an

instrumental variable approach similar to those in Chapter 3 and Kim *et al.* (2015) on a per patient basis to address the endogeneity of the ICU admission decision.

One limitation of the traditional reduced form analysis is that it cannot recover ED physicians' preference parameters over the admission decisions nor the intertemporal externalities. Therefore, we would not be able to use the estimated parameters to evaluate counterfactual admission policies. To overcome these limitations, we further model the ED physician's admission decision using a dynamic discrete choice structural model. This structural estimation approach allows us to estimate (1) the value physicians place on each admission decision choice; and (2) the intertemporal discount factor physicians apply to future payoff given the current period admission decision. This model has been applied to different areas of economics (Rust, 1987), but to the best of our knowledge, has not been used in the Healthcare Operations Management (OM) setting.

Overall, we find that the severity levels of the patients under consideration, the ICU and ward status, and the seasonality factors all impact physician's ICU admission decision. The sicker the patient, the more likely he/she will be admitted to the ICU rather than rerouted to the ward. Both ICU congestion and the severity of ICU occupants discourage patients to be admitted to the ICU from the ED, and such patients are more likely to continue boarding in the ED. On the other hand, with more recent ICU discharges, patients are more likely to be admitted to the ICU, possibly to preserve working hours from the nurses who would otherwise be discharged with the discharging patient. While the ward census does not seem to discourage rerouting patients to the ward, the average severity of patients in the ward has a high impact in discouraging ward admission. Regarding the benefits of ICU admission, we find that being admitted to the ICU significantly reduces patients' in-hospital mortality, LOS, and readmission rates within one month.

Note that the structural model estimation is still an ongoing process and more investigation is required to fine tune the details. Therefore, we will not discuss the structural model estimation results in this chapter, but only present the modeling framework and key estimation strategy. Our main contributions can be summarized as:

- We provide a framework to empirically estimate the intertemporal externalities on the inpatient admission decision from the ED using an extensive dataset of patient-level information

across 21 KPNC hospitals. To the best of our knowledge, this is the first study to quantify the intertemporal externalities on the ED admission decision. Our results could provide operational insights on how to improve capacity allocation of ICU beds and overall patient outcomes.

- We pioneer in applying the dynamic discrete choice structural model to the Healthcare OM setting to study ED admission decisions. Such structural estimation methods, to the best of our knowledge, have not been used in the healthcare OM literature on EDs previously.

4.1.1. Related Literature

Our work is related to two main bodies of research: 1) the inpatient admission process from the ED in the medical and healthcare operations management fields, and 2) the use of dynamic discrete choice structural models in economics and operations management to understand human decision-making and evaluate counterfactual policies.

In both the medical and healthcare operations management literature, a number of works have examined the inpatient admission process from the ED. One area of focus has been on the prolonged waiting time in the ED for admitted patients, typically due to the downstream congestion in inpatient units (Louriz *et al.*, 2012). Allon *et al.* (2013) show that ED boarding caused by congested inpatient units is a key factor contributing to ambulance diversion. Shi *et al.* (2016) focus on the impact of inpatient discharge policies on the time-of-day boarding. A specific area more relevant to our study is ICU admission control for patients admitted from the ED. On one hand, various studies show that when the ICU is congested, physicians ration ICU beds for more severe patients (Staiger and Stock, 1997; Strauss *et al.*, 1986). Patient rerouting or ED boarding due to ICU congestion are found to be associated with increased risk of death and extended LOS in hospital (Robert *et al.*, 2012; Shmueli *et al.*, 2003; Chalfin *et al.*, 2007; Kim *et al.*, 2015; Chan *et al.*, 2017). On the other hand, when the ICU is full, discharging a current ICU occupant due to admitting a new and more critical patient from the ED may increase the likelihood of readmission (Kc and Terwiesch, 2012).

Our work differentiates from previous literature in the scope of the question and the methodology. First, rather than estimating the impact of patient and hospital characteristics on an individual

patient's routing decision, we take the ED physician's point of view on the entire system and estimate the cost that physicians associate to each admission/waiting decision. More imporatantly, we focus on understanding the intertemporal externalities on the ICU admission decisions (i.e., how admitting a patient now affects the system status, and its ability in admitting patients with possibly more severe conditions in the future). To the best of our knowledge, we are the first to estimate the intertemporal externalities on the ICU admission decisions. Second, traditional reduced form analyses used in the medical and healthcare OM literature do not recover ED physicians' preference parameters over the decision choices; nor can it recover the intertemporal externalities of their ICU admission decisions, and therefore we would not be able to use the estimated parameters to evaluate counterfactual admission policies.

In contrast, we take a structural estimation approach which is prevalent in the economics literature and has been used to study fertility and child mortality (Wolpin, 1984), replacement of bus engines (Rust, 1987), and retirement from a firm (Lumsdaine *et al.*, 1992; Rust, 1989), etc. See Eckstein and Wolpin (1989); Rust (1994); Reiss and Wolak (2007); Aguirregabiria and Mira (2010) for an extensive survey of this literature.

There is also a growing body of structural work in operations management that study the overage/underage costs of a newsvendor with application to reserving operating room time (Olivares *et al.*, 2008), customer waiting costs in the fast food industry (Allon *et al.*, 2011), consumer strategic purchase delay in the airline industry (Li *et al.*, 2014), production smoothing in automotive manufacturing (Bray and Mendelson, 2015), and auto manufacturer and regulator's joint recall decisions (Colak and Bray, 2016). We contribute to this stream of research by applying structural estimation in the ED admission decision setting, an important area in healthcare OM where this approach has not been used.

Specifically, we use a single-agent dynamic discrete choice structural model on ED physicians' ICU admission decision process. Our model is based on the influential work of Rust (1987), and we use a similar nested fixed-point maximum likelihood algorithm to numerically solve the costs physicians associate with each of the three choices and choose the discount factor that maximizes the log-likelihood. The nested fixed-point algorithm is essentially a dynamic programming procedure

nested in the maximization likelihood estimation. Previous literature on structural models has documented that the discount factor is generally not identified nonparametrically (Rust, 1994; Magnac and Thesmar, 2002; Abbring and Daljord, 2016) unless under certain regularity conditions (Abbring and Daljord, 2016). We make parametric assumptions common in the literature and follow the identification condition in Magnac and Thesmar (2002), which is the existence of a pair of states that affects expected future payoffs but not the static payoffs. Moreover, we exploit the richness of our data and use simulations to show that the discount factor can be recovered in our data.

A limitation of this method lies in the computational complexity when the dimension (state space) of the problem increases. Several methods for accelerating the computation have been presented, including conditional choice probability estimator as a non-/semi-parametric approximation (Hotz and Miller, 1993), relative value iteration when the underlying stochastic process is ergodic (Bray, 2018a), and the endogenous value iteration which disregards the utility calculation for exogenous states unrelated to finding the stationary policy (Bray, 2018b). In this work, we focus on the intertemporal externalities and maintain a parsimonious model of the emergency physician's admission decision. This approach gives us the exact solution to the cost parameters which we would need to evaluate any counterfactual admission policy experiments.

4.2. Study Setting

In this work, we consider a retrospective dataset of all 312,306 hospitalizations which began at one of the 21 hospitals in a single hospital network.

On the operational level, our data provide the admission venue (ED or non-ED), and the admission and discharge date and time for each unit a patient 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).

For every hospitalization episode, we have patient-level admission data which includes the pa-

tient's age, gender, admitting hospital, admitting diagnosis, classification of disease 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).

In addition to the three scores assigned at the time of hospital admission, we have two additional severity of illness scores assigned during inpatient stay. First, all patients are assigned a *Simplified Acute Physiology Score 3 (SAPS3)* every time they are admitted to the ICU. The SAPS3 score is a predictor for in-hospital death for ICU patients using patient demographics, diagnosis, vital signs, and laboratory tests preceding ICU admission (Metnitz *et al.*, 2005; Liu *et al.*, 2013). Second, all patients in our dataset have an *Early Detection of Impending Physiologic Deterioration score, version 2 (EDIP2)* (Escobar *et al.*, 2012, 2013) assigned every six hours (at 4am, 10am, 4pm, 10pm) 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 predict the in-hospital deterioration risks. More details can be found in Escobar *et al.* (2012) and Kipnis *et al.* (2016).

4.2.1. Data Selection

We utilize data from all 312,306 hospitalizations to derive the maximum capacity and hourly occupancy level of all inpatient units 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 find that the maximum ICU occupancy varies from 6 to 34 for the 21 hospitals over our study period.

We now describe our data selection process for the final study cohort. We focus on patients who are admitted to a medical service via the ED because 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 admission decisions due to operational factors, such as the availability of resources. In the patient flow data, 16.8% of the total medical inpatient admissions from the ED are ever admitted to the ICU, and 11.9% are admitted directly from the ED to the ICU.

Among all 312,306 hospitalizations, 189,316 are admitted to a medical service via the ED. We first eliminate 12 hospitalizations with unknown patient gender and 88 hospitalizations transferred in from outside of KPNC. Next, we drop 2,160 hospitalizations for patients who are transferred to another hospital from the ED. We then eliminate 38 hospitalizations with inconsistent records on the hospital ID patients are located in. Finally, we remove 9,274 episodes admitted in the first month of our dataset to avoid censored estimates of the inpatient units occupancy level.

The final study cohort consists of 177,744 hospitalizations from 21 hospitals. Out of all hospitalizations, 2.2% are admitted to the inpatient units from the ED immediately after the admission decision is made. Among the 97.8% of patients who experienced ED boarding, the average boarding time is 1.4 hours and the maximum time waiting for admission is 6.1 days. The patient characteristics at the time of hospital admission for the final study cohort are summarized in Table 4.1.

Table 4.1: Characteristics of the final study cohort, N=177,744

	Min	Max	Mean	Median	Std. Dev.
Female (%)			53.20		
CHMR (%)	0.00	98.73	4.28	1.55	8.20
COPS2	0.00	306.00	44.85	28.00	43.17
LAPS2	0.00	294.00	73.80	69.00	37.44
Age	18.00	113.00	67.15	70.00	17.64

4.2.2. Missing Data Imputation

The focus of this work is the admission decision on patients admitted to the medical service from the ED, which may be impacted mostly by the patient severity conditions and level of care needed. It is also conceivable that operational factors, such as the availability of beds and the workload in inpatient units, are likely to influence the admission decision. Therefore, we will be using *all* patient hospitalizations in the raw data to calculate the hourly system census and workload in all inpatient units. In this section, we describe how we deal with the missing data in three severity scores (CHMR, SAPS3, EDIP2) when calculating the hourly system status.

CHMR for patients transferred in from outside of KPNC

In the raw data, there are in total 6,113 hospitalizations who are transported in from outside of the KPNC network. Such patients do not have the in-hospital mortality prediction (CHMR) score due to the limited knowledge of their previous hospitalizations and health conditions at the time of transfer. Because CHMR is a probability estimate falling in $[0, 1]$, we use Eq. 4.1 to model the logit-transformed CHMR scores on all hospitalizations, which also helps to correct the extremely high kurtosis (= 42.8) and right skew (= 5.4). The *AdmitCategory* includes ED or non-ED patients admitted for medical or surgical services.

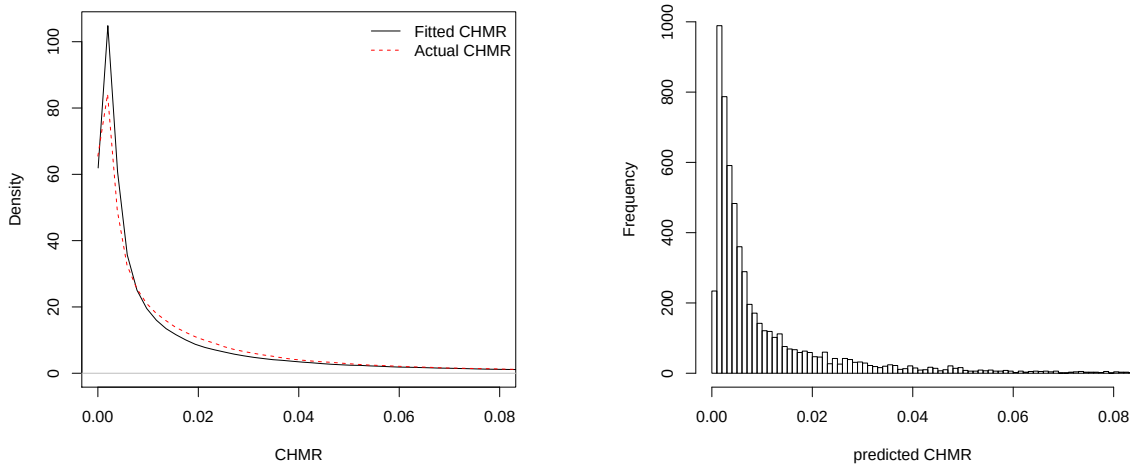
$$\begin{aligned}\log(\frac{\text{CHMR}}{1-\text{CHMR}}) = \beta_0 &+ \beta_1 \text{LAPS2} + \beta_2 \text{COPS2} + \beta_3 \text{Age} + \beta_4 \text{Sex} \\ &+ \beta_5 \text{AdmitCategory} + \beta_6 \text{DiseaseGroup} + \beta_7 \text{HospitalID}\end{aligned} \tag{4.1}$$

Table 4.2 presents the estimated coefficients in Eq. 4.1. For conciseness of presentation, we do not show the coefficients associated with 21 hospital IDs and 38 disease groups. Figure 4.1(a) compares the actual vs. fitted CHMR score in the estimation sample. Overall, we consider it to be a reasonable fit of the CHMR score, with slight under-prediction on the number of patients with CHMR score near 0 due to the distribution of the logit-transformed CHMR score being slightly skewed to the left. Therefore, we use the estimated coefficients to predict the CHMR score for the 6,113 hospitalizations transported in, and Figure 4.1(b) shows the histogram of their predicted CHMR scores.

Table 4.2: Regression table for CHMR, N=306,176, $R^2 = 0.86$

	Estimate	Std. Err.	t-value	p-val
LAPS2	0.03	0.00	582.28	0.000
COPS2	0.01	0.00	327.47	0.000
Age	0.02	0.00	249.30	0.000
Sex	-0.14	0.00	-52.98	0.000
Non-ED, Surgical	-0.03	0.01	-5.14	0.000
ED, Medical	-0.05	0.01	-8.19	0.000
Non-ED, Medical	0.45	0.01	61.09	0.000

Figure 4.1: Compare actual, fitted, and predicted CHMR scores



(a) Fitted vs. actual CHMR scores

(b) Histogram of predicted CHMR scores

SAPS3 for ICU visits

In total, we have missing SAPS3 scores for 36 ICU visits of 13 patient episodes. There are two cases: (1) for six patient episodes each with over 10 ICU visits, the raw data only records their SAPS3 scores for the first 10 ICU visits of each patient episode; (2) seven hospitalizations have completely missing SAPS3 scores although the total number of ICU visits during that hospitalization is no more than 10.

Note that SAPS3 scores vary between 8 and 112 in our data with skew=0.54 and kurtosis=3.57. A simple linear regression of crude SAPS3 scores on all other risk covariates shows that the distribution of SAPS3 has a heavy tail (see Figure C.1 in Appendix C.1). Therefore, we run a linear regression of the log-transformed SAPS3 on all other severity scores and patient demographics, as

is specified in Eq. 4.2.

$$\begin{aligned}\log(\text{SAPS3}) = \beta_0 + \beta_1 \text{LAPS2} + \beta_2 \text{COPS2} + \beta_3 \text{Age} + \beta_4 \text{Sex} + \beta_5 \text{TransportIn} \\ + \beta_6 \text{AdmitCategory} + \beta_7 \text{DiseaseGroup} + \beta_8 \text{HospitalID}\end{aligned} \tag{4.2}$$

Table 4.3 presents the estimated coefficients in Eq. 4.2. For conciseness of presentation, again, we do not show the coefficients associated with 21 hospital IDs and 38 disease groups. Figure 4.2(a) compares the actual vs. fitted SAPS3 score in the estimation sample. The log-normal transformation still cannot completely remove the heavy tail of SAPS3 distribution. However, since this only affects 36 ICU visits of 13 hospitalizations, it is not likely to impact the estimation of structural parameters in Section 4.3. To focus our attention on the modeling part, for the moment, we use the predicted SAPS3 scores from Eq. 4.2 to approximate the 13 hospitalizations with missing SAPS3 scores. Figure 4.2(b) shows the histogram of their predicted SAPS3 scores for each ICU visit.

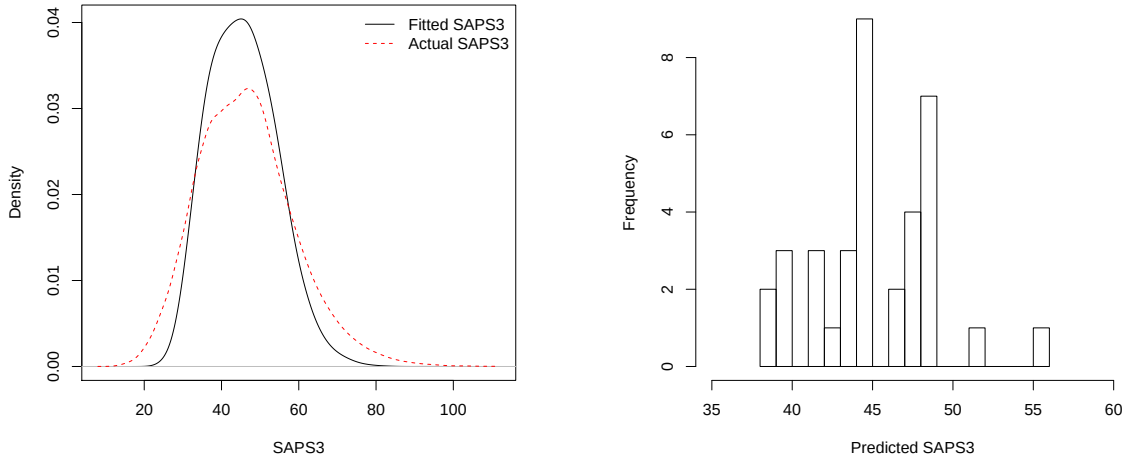
Table 4.3: Regression table for log(SAPS3), N=59,125, $R^2 = 0.50$

	Estimate	Std. Error	t value	P-val
LAPS2	0.002	0.000	91.13	0.000
COPS2	0.001	0.000	34.02	0.000
Age	0.007	0.000	125.58	0.000
Sex	0.007	0.002	4.41	0.000
Non-ED, Surgical	-0.140	0.004	-39.49	0.000
ED, Medical	0.024	0.003	8.57	0.000
Non-ED, Medical	0.040	0.004	9.69	0.000
Transport in	0.194	0.005	42.92	0.000

EDIP2 for general ward/TCU stays

Among the 312,289 hospitalizations with known sex, 297,492 hospitalizations have ever been to the general ward/TCU. For 295,553 hospitalizations, there is at least one EDIP2 score for each patient episode. 1,939 hospitalizations from all 21 hospitals do not have any EDIP2 scores throughout the entire hospitalization. There are two reasons for this observation: 1) 42 (2.17%) patients are discharged from the ward quickly after admission before the next EDIP2 update time; 2) 1,897 (97.83%) hospitalizations have inconsistent records of the inpatient units they stay in.

Figure 4.2: Compare actual, fitted, and predicted SAPS3 scores



(a) Fitted vs. actual SAPS3 scores

(b) Histogram of predicted SAPS3 scores

Patients without EDIP2 scores are more severe at admission, stayed in hospital for significantly shorter amount of time, and are more likely to die in hospital. See Table C.2, C.3, C.4, and Figure C.2 in Appendix C.1 for detailed comparisons of the hospitalizations with and without EDIP2 scores.

Recall that the EDIP2 score is updated every six hours while patients are staying in the general ward/TCU. Similar to the model on CHMR, we impute the EDIP2 score for the 1,939 hospitalizations by regressing the logit-transformed *first* EDIP2 score for each ward/TCU stay on other severity scores and patient demographics, as is specified in Eq. 4.3.

$$\begin{aligned}\log\left(\frac{\text{1st EDIP2}}{1-\text{1st EDIP2}}\right) = \beta_0 + \beta_1\text{LAPS2} + \beta_2\text{COPS2} + \beta_3\text{Age} + \beta_4\text{Sex} + \beta_5\text{TransportIn} \\ + \beta_6\text{AdmitCategory} + \beta_7\text{DiseaseGroup} + \beta_8\text{HospitalID}\end{aligned} \tag{4.3}$$

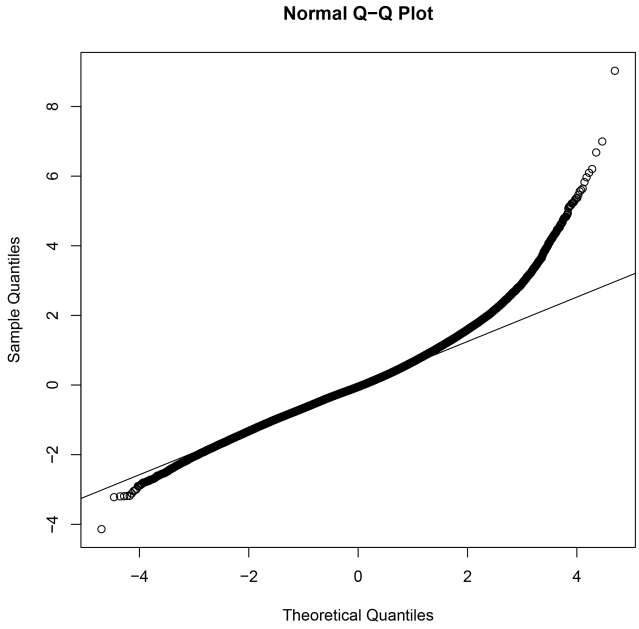
Table 4.4 presents the estimated coefficients and Figure 4.3 shows the QQ plot of the residuals from the regression. Similar to SAPS3, the distribution of EDIP2 scores has a heavy tail toward the higher scores, as is reflected from the QQ plot.

We also attempt to fine tune the estimation model in subgroups of patients with short LOS, divide patients by severity of illness, and pre-process the data with propensity score matching.

Table 4.4: Regression results for first EDIP2, N=373,548, $R^2 = 0.40$

	Estimate	Std. Error	t value	P-value
LAPS2	0.011	0.000	333.15	0.000
COPS2	0.001	0.000	34.44	0.000
Age	-0.003	0.000	-44.89	0.000
Sex	-0.117	0.002	-49.59	0.000
Non-ED, Surgical	-0.396	0.005	-75.76	0.000
ED, Medical	-0.009	0.005	-2.02	0.043
Non-ED, Medical	0.025	0.006	3.94	0.000
Transport in	0.102	0.009	11.23	0.000

Figure 4.3: QQ plot of residuals from regressing first EDIP2 on all other risk covariates



However, none of the above methods leads to significant increase in explanation power. Although it is difficult to accurately predict each individual patient's first EDIP2 score, ultimately we only care about the distribution of EDIP2 scores for all patients staying in each inpatient unit at a system snapshot. While each patient's individual EDIP2 score could vary extensively during the entire hospitalization, the distribution of EDIP2 scores on all patients in the ward/TCU has infinitesimal hourly change. Overall only 8% of the average EDIP2 scores change between consecutive hours and the average change is only 0.0003. As such, we decide to simply use the previous hour's average/min/max EDIP2 score in the unit as the current hour's if there are incoming patients missing

EDIP2 scores in the current period. As a robustness check, we also drop the 23% observations when there are patients missing EDIP2 scores in the general ward/TCU and our findings from the reduced-form regressions are similar.

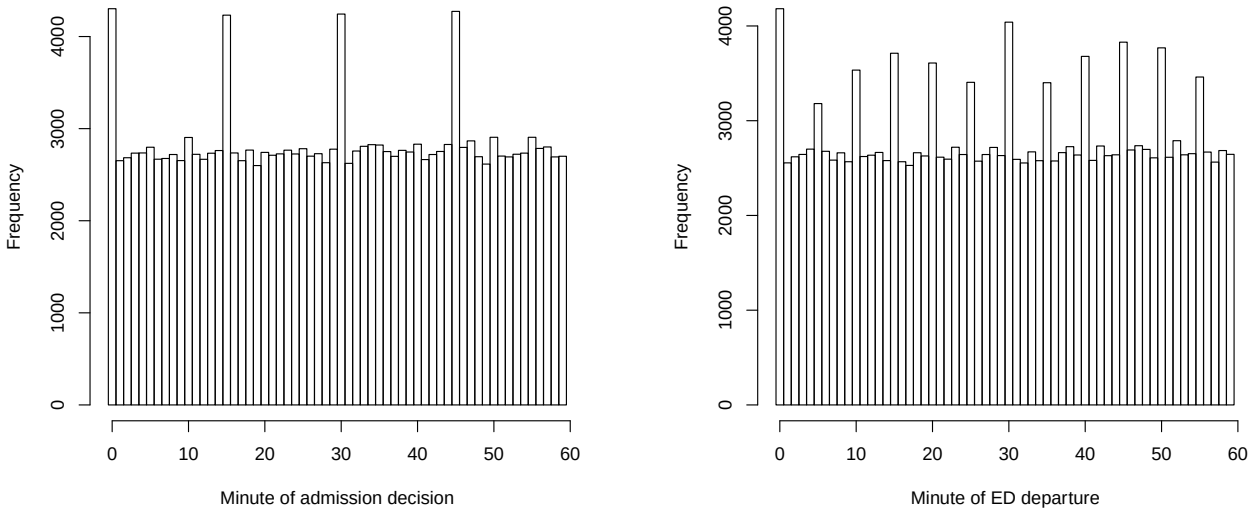
4.3. Empirical Models

Our goal is to estimate the intertemporal discount factor and the cost parameters physicians associate to their ED admission decisions. We first use reduced-form regressions on the individual patient level to understand qualitatively what patient and system factors are impacting the routing decisions of patients in the ED. Furthermore, we use rigorous econometric models to estimate the potential impact of ICU admission from the ED on patient outcomes. We then formally introduce the dynamic discrete choice structural model on the system level to quantify the structural parameters associated with physicians' admission decisions.

Recall that we focus on three choices ED physicians can make once they decide a patient needs to be hospitalized. They can admit the patient to the ICU, reroute the patient to the general ward/TCU, or let the patient wait in the ED for a downstream bed to be available. Since this is a dynamic decision made repeatedly over the time period when patients remain in the ED, we first define the time interval (or frequency) of the decision. We use two hours as the length of time between any two admission decisions, for two main reasons. First, we find a behavioral pattern on the admission and discharge time from the ED in our data, where most admission decisions are made every 15 minutes of the hour, whereas the final departure of patients from the ED to inpatient units happens mostly hourly, followed by every 30 minutes, as is shown in Figure 4.4. Hence, using one or two hours as the decision interval helps to smooth out the "seasonality" in the admission decision and departure time pattern. Table 4.5 shows that patient severity conditions are clinically similar when comparing the groups of patients who wait no more than one or two hours. Second, once an ED physician issues the admission request, it typically takes time to sort out the bed availability with various inpatient units and discuss the patient conditions and potential treatment plans with hospitalists. Hospitalists often need to come down to the ED to assess the patient prior

to finally transferring the patient to an inpatient unit. Therefore, we believe every two hours is a reasonable frequency for examining admission decisions from the ED.

Figure 4.4: Histogram for the time of ED admission decision and departure



(a) Minute-of-hour for ED admission decision

(b) Minute-of-hour for ED departure

Table 4.5: Patient characteristics between admitted immediately, wait $\leq$ 1hr, and wait $\leq$ 2hrs

	# of obs	LAPS2	COPS2	CHMR	Age	Flu season	Female
Wait$\leq$1hr	114,419	71.30	44.26	4.1%	67.05	26.5%	53.1%
Wait$\leq$2hr	157,668	71.87	44.51	4.2%	67.16	28.5%	53.2%
Do not wait	4,228	64.88	38.77	3.2%	64.75	28.5%	48.0%

4.3.1. Reduced Form Regressions

4.3.1.1. Factors Impacting the ICU Admission Decision

To model physicians' ICU admission decision from the ED, we first use two reduced form regressions to understand what patient risk factors and system controls are impacting the admission decision. Specifically, we use the logistic regression to model whether a patient is admitted to the ICU or not, and use the multinomial logit model on all three admission choices simultaneously. The unit

of analysis for both models is per patient per decision epoch.

Logistic regression

For each patient i in time period t, let $ICUAdmit_{i,t}$ be a binary variable indicating whether patient i is admitted to the ICU from the ED in time period t. Let $X_{i,t}$ be the system status and seasonality controls faced by patient i at time t, which include the census and aggregate patient characteristics in all inpatient units, hospital IDs, day-of-week, month, and nurse shifts. Let Y_i be the patient characteristics that are constant over t, including patient demographics (age, gender), severity scores (CHMR, COPS2, LAPS2) and 38 disease categories (see Table C.5 in Appendix C.2 for a list of the control variables used). Let $\epsilon_{i,t}$ be an error following the standard logistic distribution. The logistic regression model is specified as follows.

$$ICUAdmit_{i,t} = \mathbb{1}\{ICUAdmit^*_{i,t} > 0\} \quad \text{where} \quad ICUAdmit^*_{i,t} = X_{i,t}^\top \beta + Y_i^\top \alpha + \epsilon_{i,t} \tag{4.4}$$

Multinomial logit regression

The multinomial logit model allows us to examine all three admission choices (denoted as option j, $j = 0, 1, 2$) at the same time: reroute to the ward/TCU, admit to the ICU, or continue waiting in the ED. Using similar notation as in Eq. 4.4, for each patient hospitalization i at system time t, the utility of choosing option j is

$$U_{(i,t),j} = X_{(i,t)}^\top \beta_j + Y_i^\top \alpha_j + \varepsilon_{(i,t),j},$$

where $\varepsilon_{(i,t),j}$ follows the standard Gumbel distribution. Normalizing at one of the three choices "reroute to the ward" ($j = 0$) for identification, the probability of choosing j for patient i at time t is

$$P_{(i,t)}(j) = \frac{e^{X_{(i,t)}^\top \beta_j + Y_i^\top \alpha_j}}{1 + \sum_{k=1}^{2} e^{X_{(i,t)}^\top \beta_k + Y_i^\top \alpha_k}}$$

The coefficients β_j and α_j are estimated via maximum likelihood estimation.

4.3.1.2. Impact of ICU Admission on Patient Outcomes

In Section 4.3.1.1, we attempt to understand what factors impact the ICU admission decision. Physicians implicitly weigh the cost of taking up ICU resources and the potential benefit of improving patient outcomes when making the admission decision. In this section, we present the model to estimate the potential benefits of ICU admission on three patient oucomes: 1) in-hospital mortality, 2) length-of-stay (LOS, measured in number of days) in hospital, and 3) hospital readmission within two weeks or one month. Table 4.6 summarizes the statistics for patient outcomes. Different from that in section 4.3.1.1, the unit of analysis here is per patient because there is only one realized admission decision and outcome per patient.

Table 4.6: Summary of patient outcomes

Outcome	N	Mean	Median	Std. Dev.
Mortality	170,405	4.05%		
LOS (days)	170,405	4.24	3	6.17
Re-admit (1m)	163,502	16.70%		
Re-admit (2w)	163,502	10.40%		

A naive approach to estimate the effect of ICU admission on patient outcomes is to regress patient outcome on $ICUAdmit_i$ and control variables in a way similar to Eq. 4.4. This approach ignores the fact that 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. To address this concern, we utilize an instrumental variable approach similar to that in Chapter 3 Section 3.3.3 and that in Kim *et al.* (2015).

Similar to Chapter 3 Section 3.3.3 and Kim *et al.* (2015), we use ICU occupancy as an instrumental variable (IV). Formally, we define an ICU to be "busy" ($ICUBusy = 1$) when the ICU occupancy is no less than the 95^{th} percentile of its occupancy distribution. We use logistic regression to verify that ICU occupancy level has a significant impact on ICU admission. We also find that observed patient severity scores are not correlated to ICU occupancy, and therefore ICU occupancy is unlikely to be related to unobservable risk measures. Similar to Chapter 3 Section 3.3.3, we include a covariate $AvgOccVisited_i$ that measures the average occupancy of every unit a patient

visits during their hospital stay. This approach is required 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. The average $AvgOccVisited_i$ was 0.78, with median of 0.79, in our dataset. We find that our instrumental variable, $ICUBusy$ before patient i's admission, is not highly correlated with the average occupancy experienced by a patient after admission, with a correlation coefficient of 0.13. We examine other cutoffs on the ICU occupancy as robustness tests in Section 4.4.1.2.

In addition to $ICUBusy$, we include 3 additional instrumental variables that were used as potential factors affecting ICU admission decisions in Kim *et al.* (2015). The first auxiliary IV, $RecentDischarge_i$, accounts for the number of all discharges from the ICU within three hours prior to patient i's inpatient admission decision. This auxiliary is relevant to the ICU admission pattern because ICU discharges may release the nurse in charge of the discharged patient, but intensivists may have an incentive to "preserve the nurse hours" by admitting new patients to the ICU. In our sample, 34% of the patients saw zero ICU discharge, 25% saw one ICU discharge, and 17% saw two ICU discharges within three hours prior to admission. We divide the number of ICU discharges by the ICU capacity to normalize the IV. The second auxiliary IV, $RecentAdmission_i$, accounts for the number of patients admitted to the ICU via the ED for medical service within two hours before patient i's inpatient admission decision. The rationale is that a high number of recent admissions from the ED may reduce the bargaining power of the ED physician to further request more ICU admissions. In our sample, 75% of the patients saw zero recent admission and 21% saw one recent admission. Again, we divide the number of recent admissions by the ICU capacity to normalize the IV. The third auxiliary IV, $PctSevere_i$, accounts for the percentage of ICU capacity occupied by the most severe patients with LAPS2 ≥ 152 (which is the 97th percentile of the LAPS2 distribution, typically with over 50% ICU admission rates). The idea is that when the current ICU occupants are sicker, intensivists might require an incoming patient to also be highly severe. In our sample, 11% of the patients saw zero highly severe patients, 18% saw one, 19% saw two, and 17% saw three highly severe ICU occupants. All three auxiliary IVs exhibit no correlation with the LAPS2 score of the incoming patient, suggesting that they are unrelated to patients' severity of illness and therefore appear to be exogenous.

Mortality/Readmission: Because $Mortality_i$ and $Readmit_i$ are binary outcomes, it is more efficient to model the joint determination of mortality (or readmisison) and the ICU admission decision by a bivariate probit model and use maximum likelihood estimation rather than two-stage-least-squares (Wooldridge, 2010).

Let Z_i be the set of 4 IVs, and X_i be patient, hospital, and seasonality controls that include patient demographics (age, gender), severity scores (EDIP2, CHMR, LAPS2), 38 disease categories, and other indicators for hospital, day of the week, and month (see Table C.6 in the Appendix for more details). We have that

$$ICUAdmit_i = \mathbb{1}\{ICUAdmit_i^* > 0\} \quad \text{where} \quad ICUAdmit_i^* = X_i^\top \beta_1 + Z_i^\top \beta_2 + \epsilon_i \tag{4.5}$$

$$Y_i = \mathbb{1}\{Y_i^* > 0\} \quad \text{where} \quad Y_i^* = X_i^\top \beta_3 + \beta_4 ICUAdmit_i + \beta_5 AvgOccVisited_i + \nu_i. \tag{4.6}$$

(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 $ICUAdmit_i$ is indeed endogenous).

LOS: LOS_i is a count variable of the number of nights a patient stays in the hospital. We use the negative binomial regression, which can model the over-dispersion of the LOS distribution (as is shown in Table 4.6), with Eq. 4.5 in the first stage to estimate the impact of ICU admission on LOS using the method in Deb *et al.* (2006). Our estimates include patients who do not survive to hospital discharge, but our results are robust to excluding them.

4.3.2. Structural Model

4.3.2.1. Model Formulation

In this section, we define a stylized structural model of patient admission decisions from the ED. We first classify patients in the ED into K_{ED} groups by their severity levels at the time of ED admission decision. Similarly, we classify patients in the ICU into K_{ICU} groups by their severity levels during the ICU care.

Let $A_i, i = 1, \ldots, K_{ED}$ be the number of class i patients requiring inpatient admission in the

ED during each time period. A_i follows a Poisson distribution with rate λ_i. While theoretically A_i could be unbounded, we truncate A_i at the maximum number of patients requiring admission per period (denoted as N_{A_i}) observed in our data. This approach helps to limit the state space of the model. Similarly, let $D_i, i = 1, \ldots, K_{ICU}$ be the number of class i patients departing from the ICU per time period. For each individual patient from class i, his/her service completion probability per period is $\mu_i, i = 1, \ldots, K_{ICU}$.

At each time period $t, t = 1, 2, \ldots$, an ED physician observes the hospital system state s_t and chooses an action d_t from a finite choice set $\Pi(s_t)$, resulting in a utility $U(s_t, d_t)$. We define the system state at the beginning of time period t as

$$s_t = \left[n_{1,t}^{ED}, \ldots, n_{K_{ED},t}^{ED}; n_{1,t}^{ICU}, \ldots, n_{K_{ICU},t}^{ICU}\right],$$

where $n_{k,t}^{ED}$ is the number of class k patients in the ED at the beginning of time period t, and vice versa for $N_{k,t}^{ICU}$. Again, to limit the possible state space and align with the real hospital practice, we consider an ICU with N_{ICU} beds, and an ED with the maximum boarding space of N_{ED}. A capacitated ED bears similarity to the ambulance diversion when the ED is full. To focus our attention on the ICU admission decision, we consider a ward with ample capacity in the hospital. Therefore, the entire state space at the beginning of time period t is

$$\mathcal{S}_t = \left\{ \left[n_{1,t}^{ED}, \ldots, n_{K_{ED},t}^{ED}; n_{1,t}^{ICU}, \ldots, n_{K_{ICU},t}^{ICU}\right] : \sum_{i=1}^{K_{ED}} n_{i,t}^{ED} \leq N_{ED} + \sum_{i=1}^{K_{ED}} N_{A_i}, \sum_{i=1}^{K_{ICU}} n_{i,t}^{ICU} \leq N_{ICU} \right\}$$

For simplicity of presentation, we index all possible states as $1, 2, \ldots, N_s$.

We define the action taken at a given state at time period t as

$$d_t = \left[Admit_{1,t}, Ward_{1,t}; \ldots; Admit_{K_{ED},t}, Ward_{K_{ED},t}\right],$$

where $Admit_{k,t}$ indicates the number of ICU admissions for class k patients from the ED, and vice versa for $Ward_{k,t}$. Because the space of waiting in the ED is limited to N_{ED}, instead of forcing ambulance diversion, which is rarely the case in KPNC, we reroute patients to the ward

immediately to relieve the congestion in the ED. Therefore, the entire choice set given the system state s_t at time period t is

$$\Pi(s_t) = \Bigg\{ [Admit_{1,t}, Ward_{1,t}; \ldots; Admit_{K_{ED},t}, Ward_{K_{ED},t}] : \\ \sum_{i=1}^{K_{ED}} Admit_{i,t} \leq N_{ICU} - \sum_{j=1}^{K_{ICU}} n_{j,t}^{ICU}, \\ Admit_{i,t} + Ward_{i,t} \leq n_{i,t}^{ED}, \quad i = 1, \ldots, K_{ED} \\ \sum_{i=1}^{K_{ED}} Admit_{i,t} + Ward_{i,t} \geq \left(\sum_{i=1}^{K_{ED}} n_{i,t}^{ED} - N_{ED} \right)^{+} \Bigg\},$$

where the first two constraints are feasibility constraints on how many patients can be admitted to the ICU or non-ICU units, and the third constraint is corresponding to the mandatory ward rerouting to relieve the ED congestion. For simplicity of presentation, we index all possible actions as $1, 2, \ldots, N_d$.

Finally, we aim to estimate the cost associated with admitting a patient from class k in the ED to the ICU (denoted as i_k), the general ward (denoted as ϕ_k), or waiting in the ED (denoted as h_k). To achieve identification, without loss of generality, we set the choice of ICU admission as the outside option and normalize i_k to zero for all k. We use $\beta \in (0, 1)$ to denote the intertemporal discount factor, which indicates how much physicians take into account the intertemporal externality in their decision-making. The ED physician chooses a sequence of decision rules to maximize the expected discounted utility over an infinite time horizon

$$\sup \mathsf{E} \left\{ \sum_{j=t}^{\infty} \beta^{j-t} U\left(s_j, d_j\right) \mid s_t \right\}. \tag{4.7}$$

Similar to Rust (1987), we make the following four main assumptions to simplify the model:

1. The utility function $U(s_t, d_t)$ is additively separable (i.e., $U(s_t, d_t) = -c(s_t, d_t) + \epsilon_t(d_t)$), where $c(s_t, d_t)$ is the one-period cost of taking action d_t on an observed system state s_t, and $\epsilon_t(d_t)$

is a random utility component observed by the ED physician when making choice d_t, but unobservable in our choice data. The one-period cost of taking action d_t on the observed state s_t is the one-time cost of rerouting plus the cost of waiting (if any)

$$c(s_t, d_t) = \sum_{i=1}^{K_{ED}} \phi_i Ward_{i,t} + h_i \left(n_{i,t}^{ED} - Admit_{i,t} - Ward_{i,t} \right)^+ .$$

2. The state transition density is Markovian, i.e.,

$$Pr\{s_{t+1}, \epsilon_{t+1} \mid s_t, \epsilon_t, d_t; s_{t-1}, \epsilon_{t-1}, d_{t-1}; \ldots\} = Pr\{s_{t+1}, \epsilon_{t+1} \mid s_t, \epsilon_t, d_t\}.$$

3. Conditional on s_t, ϵ_t is independent from the past observation history (i.e., ϵ_t is a noise superimposed on s_t)

$$Pr\{s_{t+1}, \epsilon_{t+1} | s_t, \epsilon_t, d_t\} = p(s_{t+1}|s_t, d_t) q(\epsilon_{t+1}|s_{t+1})$$

4. ϵ_t follows a multivariate extreme value distribution with mean θ and standard deviation σ

$$q(\epsilon_t \mid s_t) = \prod_{d \in \Pi(s_t)} \exp\{-\frac{\epsilon_t(d) - \theta}{\sigma}\} \exp\{-e^{-\frac{\epsilon_t(d)-\theta}{\sigma}}\}$$

Under the four assumptions above, Rust (1987) shows that the optimal solution to Eq. 4.7 is given by a stationary decision rule. The probability of choosing action d given the state s is given by the multinomial logit choice model

$$P(d \mid s) = \frac{\exp\{-c(s,d) + \beta \sum_{s'} p(s' \mid s, d) v(s')\}}{\sum_{j \in \Pi(s)} \exp\{-c(s,j) + \beta \sum_{s'} p(s' \mid s, j) v(s')\}} \tag{4.8}$$

where the expected value function $v(s)$ at any given state s is given by the fixed point to the contraction mapping

$$v(s) = \sum_{d \in \Pi(s)} P(d \mid s) \left[-c(s,d) + \beta \sum_{s'} p(s' \mid s, d) v(s') \right] \tag{4.9}$$

With the observed pairs of hospital state and action chosen, denoted as (s_t, d_t), we aim to back out the structural cost parameters associated with physicians' decision-making via finding the set of $\{\phi_k, h_k, \beta, k = 1, 2, \ldots, K_{ED}\}$ that maximizes the likelihood of choosing action d_t on a given state s_t over the entire observation history:

$$L(s_t, d_t, t = 1, 2, \ldots, T \mid s_0, d_0) = \prod_{t=1}^{T} P(d_t \mid s_t) p(s_t \mid s_{t-1}, d_{t-1}) \tag{4.10}$$

4.3.2.2. Estimation Procedures

We adopt the *nested fixed point algorithm* similar to that used in Rust (1987) to solve the above structural model, which is essentially a fixed point algorithm (aka, "inner" algorithm) nested in a maximum likelihood estimation (aka, "outer" algorithm). Specifically, given a set of structural cost parameters $\{\boldsymbol{\phi} = \{\phi_k\}, \boldsymbol{h} = \{h_k\}, k = 1, 2, \ldots, K_{ED}\}$ and the intertemporal discount factor β, the "inner" fixed point algorithm computes the unknown $v(s)$ as the fixed point to Eq. 4.9. Then, given $v(s)$, the "outer" algorithm searches for another set of cost parameters which increases the likelihood in Eq. 4.10. We iterate the "inner" and "outer" algorithm until it converges to a set of cost parameters that gives the maximum likelihood. For speed of convergence, we fix β in the algorithm to find the optimal set of cost parameters corresponding to the given β. We then do a grid search over β to find the set of cost parameters and β that maximize the likelihood. The estimation steps are illustrated in detail below. Fix a β, at the k^{th} iteration,

1. Given the set of unknown parameters $\boldsymbol{\alpha^k} = (\boldsymbol{\phi^k}, \mathbf{h^k}; \beta)$, calculate the corresponding expected value-to-go function $v^k(s)$ for each state s as the fixed point in the Bellman Equation 4.9 through value iteration with the convergence criteria being the maximum change in $v^k(s) < 0.0001$. Such convergence criteria strike a good balance between the accuracy of the value-to-go function and the speed of computation. From $v^k(s)$, we can calculate the probability of taking an action d on the state s, denoted as $P^k(d \mid s)$, using Eq. 4.8. For notation simplicity,

let $A^k(s,d) = -c^k(s,d) + \beta \sum_{s'} p\left(s' \mid s,d\right) v^k(s')$, and therefore

$$\begin{aligned} v^k(s) &= \sum_{d\in\Pi(s)} P^k(d \mid s) \left[-c^k(s,d) + \beta \sum_{s'} p(s' \mid s,d) v^k(s')\right] \\ &= \sum_{d\in\Pi(s)} \frac{e^{-c^k(s,d)+\beta\sum_{s'} p(s'|s,d)v^k(s')}}{\sum_{j\in\Pi(s)} e^{-c^k(s,j)+\beta\sum_{s'} p(s'|s,j)v^k(s')}} \left[-c^k(s,d) + \beta \sum_{s'} p(s' \mid s,d) v^k(s')\right] \\ &= \sum_{d\in\Pi(s)} \frac{e^{A^k(s,d)}}{\sum_{j\in\Pi(s)} e^{A^k(s,j)}} A^k(s,d) \end{aligned}$$

2. Calculate the likelihood L^k of the observation history $(s_t, d_t), t = 1, \ldots, T$. Specifically, because the $p(s_t \mid s_{t-1}, d_{t-1})$ is exogenously determined by the arrival rates to the ED and departure probabilities from the ICU, but is unrelated to the unknown parameters $\boldsymbol{\alpha^k} = (\boldsymbol{\phi^k}, \mathbf{h^k}; \beta)$, we simply need to calculate the partial log-likelihood l as

$$\begin{aligned} & l^k(s_1, \cdots, s_T, d_1, \cdots, d_T \mid \boldsymbol{\alpha^k}) \\ =& \sum_{t=1}^{T} \log P^k(d_t \mid s_t) \\ =& \sum_{t=1}^{T} \left[-c^k(s_t,d_t) + \beta \sum_{s'} p\left(s' \mid s_t, d_t\right) v^k(s') - \log \sum_{j\in\Pi(s_t)} e^{-c^k(s_t,j)+\beta\sum_{s'} p(s'|s_t,j)v^k(s')}\right] \\ =& \sum_{t=1}^{T} \left(A^k(s_t,d_t) - \log \sum_{j\in\Pi(s_t)} e^{A^k(s_t,j)}\right) \end{aligned} \tag{4.11}$$

3. Calculate the gradient of the partial log-likelihood over all structural cost parameters: $\frac{\partial l^k(\boldsymbol{\alpha^k})}{\partial \boldsymbol{\alpha}}$. Using $\frac{\partial l^k}{\partial \phi_1}$ as an example,

$$\begin{aligned} \frac{\partial l^k(\boldsymbol{\alpha^k})}{\partial \phi_1} &= \sum_{t=1}^{T} \frac{\partial A^k(s_t,d_t)}{\partial \phi_1} - \sum_{d\in\Pi(s_t)} P^k\left(d \mid s_t\right) \frac{\partial A^k(s_t,d)}{\partial \phi_1}, where \\ \frac{\partial A^k(s_t,d_t)}{\partial \phi_1} &= -\frac{\partial c^k(s_t,d_t)}{\partial \phi_1} + \beta \sum_{s'} p\left(s' \mid s_t, d_t\right) \frac{\partial v^k(s')}{\partial \phi_1} \end{aligned}$$

4. Use the BFGS and backtracking line search algorithm to find a descent direction γ^k and step length δ^k, and update $\boldsymbol{\alpha^{k+1}} = \boldsymbol{\alpha^k} + \delta^k\gamma^k$, then repeat from step 1 with $k = k + 1$. The BFGS algorithm (Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970) essentially approximates the Hessian matrix H of the partial log-likelihood $l(\boldsymbol{\alpha})$ by its gradient ∇l to reduce the computational burden of calculating the second derivatives of $l(\boldsymbol{\alpha})$. Then, the backtracking line search algorithm (Armijo, 1966) attempts to find a step length along the descent direction $-H * \nabla l$ to increase the log-likelihood.

The above iteration process stops when $l^k(\boldsymbol{\alpha^k}) - l^{k-1}(\boldsymbol{\alpha^{k-1}}) < 10^{-6}$. Again, we use this stopping criteria to strike a balance between the accuracy of the results and the computational speed.

4.4. Empirical Results

In this section, we present and discuss our main empirical results for reduced form regressions. The structural estimation is still an ongoing process and further investigation is needed to fine tune the details. Therefore, we will not discuss the empirical results from the structural model in this chapter.

4.4.1. Reduced Form Regressions

4.4.1.1. Factors Impacting the ICU Admission Decision

We use the logistic and multinomial logit regressions to obtain a preliminary understanding on what and how patient characteristics and hospital system status measures may impact the ICU admission decision in the ED. Table C.7 in Appendix C.2 presents the estimated coefficients from the logistic regression on the ICU admission decision. Table C.8 in Appendix C.2 shows the estimated coefficients from the multinomial logit model on all three decision choices (normalizing at the choice "reroute to the ward"). Overall, the severity levels of the patients under consideration, the ICU and ward status, and the seasonality factors all impact physicians' ICU admission decisions. The higher the severity of the patients under consideration, the more likely they are going to be admitted to the ICU rather than rerouted to the ward.

When examining the impact of the ICU status on admission choices, we find that a higher ICU occupancy level or an ICU with more severe patients (measured by the average LAPS2 and CHMR scores) at the time of the admission decision discourages patients to be admitted to the ICU, but encourages them to continue waiting in the ED. The more recent discharges from the ICU, which frees up the ICU space, the more likely patients would be admitted to the ICU from the ED. The more severe patients recently admitted to the ICU, the less likely new patients would be admitted to the ICU from the ED.

In terms of the ward status, the ward census does not seem to discourage rerouting patients to the general ward/TCU. This observation may be due to the fact that the ICU census is highly correlated with the ward census with a correlation coefficient of 0.8. Therefore, when both the ward and the ICU are congested, more patients are being rerouted to the ward because the ICU is more capacity constrained. On the other hand, the average EDIP2 score of the patients in the ward has a high impact in discouraging ward admission.

Regarding the number of patients waiting for admission in the ED, the more patients boarding, and the higher average severity of patients boarding, the more likely such patients would be waiting in the ED for a decision, which may be because more time would be needed to sort out the right allocation of patients to inpatient resources.

Finally, the operating room status does not have a significant impact on the ICU admission decision on patients in the ED, probably because ICU beds are typically already reserved for the patients finishing the surgery.

4.4.1.2. Impact of ICU Admission on Patient Outcomes

We estimate the impact of ICU admission on patient outcomes in two samples: 1) the entire patient population, and 2) a subset of more severe patients with LAPS2 $\geq$ 110, which is the 85th percentile of the LAPS2 distribution. The reason to examine a subsample of more severe patients is that low-severity patients typically would not be considered for ICU admission regardless of the ICU congestion level, but the IV approach applies fundamentally to patients whose admission decision would change with the ICU congestion level. Patients with LAPS2 score above 110 have an ICU

admission rate of 30.5%, which is significantly higher than the ICU admission rate of 11.7% in the entire patient population. Table 4.7 compares the mean patient characteristics and outcome measures between the entire population and the subsample with LAPS2 $\geq$ 110.

Table 4.7: Mean patient characteristics and outcome measures in entire population and subsample with LAPS2 $\geq$ 110

	Entire population	LAPS2 $\geq$ 110 subsample
N	170,405	26,362
Female (%)	53.25	50.16
CHMR (%)	4.25	16.17
LAPS2	71.99	135.12
Age	67.13	74.74
Mortality (%)	4.05	15.62
LOS (days)	4.24	6.01
Readmit (1m) (%)	16.70	22.59
Readmit (2w) (%)	10.40	14.57

N for the entire population is smaller than the sample size in Table 4.1 due to the drop of patients staying over the stable system occupancy time period

Tables 4.8 and 4.9 summarize estimation results for both the entire patient population and the subsample with LAPS2 $\geq$ 110.

Table 4.8: Effect of ICU admission on patient outcomes, all patients

	With IV[1]			Without IV
Outcome	Coef. (Std. Err.)	Avg. change	Pct. change	Coef. (Std. Err.)
Mortality	-0.193** (0.061)	-0.012	-38.0%	0.210*** (0.018)
LOS (days)	-0.337*** (0.007)	-1.037	-28.6%	0.262*** (0.007)
Readmit (1m)	-0.128* (0.058)	-0.030	-17.9%	-0.006 (0.013)
Readmit (2w)	-0.050 (0.066)	-0.009	-8.5%	0.008 (0.014)

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[1] First-stage F-statistic for the IVs = 97.3

First of all, we find evidence that the ICU admission decision is indeed endogenous, and hence an IV estimation approach is necessary to obtain unbiased inferences. Both Tables 4.8 and 4.9 show that without IV, ICU admission is estimated to result in higher in-hospital mortality and longer LOS in hospital. 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. Table C.11

Table 4.9: Effect of ICU admission on patient outcomes, LAPS2 $\geq$ 110

	With IV[1]			Without IV
Outcome	Coef. (Std. Err.)	Avg. change	Pct. change	Coef. (Std. Err.)
Mortality	-0.163 (0.115)	-0.017	-29.0%	0.158*** (0.025)
LOS (days)	-0.347*** (0.017)	-0.990	-29.3%	0.292*** (0.012)
Readmit (1m)	-0.071 (0.165)	-0.019	-9.8%	-0.009 (0.024)
Readmit (2w)	-0.037 (0.167)	-0.007	-6.1%	-0.007 (0.026)

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[1] First-stage F-statistic for the IVs = 28.4

and C.12 in Appendix C.2 provide more detailed regression results without using an IV.

Second, we find that the set of IVs is a quite strong ensemble with the first-stage F-statistic equal to 97.3 in the entire population and 28.4 in the more severe subsample, both well over the rule-of-thumb of 10. Specifically, $ICUBusy$ and $RecentDischarge$ have a significant impact on the ICU admission decision in both the entire patient population and the more severe subsample. $PctSevere$ and $RecentAdmission$ mainly have a significant impact on the ICU admission decision in the more severe subsample. When the ICU is busy, the likelihood of being admitted to the ICU decreases by 40% on average in the entire population and 32% in the more severe subsample. Table C.9 and C.10 in Appendix C.2 provide more detailed regression results using the IVs.

Third, we find that ICU admission has a significant impact on reducing in-hospital mortality, LOS, and readmission rates within one month in the entire population. For instance, admitting a patient to the ICU could reduce in-hospital mortality on average by 38%, LOS by 1 day, and readmission within one month by 18% in the entire population. Interestingly, in the more severe patient subsample, we only find significant impact of ICU admission on LOS, and its marginal impact (-1 day) is essentially the same as that in the entire population. We do not find a significant impact of ICU admission on in-hospital mortality, nor on hospital readmission.

One possible explanation may be that for such a group of patients with high severity of illness, their ICU admission decision is less likely to be affected by the ICU congestion level, but more based on their medical needs (recall that the first-stage F-statistic and the marginal impact of the IV in the more severe subsample is lower than that in the entire population). If ICU care could substantially reduce the mortality and readmission risks of such patients, they might be admitted

anyway even when the ICU is highly congested. In this case, our IV estimation approach cannot capture the potential benefits of ICU admission on such "non-complying" highly severe patients whose admission decisions are less likely to be affected by the instrumental variable.

In defining the binary $ICUBusy$ from the continuous ICU occupancy levels, we use the 95^{th} percentile of the ICU occupancy distribution for each hospital as the threshold for "busy". As a robustness check, we also attempted different thresholds varying between 90th and 97th percentiles. The estimation results are similar with only slight changes in the coefficient estimates.

4.5. Conclusion and Future Work

The inpatient admission decision from the ED exhibits a wide range of practice and remains a personalized clinical decision for ED physicians. In this work, we empirically estimate whether and how physicians account for the severity of patients and the system status in making ICU admission decisions from the ED. We further estimate the potential benefits of ICU admission on patient outcomes. We find that the sicker the patient, the less congested the ICU, the fewer severe patients in the ICU, or the more recent ICU discharges, all contribute to increasing the likelihood of ICU admission from the ED. We show that being admitted to the ICU from the ED could significantly reduce patients' in-hospital mortality, LOS, and readmission rates within one month.

Next, we propose a dynamic discrete choice structural model which allows us to empirically estimate the extent to which physicians account for the intertemporal externalities of their admission decisions and the cost parameters associated with their decision choices. Note that the structural estimation is still an ongoing process and further work is needed to finalize the results. Therefore, we defer the discussion on structural estimation results to a future paper.

There are several limitations to our work. First, we do not consider elective admissions to the ICU in modeling the transition of system states. Because the admission rates to the ICU for all other patients vary between 0.09 and 0.58 per period (two hours), therefore, with 98% chance there would be no more than two other patients admitted to the ICU per period. It would be similar to reducing the ICU size, but our findings should be similar. Second, we do not include the possibility

that patients admitted to the general ward from the ED may in the future require unplanned ICU transfers due to deterioration. It is however difficult to distinguish between patients who should have been admitted to the ICU but are misplaced to the ward, and the patients who are correctly admitted to the general ward, but develop complications later in their hospitalization, the former being avoidable but not the latter. Third, our data does not include patient code status. The estimated effect of ICU admission on patient outcomes may be overestimated for patients who are not full code. Fourth, our instrumental variable approach essentially estimates the impact of ICU admission on patients whose admission decision depends on ICU occupancy. This process excludes the set of patients who are very mildly sick (and hence never admitted to the ICU) and those who are extremely sick (and hence always admitted to the ICU).

For future work, it would be helpful to conduct counterfactual analyses to estimate how the ED admission practice and patient outcomes would change by modifying the estimated cost parameters and intertemporal discount factors estimated through the structural model. Because of the computational burden, we are confining the system states to census and classifying patients by their LAPS2 scores. It would be interesting for future work to also include the average patient severity characteristics into the system states.

Part III

Bibliography

Bibliography

Jaap H Abbring and Øystein Daljord. Identifying the discount factor in dynamic discrete choice models. *working paper*, 2016.

Alan Agresti. *Categorical data analysis*. Wiley, Hoboken, New Jersey, 3 edition, 2012.

Victor Aguirregabiria and Pedro Mira. Dynamic discrete choice structural models: A survey. *Journal of Econometrics*, 156(1):38 – 67, 2010.

Gad Allon, Awi Federgruen, and Margaret Pierson. How much is a reduction of your customers' wait worth? an empirical study of the fast-food drive-thru industry based on structural estimation methods. *Manufacturing & Service Operations Management*, 13(4):489–507, 2011.

Gad Allon, Sarang Deo, and Wuqin Lin. The impact of size and occupancy of hospital on the extent of ambulance diversion: Theory and evidence. *Operations Research*, 61(3):544–562, 2013.

Eitan Altman, Tania Jiménez, and Ger Koole. On optimal call admission control in resource-sharing system. *IEEE Transactions on Communications*, 49(9):1659–1668, 2001.

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

Larry Armijo. Minimization of functions having lipschitz continuous first partial derivatives. *Pacific J. Math.*, 16(1):1–3, 1966.

Australian Commission on Safety and Quality in Health Care. Australian atlas of healthcare variation. Report, National Health Performance Authority, 2015.

BIBLIOGRAPHY

Mike Baiocchi, Dylan S. Small, Scott Lorch, and Paul R. Rosenbaum. Building a stronger instrument in an observational study of perinatal care for premature infants. *Journal of the American Statistical Association*, 105(492):1285–1296, 2010.

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

Mitchell J. Barnett, Peter J. Kaboli, Carl A. Sirio, and Gary E. Rosenthal. Day of the week of intensive care admission and patient outcomes: A multisite regional evaluation. *Medical Care*, 40(6):530–539, 2002.

Eta S. Berner, Don E. Detmer, and Donald Simborg. Will the wave finally break? a brief view of the adoption of electronic medical records in the united states. *Journal of the American Medical Informatics Association*, 12(1):3–7, 2005.

John Bound, David A. Jaeger, and Regina M. Baker. 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, 1995.

Theologos Bountourelis, David Eckman, Louis Luangkesorn, Andrew Schaefer, Spencer G. Nabors, and Gilles Clermont. Sensitivity analysis of an ICU simulation model. In *Proceedings of the 2012 Winter Simulation Conference (WSC)*, pages 1–12, Dec 2012.

Robert L Bray and Haim Mendelson. Production smoothing and the bullwhip effect. *Manufacturing & Service Operations Management*, 17(2):208–220, 2015.

Robert L Bray. Strong convergence and dynamic economic models. *Quantitative Economics to appear*, pages 1–28, 2018.

Robert Louis Bray. Markov decision processes with exogenous variables. *Management Science to appear*, pages 1–28, 2018.

Guy N. Brock, Christopher Barnes, Julio A. Ramirez, and John Myers. How to handle mortality

when investigating length of hospital stay and time to clinical stability. *BMC Medical Research Methodology*, 11(1):144, 2011.

Charles George Broyden. The convergence of a class of double-rank minimization algorithms. *IMA Journal of Applied Mathematics*, 6(1):76–90, 1970.

Brad W. Butcher, Eric Vittinghoff, Judith Maselli, and Andrew D. Auerbach. 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, 2013.

Rodrigo Cavallazzi, Paul E. Marik, Amyn Hirani, Monvasi Pachinburavan, Tajender S. Vasu, and Benjamin E. Leiby. Association between time of admission to the ICU and mortality: A systematic review and metaanalysis. *Chest*, 138(1):68–75, 2010.

Donald B Chalfin, Stephen Trzeciak, Antonios Likourezos, Brigitte M Baumann, R Phillip Dellinger, DELAY-ED study group, et al. Impact of delayed transfer of critically ill patients from the emergency department to the intensive care unit. *Critical care medicine*, 35(6):1477–1483, 2007.

Carri W. Chan, Vivek F. Farias, Nicholas Bambos, and Gabriel J. Escobar. Optimizing intensive care unit discharge decisions with patient readmissions. *Operations Research*, 60(6):1323–1341, 2012.

Carri W. Chan, Vivek F. Farias, and Gabriel J. Escobar. The impact of delays on service times in the intensive care unit. *Management Science*, 63(7):2049–2072, 2017.

Carla A. Chrusch, Kendiss P. Olafson, Patricia M. McMillan, Daniel E. Roberts, and Perry R. Gray. High occupancy increases the risk of early death or readmission after transfer from intensive care. *Critical Care Medicine*, 37(10):2753–2758, 2009.

Ahmet Colak and Robert Bray. Why do automakers initiate recalls? a structural econometric game. *working paper*, 2016.

BIBLIOGRAPHY

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

Partha Deb, Pravin K Trivedi, et al. Maximum simulated likelihood estimation of a negative binomial regression model with multinomial endogenous treatment. *Stata Journal*, 6(2):246–255, 2006.

Elizabeth R. DeLong, David M. DeLong, and Daniel L. Clarke-Pearson. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*, 44(3):837–845, 1988.

Department of Health. The nhs atlas of variation in healthcare. Report, Public Health England, September 2015.

Kathy D. Duncan, Christine McMullan, and Barbara M. Mills. Early warning systems: The next level of rapid response. *Nursing*, 42(2):38–44, 2012.

Zvi Eckstein and Kenneth I. Wolpin. The specification and estimation of dynamic stochastic discrete choice models: A survey. *The Journal of Human Resources*, 24(4):562–598, 1989.

Gabriel J. Escobar, John D. Greene, Peter Scheirer, Marla N. Gardner, David Draper, and Patricia Kipnis. Risk-adjusting hospital inpatient mortality using automated inpatient, outpatient, and laboratory databases. *Medical Care*, 46(3):232–239, 2008.

Gabriel J. Escobar, Juan Carlos LaGuardia, Benjamin J. Turk, Arona Ragins, Patricia Kipnis, and David Draper. 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, 2012.

Gabriel J. Escobar, Marla N. Gardner, John D. Greene, David Draper, and Patricia Kipnis. Risk-adjusting hospital mortality using a comprehensive electronic record in an integrated health care delivery system. *Medical Care*, 51(5):446–453, 2013.

BIBLIOGRAPHY

Gabriel J Escobar, Vincent Liu, Yan S Kim, Benjamin J Turk, Arona Ragins, Brian A Dummett, Carnen L Adams, Elizabeth A Scruth, and Patricia Kipnis. Early detection of impending deterioration outside the ICU: a difference-in-differences (DiD) study. In *D104. CRITICAL CARE: PRE-ICU, RAPID RESPONSE, AND INITIAL TREATMENT FOR SEPSIS*, pages A7614–A7614. Am Thoracic Soc, 2016.

Antoine; Eskander, Jonathan C.; Irish, David R.; Urbach, and David P.; Goldstein. Head and neck cancer surgery in ontario, 2003–2010: An ices atlas. Technical report, Institute for Clinical Evaluative Sciences, Toronto, Ontario, 2015.

R Scott Evans, Kathryn G Kuttler, Kathy J Simpson, Stephen Howe1, Peter F Crossno, Kyle V Johnson, Misty N Schreiner, James F Lloyd, William H Tettelbach, Roger K Keddington, Alden Tanner, Chelbi Wilde, and Terry P Clemmer. Automated detection of physiologic deterioration in hospitalized patients. *Journal of the American Medical Informatics Association*, 22(2):350–360, 2015.

Peter D. Faris, William A. Ghali, and Rollin Brant. Bias in estimates of confidence intervals for health outcome report cards. *Journal of Clinical Epidemiology*, 56(6):553–558, 2003.

Joseph L Fleiss, Bruce Levin, and Myunghee Cho Paik. *Statistical methods for rates and proportions*. John Wiley & Sons, Hoboken, New Jersey, 2013.

Roger Fletcher. A new approach to variable metric algorithms. *The computer journal*, 13(3):317–322, 1970.

Gregg C Fonarow, Eric E Smith, Mathew J Reeves, Wenqin Pan, DaiWai Olson, Adrian F Hernandez, Eric D Peterson, and Lee H Schwamm. Hospital-level variation in mortality and rehospitalization for medicare beneficiaries with acute ischemic stroke. *Stroke*, 42(1):159–166, 2011.

Donald E Fry, Michael Pine, Harmon S Jordan, Anne Elixhauser, David C Hoaglin, Barbara Jones, David Warner, and Roger Meimban. Combining administrative and clinical data to stratify surgical risk. *Annals of surgery*, 246(5):875–885, 2007.

Noah Gans, Haipeng Shen, Yong-Pin Zhou, Nikolay Korolev, Alan McCord, and Herbert Ristock. Parametric forecasting and stochastic programming models for call-center workforce scheduling. *Manufacturing & Service Operations Management*, 17(4):571–588, 2015.

Haiyan Gao, Ann McDonnell, David A.Harrison, Tracey Moore, Sheila Adam, Kathleen Daly, Lisa Esmonde, David R. Goldhill, Gareth J. Parry, Arash Rashidian, Christian P. Subbe, and Sheila Harvey. 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, 2007.

Laurent G. Glance, Andrew Dick, Turner M. Osler, Yue Li, and Dana B. Mukamel. Impact of changing the statistical methodology on hospital and surgeon ranking: The case of the new york state cardiac surgery report card. *Medical Care*, 44(4):311–319, 2006.

Laurent G. Glance, Andrew W. Dick, Turner M. Osler, and Dana B. Mukamel. Accuracy of hospital report cards based on administrative data. *Health Services Research*, 41(4p1):1413–1437, 2006.

Donald Goldfarb. A family of variable-metric methods derived by variational means. *Mathematics of computation*, 24(109):23–26, 1970.

Harvey Goldstein and David J. Spiegelhalter. League tables and their limitations: Statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 159(3):385–443, 1996.

Jesse Green and Neil Wintfeld. How accurate are hospital discharge data for evaluating effectiveness of care? *Medical Care*, 31(8), 1993.

Linda V. Green. How many hospital beds? *Inquiry*, 39(4):400–412, 2002.

Faheem W. Guirgis, Cynthia Gerdik, Robert L. Wears, Deborah J. Williams, Colleen J. Kalynych, Joseph Sabato, and Steven A. Godwin. Proactive rounding by the rapid response team reduces inpatient cardiac arrests. *Resuscitation*, 84(12):1668–1673, 2013.

J. L. Hodges and E. L. Lehmann. *Estimates of Location Based on Rank Tests*, pages 287–300. Springer US, Boston, MA, 2012.

V Joseph Hotz and Robert A Miller. Conditional choice probabilities and the estimation of dynamic models. *The Review of Economic Studies*, 60(3):497–529, 1993.

Wenqi Hu, Carri W. Chan, José R. Zubizarreta, and Gabriel J. Escobar. An examination of early transfers to the ICU based a physiological risk score. *Manufacturing & Service Operations Management, to appear*, 20(3):531–549, 2018.

Wenqi Hu, Carri W. Chan, José R. Zubizarreta, and Gabriel J. Escobar. Incorporating longitudinal comorbidity and acute physiology data in template matching for assessing hospital quality: an exploratory study in an integrated health care delivery system. *Medical Care*, 56(5):448–454, 2018.

L. I. Iezzoni, A. S. Ash, M. Shwartz, J. Daley, J. S. Hughes, and Y. D. Mackiernan. Judging hospitals by severity-adjusted mortality rates: the influence of the severity-adjustment method. *American Journal of Public Health*, 86(10):1379–1387, 1996.

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

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

Institute of Medicine. *Crossing the quality chasm: a new health system for the 21st century.* Washington, DC, 2001.

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

Ashish K. Jha, Zhonghe Li, E. John Orav, and Arnold M. Epstein. Care in U.S. hospitals—the hospital quality alliance program. *New England Journal of Medicine*, 353(3):265–274, 2005.

J. G. Jollis, M. Ancukiewicz, E. R. DeLong, D. B. Pryor, L. H. Muhlbaier, and D. B. Mark. Discordance of databases designed for claims payment versus clinical information systems: Implications for outcomes research. *Annals of Internal Medicine*, 119(8):844–850, 1993.

BIBLIOGRAPHY

Daryl A. Jones, Michael A. DeVita, and Rinaldo Bellomo. Rapid-response teams. *New England Journal of Medicine*, 365(2):139–146, 2011.

David Jones, Eric Schone, Frank Yoon, Alex Bohl, Sheng Wang, and Mariel Finucane. Improving the AHRQ quality indicators: Summary of findings and recommendations for improving the methodological approach. Technical report, Mathematica Policy Research, Cambridge, MA, 2014.

Diwas S. Kc and Christian Terwiesch. An econometric analysis of patient flows in the cardiac intensive care unit. *Manufacturing & Service Operations Management*, 14(1):50–65, 2012.

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

Song-Hee Kim, Carri W. Chan, Marcelo Olivares, and Gabriel Escobar. ICU admission control: An empirical study of capacity allocation and its implication for patient outcomes. *Management Science*, 61(1):19–38, 2015.

Patricia Kipnis, Benjamin J. Turk, David A. Wulf, Juan Carlos LaGuardia, Vincent Liu, Matthew M. Churpek, Santiago Romero-Brufau, and Gabriel J. Escobar. 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, 2016.

Marin H. Kollef, Yixin Chen, Kevin Heard, Gina N. LaRossa, Chenyang Lu, Nathan R. Martin, Nelda Martin, Scott T. Micek, and Thomas Bailey. A randomized trial of real-time automated clinical deterioration alerts sent to a rapid response team. *Journal of Hospital Medicine*, 9(7):424–429, 2014.

Harlan M. Krumholz, Yun Wang, Jennifer A. Mattera, Yongfei Wang, Lein Fang Han, Melvin J. Ingber, Sheila Roman, and Sharon-Lise T. Normand. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with heart failure. *Circulation*, 113:1693–1701, 2006.

William H Kruskal and W Allen Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260):583–621, 1952.

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

L. I. lezzoni. The risks of risk adjustment. *JAMA*, 278(19):1600–1607, 1997.

Jun Li, Nelson Granados, and Serguei Netessine. Are consumers strategic? structural estimation from the air-travel industry. *Management Science*, 60(9):2114–2137, 2014.

Winston Lin, Scott D Halpern, Meeta Prasad Kerlin, and Dylan S Small. A "placement of death" approach for studies of treatment effects on ICU length of stay. *Statistical Methods in Medical Research*, 26(1):292–311, 2017.

Peter K. Lindenauer, Denise Remus, Sheila Roman, Michael B. Rothberg, Evan M. Benjamin, Allen Ma, and Dale W. Bratzler. Public reporting and pay for performance in hospital quality improvement. *New England Journal of Medicine*, 356(5):486–496, 2007.

Peter K. Lindenauer, Susannah M. Bernheim, Jacqueline N. Grady, Zhenqiu Lin, Yun Wang, Yongfei Wang, Angela R. Merrill, Lein F. Han, Michael T. Rapp, Elizabeth E. Drye, Sharon-Lise T. Normand, and Harlan M. Krumholz. The performance of us hospitals as reflected in risk-standardized 30-day mortality and readmission rates for medicare beneficiaries with pneumonia. *Journal of Hospital Medicine*, 5(6):E12–E18, 2010.

Vincent Liu, Patricia Kipnis, Michael K. Gould, and Gabriel J. Escobar. Length of stay predictions: Improvements through the use of automated laboratory and comorbidity variables. *Medical Care*, 48(8):739–744, 2010.

Vincent Liu, Benjamin J Turk, Arona I Ragins, Patricia Kipnis, and Gabriel J Escobar. An electronic simplified acute physiology score-based risk adjustment score for critical illness in an integrated healthcare system. *Critical care medicine*, 41(1):41–48, 2013.

Maha Louriz, Khalid Abidi, Mostafa Akkaoui, Naoufel Madani, Kamal Chater, Jihane Belayachi, Tarek Dendane, Amine Ali Zeggwagh, and Redouane Abouqal. Determinants and outcomes associated with decisions to deny or to delay intensive care unit admission in morocco. *Intensive Care Medicine*, 38(5):830–837, May 2012.

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

Robin L Lumsdaine, James H Stock, and David A Wise. Three models of retirement: Computational complexity versus predictive validity. In *Topics in the Economics of Aging*, pages 21–60. University of Chicago Press, 1992.

Charles-Edouard Luyt, Alain Combes, Philippe Aegerter, Bertrand Guidet, Jean-Louis Trouillet, Claude Gibert, Jean Chastre, and for the CUB-Réa Network. Mortality among patients admitted to intensive care units during weekday day shifts compared with "off" hours. *Critical Care Medicine*, 35(1):3–11, 2007.

Thierry Magnac and David Thesmar. Identifying dynamic discrete decision processes. *Econometrica*, 70(2):801–816, 2002.

Lionel A. Mandell, Richard G. Wunderink, Antonio Anzueto, John G. Bartlett, G. Douglas Campbell, Nathan C. Dean, Scott F. Dowell, Thomas M. File, Jr., Daniel M. Musher, Michael S. Niederman, Antonio Torres, and Cynthia G. Whitney. Infectious diseases society of america/american thoracic society consensus guidelines on the management of community-acquired pneumonia in adults. *Clinical Infectious Diseases*, 44(Supplement_2):S27–S72, 2007.

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

Elizabeth A. McGlynn, Steven M. Asch, John Adams, Joan Keesey, Jennifer Hicks, Alison DeCristofaro, and Eve A. Kerr. The quality of health care delivered to adults in the united states. *New England Journal of Medicine*, 348(26):2635–2645, 2003.

Philipp G. H. Metnitz, Rui P. Moreno, Eduardo Almeida, Barbara Jordan, Peter Bauer, Ricardo Abizanda Campos, Gaetano Iapichino, David Edbrooke, Maurizia Capuzzo, and Jean-Roger Le Gall. Saps 3—from evaluation of the patient to evaluation of the intensive care unit. part 1: Objectives, methods and cohort description. *Intensive Care Medicine*, 31(10):1336–1344, Oct 2005.

Eric B. Milbrandt, Alexander Kersten, Malik T. Rahim, Tony T. Dremsizov, Gilles Clermont, Liesl M. Cooper, Derek C. Angus, and Walter T. Linde-Zwirble. Growth of intensive care unit resource use and its estimated cost in medicare. *Critical Care Medicine*, 36(9):2504–2510, 2008.

Mauro Moscucci, Kim A. Eagle, David Share, Dean Smith, Anthony C. De Franco, Michael O'Donnell, Eva Kline-Rogers, Sandeep M. Jani, and David L. Brown. Public reporting and case selection for percutaneous coronary interventions: An analysis from two large multicenter percutaneous coronary intervention databases. *Journal of the American College of Cardiology*, 45(11):1759–1765, 2005.

Peter M. Mullins, Munish Goyal, and Jesse M. Pines. 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, 2013.

National Center for Health Statistics. Health, united states, 2016: With chartbook on long-term trends in health. Technical report, U.S. Department of Health and Human Services, 2017.

OECD. *Geographic Variations in Health Care: What Do We Know and What Can Be Done to Improve Health System Performance?* OECD Publishing, Paris, 2014.

Marcelo Olivares, Christian Terwiesch, and Lydia Cassorla. Structural estimation of the newsvendor model: An application to reserving operating room time. *Management Science*, 54(1):41–55, 2008.

E. Lerzan Ormeci, Evrim Didem Gunes, and Derya Kunduzcu. A modeling framework for control of preventive services. *Manufacturing & Service Operations Management*, 18(2):227–244, 2016.

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

Rolla Edward Park, Robert H. Brook, Jacqueline Kosecoff, and et al. Explaining variations in hospital death rates: Randomness, severity of illness, quality of care. *JAMA*, 264(4):484–490, 1990.

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

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

M. Pine, M. Norusis, B. Jones, and G. E. Rosenthal. Predictions of hospital mortality rates: A comparison of data sources. *Annals of Internal Medicine*, 126(5):347–354, 1997.

M. Pine, H. S. Jordan, A. Elixhauser, and et al. Enhancement of claims data to improve risk adjustment of hospital mortality. *JAMA*, 297(1):71–76, 2007.

Jesse M. Pines, Ryan L. Mutter, and Mark S. Zocchi. Variation in emergency department admission rates across the united states. *Medical Care Research and Review*, 70(2):218–231, 2013.

Jacqueline M. Quail, Lisa M. Lix, Beliz Acan Osman, and Gary F. Teare. Comparing comorbidity measures for predicting mortality and hospitalization in three population-based cohorts. *BMC Health Services Research*, 11(1):146, 2011.

Hude Quan, Gerry A Parsons, and William A Ghali. Validity of information on comorbidity derived from icd-9-ccm administrative data. *Medical care*, 40(8):675–685, 2002.

Peter C Reiss and Frank A Wolak. Structural econometric modeling: Rationales and examples from industrial organization. *Handbook of econometrics*, 6:4277–4415, 2007.

René Robert, Jean Reignier, Caroline Tournoux-Facon, Thierry Boulain, Olivier Lesieur, Valérie Gissot, Vincent Souday, Mouldi Hamrouni, Cécile Chapon, and Jean-Paul Gouello. Refusal of

intensive care unit admission due to a full unit. *American Journal of Respiratory and Critical Care Medicine*, 185(10):1081–1087, 2012.

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

Paul R Rosenbaum and Jeffrey H Silber. Amplification of sensitivity analysis in matched observational studies. *Journal of the American Statistical Association*, 104(488):1398–1405, 2009.

Paul R. Rosenbaum. Sensitivity analysis for certain permutation inferences in matched observational studies. *Biometrika*, 74(1):13–26, 1987.

Paul R. Rosenbaum. *Observational Studies*. Springer-Verlag, 2 edition, 2002.

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

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

J. Rosenthal, C. Hanlon, K. Levit, and L. Holmquist. Enhancing clinical information in statewide hospital administrative data: Project summary and lessons learned. Report, U.S. Agency for Healthcare Research and Quality (AHRQ), 2010.

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

P Rui, K Kang, and Albert M. National hospital ambulatory medical care survey: 2013 emergency department summary tables. Technical report, National Center for Health Statistics, 2013.

John Rust. Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica*, 55(5):999–1033, 1987.

John P Rust. A dynamic programming model of retirement behavior. In *The economics of aging*, pages 359–404. University of Chicago Press, 1989.

John Rust. Chapter 51 structural estimation of markov decision processes. *Handbook of Econometrics*, 4:3081–3143, 1994.

Ian A Scott, Caroline A Brand, Grant E Phelps, Anna L Barker, and Peter A Cameron. Using hospital standardised mortality ratios to assess quality of care-proceed with extreme caution. *Med J Aust*, 194(12):645–648, 2011.

David M. Shahian and Sharon-Lise T. Normand. Comparison of "risk-adjusted" hospital outcomes. *Circulation*, 117(15):1955–1963, 2008.

David F Shanno. Conditioning of quasi-newton methods for function minimization. *Mathematics of computation*, 24(111):647–656, 1970.

Pengyi Shi, Mabel C. Chou, Jim G. Dai, Ding Ding, and Joe Sim. Models and insights for hospital inpatient operations: Time-dependent ed boarding time. *Management Science*, 62(1):1–28, 2016.

Amir Shmueli, CharlesL. Sprung, and EdwardH. Kaplan. Optimizing admissions to an intensive care unit. *Health Care Management Science*, 6(3):131–136, 2003.

Amir Shmueli, Mario Baras, and Charles L Sprung. The effect of intensive care on in-hospital survival. *Health Services and Outcomes Research Methodology*, 5(3-4):163–174, 2004.

Jonathan J. Shuster, Douglas W. Theriaque, and Brian M. Ilfeld. Applying hodges-lehmann scale parameter estimates to hospital discharge times. *Clinical trials (London, England)*, 5(6):631–634, 2008.

Jeffrey H Silber, Paul R Rosenbaum, Richard N Ross, Justin M Ludwig, Wei Wang, Bijan A Niknam, Nabanita Mukherjee, Philip A Saynisch, Orit Even-Shoshan, Rachel R Kelz, et al. Template matching for auditing hospital cost and quality. *Health services research*, 49(5):1446–1474, 2014.

Jeffrey H. Silber, Paul R. Rosenbaum, Richard N. Ross, Justin M. Ludwig, Wei Wang, Bijan A. Niknam, Philip A. Saynisch, Orit Even-Shoshan, Rachel R. Kelz, and Lee A. Fleisher. A hospital-

specific template for benchmarking its cost and quality. *Health Services Research*, 49(5):1475–1497, 2014.

Dylan S Small and Paul R. Rosenbaum. War and wages: The strength of instrumental variables and their sensitivity to unobserved biases. *Journal of the American Statistical Association*, 103(483):924–933, 2008.

Douglas Staiger and James H. Stock. Instrumental variables regression with weak instruments. *Econometrica*, 65(3):557–586, 1997.

C. Stenhouse, S. Coates, M. Tivey, P. Allsop, and T. Parker. 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, 2000.

Michael J Strauss, James P LoGerfo, James A Yeltatzie, Nancy Temkin, and Leonard D Hudson. Rationing of intensive care unit services: An everyday occurrence. *JAMA*, 255(9):1143–1146, 1986.

Elizabeth A. Stuart. Matching methods for causal inference: A review and a look forward. *Statistical Science*, 25(1):1–21, 2 2010.

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

Kenneth I Wolpin. An estimable dynamic stochastic model of fertility and child mortality. *Journal of Political economy*, 92(5):852–874, 1984.

Jeffrey M Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA, 2 edition, 2010.

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

Fan Yang, José R. Zubizarreta, Dylan S. Small, Scott Lorch, and Paul R. Rosenbaum. Dissonant conclusions when testing the validity of an instrumental variable. *The American Statistician*, 68(4):253–263, 2014.

Paul Young, Manoj Saxena, Rinaldo Bellomo, Ross Freebairn, Naomi Hammond, Frank van Haren, Mark Holliday, Seton Henderson, Diane Mackle, Colin McArthur, Shay McGuinness, John Myburgh, Mark Weatherall, Steve Webb, and Richard Beasley. Acetaminophen for fever in critically ill patients with suspected infection. *New England Journal of Medicine*, 373(23):2215–2224, 2015.

José R. Zubizarreta, Dylan S. Small, Neera K. Goyal, Scott A. Lorch, and Paul R. Rosenbaum. Stronger instruments via integer programming in an observational study of late preterm birth outcomes. *Annals of Applied Statistics*, 7:25–50, 2013.

José R; Zubizarreta, Çınar; Kılcıoğlu, and Juan Pablo; Vielma. designmatch: Construction of optimally matched samples for randomized experiments and observational studies that are balanced and representative by design. *CRAN, R package version 0.3.0*, 2016.

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

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

Part IV

Appendices

Appendix A

Appendix to Chapter 2

A.1. Supplemental Information on Covariate Balances

Table A.1: Covariate means when matched on administrative data only

Hospital ID	2	3	5	7	8	9	10	11	13	14	16	17	19	20	21	24	25	28	χ^2/df^*	P-val.
Age	67.7	68	67.8	67.7	67.8	67.8	67.7	67.8	67.8	68.1	67.7	67.7	67.7	67.8	67.8	67.7	67.8	67.6	0.01	1.00
Female (%)	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	0.00	1.00
Flu season (%)	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	0.00	1.00
ED surgical (%)	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	0.00	1.00
Non-ED surgical (%)	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	0.00	1.00
ED medical (%)	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	0.00	1.00
Non-ED medical (%)	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	0.00	1.00
COPS2	39.7	37.9	36.9	38.8	38.1	37.5	38.5	43.8	37.5	36.5	40.0	38.1	37.2	39.1	36.8	37.2	39.7	44.3	1.45	0.10
LAPS2	64.5	61.6	63.3	67.6	62.2	65.0	64.0	60.6	60.2	59.7	63.6	67.4	66.0	61.7	64.9	61.0	62.9	65.3	1.57	0.06

* For the continuous covariates (age, COPS2, LAPS2), if the Kruskal-Wallis chi-square test statistic is significant, it means that at least one hospital stochastically dominates another hospital on the covariate of interest. If the chi-square statistic for the Kruskal-Wallis test divided by its degrees of freedom is larger than 1, patients tend to differ more than expected by random hospital assignment, vice versa.

Table A.2: Covariate means when matched on administrative data & COPS2

Hospital ID	2	3	5	7	8	9	10	11	13	14	16	17	19	20	21	24	25	28	χ^2/df	P-val.
Age	67.8	68.1	68.0	67.8	68.0	67.8	67.8	67.9	68.0	68.4	67.8	68.1	67.9	68.1	67.8	67.7	67.8	67.7	0.06	1.00
Female (%)	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	0.00	1.00
Flu season (%)	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	0.00	1.00
ED surgical (%)	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	0.00	1.00
Non-ED surgical (%)	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	0.00	1.00
ED medical (%)	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	0.00	1.00
Non-ED medical (%)	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	0.00	1.00
COPS2	42.9	42.7	42.7	43.1	42.2	42.3	42.7	43.0	42.4	42.5	42.8	43.5	42.7	42.6	42.7	42.6	42.7	43.0	0.03	1.00
LAPS2	64.5	62.7	64.0	69.3	60.6	65.1	67.0	66.4	60.3	59.7	64.9	68.3	66.4	63.7	64.7	60.9	65.1	66.2	2.24	0.00

Table A.3: Covariate means when matched on administrative data & LAPS2

Hospital ID	2	3	5	7	8	9	10	11	13	14	16	17	19	20	21	24	25	28	χ^2/df	P-val.
Age	67.7	68.0	67.8	67.8	67.7	67.8	67.7	67.7	67.8	68.1	67.8	67.7	67.7	67.7	67.8	67.9	67.9	67.3	0.04	1.00
Female (%)	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	0.00	1.00
Flu season (%)	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	0.00	1.00
ED surgical (%)	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	0.00	1.00
Non-ED surgical (%)	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	0.00	1.00
ED medical (%)	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	0.00	1.00
Non-ED medical (%)	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	0.00	1.00
COPS2	40.2	42.2	36.7	39.7	38.2	37.3	34.6	40.6	39.4	37.9	41.0	40.6	37.7	39.8	36.1	38.3	41.6	44.9	0.99	0.46
LAPS2	64.4	64.4	64.2	64.6	64.0	64.3	64.5	64.3	63.9	63.9	64.3	64.3	64.5	63.9	64.6	63.8	64.2	64.7	0.04	1.00

Table A.4: Covariate means when matched on administrative data, COPS2 & LAPS2

Hospital ID	2	3	5	7	8	9	10	11	13	14	16	17	19	20	21	24	25	28	χ^2/df	P-val.
Age	67.6	68.1	67.9	67.9	68.0	67.7	67.9	67.7	67.8	68.2	67.6	67.9	67.7	67.8	67.7	68.0	67.5	67.1	0.13	1.00
Female (%)	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	0.00	1.00
Flu season (%)	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	36.4	0.00	1.00
ED surgical (%)	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	0.00	1.00
Non-ED surgical (%)	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	0.00	1.00
ED medical (%)	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	54.8	0.00	1.00
Non-ED medical (%)	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	10.8	0.00	1.00
COPS2	42.2	42.1	42.2	42.4	41.6	41.9	42.5	42.3	41.8	41.6	41.7	42.1	42.1	41.8	41.8	42.0	41.9	42.3	0.03	1.00
LAPS2	63.9	63.7	63.8	64.8	63.7	63.8	64.8	63.9	63.5	62.0	63.9	63.8	63.7	63.8	64.0	62.9	63.5	64.3	0.17	1.00

A.2. Supplemental Information on Hospital Ranking Changes

Table A.5: Mean and max absolute changes in hospital rankings

	Admin vs Admin + COPS2		Admin vs Admin + LAPS2		Admin vs Admin + COPS2 + LAPS2	
	Mean	Max	Mean	Max	Mean	Max
30-day mortality	2.9	12	2.4	9	4.4	16
Truncated LOS	2.4	7	1.7	6	2.0	8

To calculate the mean and max absolute changes, we compare each hospital's rank when matched on administrative information only, with its corresponding rank when adding COPS2 and/or LAPS2 into matching. We calculate the absolute difference in each hospital's two ranks, and summarize their mean and max across all 18 hospitals.

Appendix B

Appendix to Chapter 3

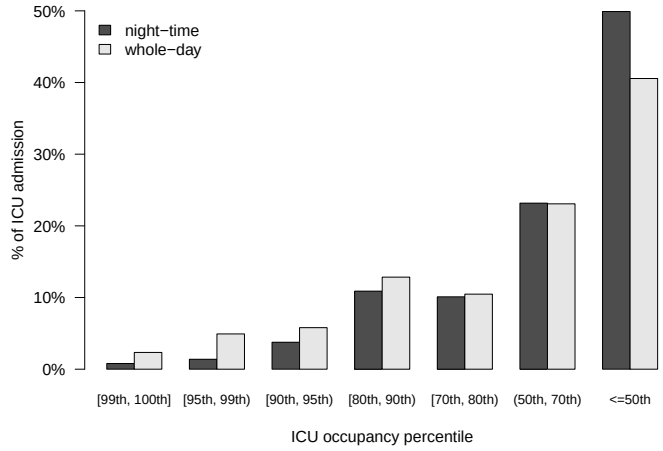
B.1. Supplemental Information on Empirical Analysis

Figure B.1 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.

B.1.1. 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 $\mathbf{x}_{t,\cdot} = \{\mathbf{x}_{t,p_{\mathbf{1}}}, ..., \mathbf{x}_{t,p_{\mathbf{P}}}\}$, and each encouraged $c \in \mathcal{C}$ has a similar vector

Figure B.1: Percentage of ICU transfer by ICU occupancy during night-time and whole-day



$\mathbf{x}_{c,\cdot} = \{\mathbf{x}_{c,p_1}, ..., \mathbf{x}_{c,p_\mathbf{P}}\}$. Let $0 \leq \delta_{t,c} < \infty$ denote the distance between each pair of discouraged and encouraged units. We solve:

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

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

Table B.1: 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

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.

B.1.2. Covariate Balance

By means of the integer program (B.1) 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 B.2–B.4, 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 B.2 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 B.2: 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 B.3: 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

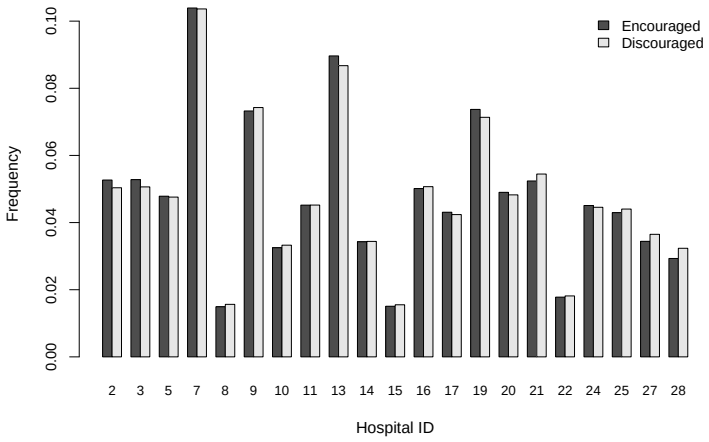
B.1.3. Robustness Checks

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

Table B.4: 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 B.2: Balance table for hospital ID after match



B.1.3.1. Alternative IV Definition.

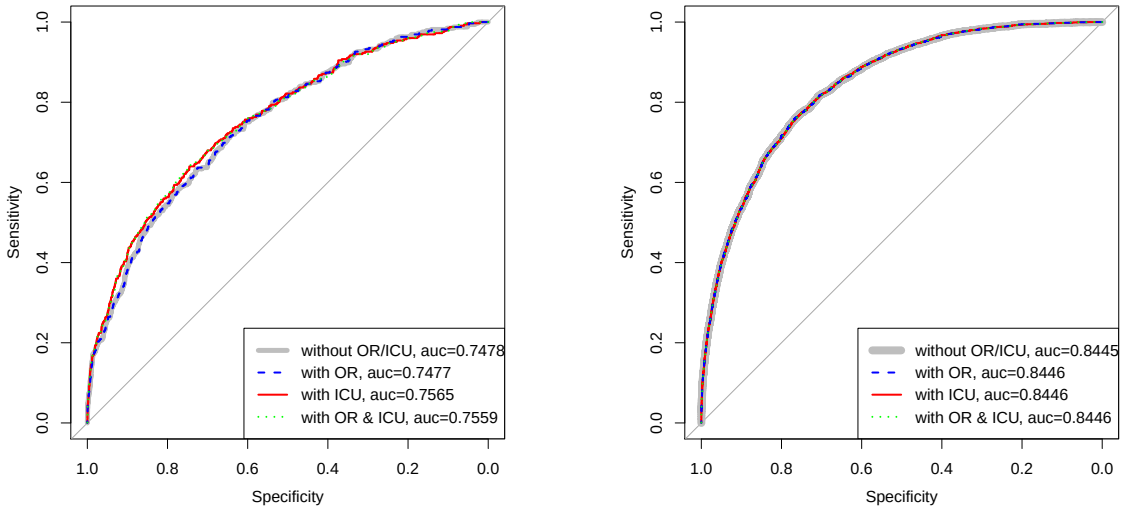
In defining the binary instrumental variable from the continuous ICU occupancy levels, we use the 90^{th} percentile and 70^{th} percentile of the ICU occupancy distribution for each hospital as the threshold for "busy" and "not-busy". We also tried different thresholds, including the 65^{th}, 67.5^{th}, 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.

B.1.3.2. 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 B.3 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 B.3: ROC Curves

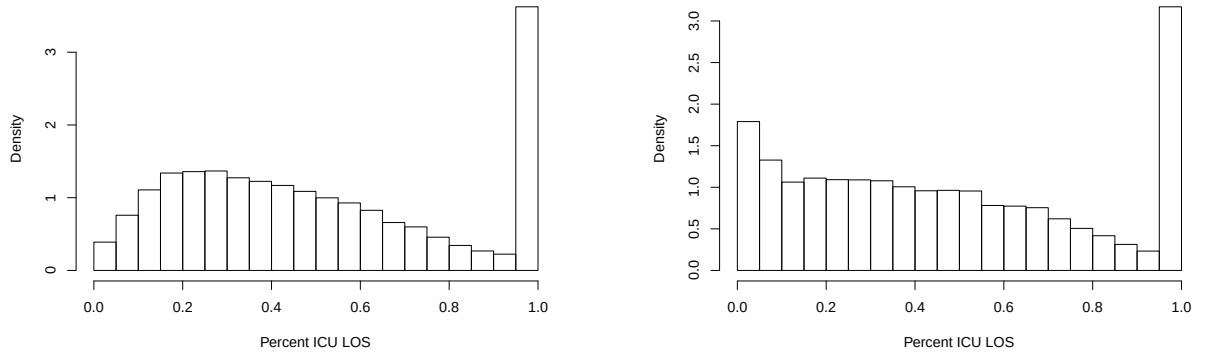


(a) Night-time ICU transfers

(b) Mortality of night-time patients

B.2. Supplemental information for Simulation

Figure B.4: 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 $f_{p_W}(p)$

B.2.1. 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: $i \in \{1, 2, \dots 10\}$, where state i denotes the patient is currently in EDIP2 group 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. $\mathbf{T}_{i,j}, i = 1, 2, \dots, 10, j = 1, 2, \dots, 13$ 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:

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

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.2.2. Optimization Problem to Calibrate Crashed Parameters

We use our empirical results in Section 3.4 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 B.5. 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:

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

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 B.5, 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 B.5. The optimization problem is formulated as:

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

We formulate and solve a similar optimization problem for LOS_C:

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

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 B.6

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 B.5: 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 B.6: 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

B.2.3. Simulation Robustness Checks

ICU size: We consider 4 different ICU sizes $\mathbf{N} = 10, 15, 20, 30$ operated at approximately 70%, 80%

Table B.7: Percentage differences between the best 3- and 4-threshold state-dependent policy static policy

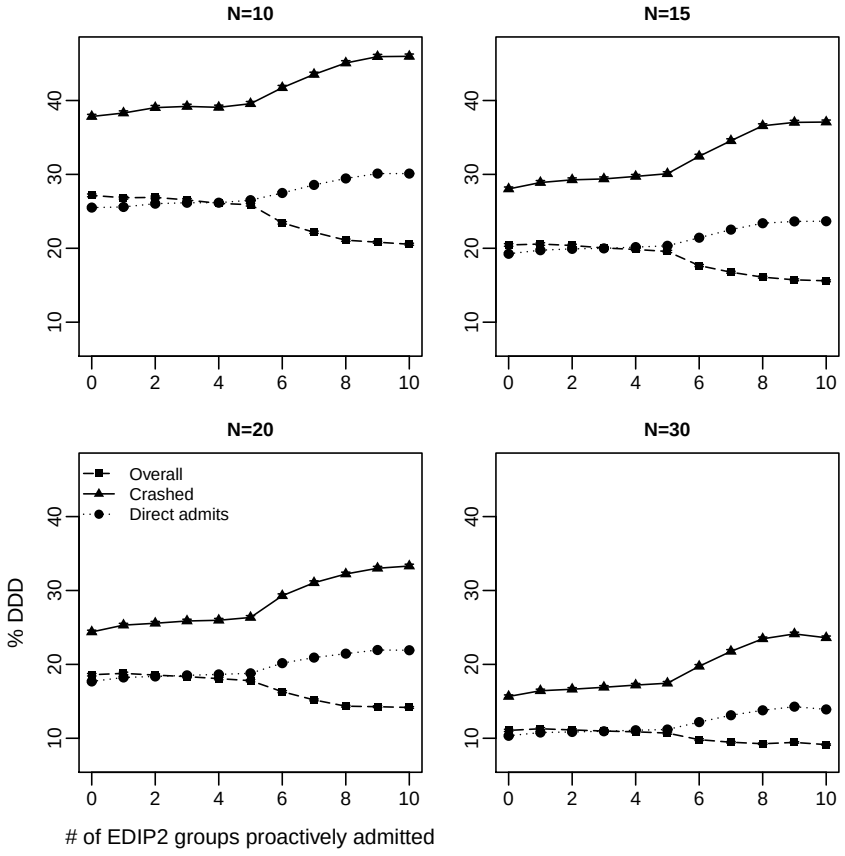
3-threshold						
# of groups	Mortality	LOS	$DDD_{crashed}$	$DDD_{directadmit}$	$r_{crashed}$	$r_{directadmit}$
Top 2	0.19	-0.04	-1.94*	-1.02	0.18	-0.50*
Top 3	1.01*	0.27*	-2.39*	-2.21*	-0.50	-0.13
Top 4	0.74*	0.23*	-2.01*	-1.39*	-0.55	-0.17
Top 5	1.09*	0.26*	-3.25*	-3.17*	-0.25	-0.32
Top 6	2.83*	1.73*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

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 B.5). 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.

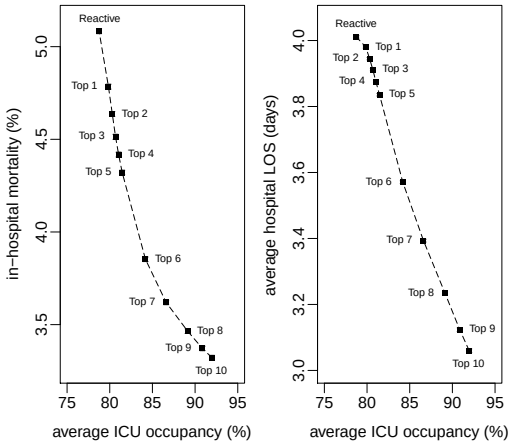
Parameter calibration: We also vary the calibration of some of our model primitives. Specifically, we vary $\beta = [0.1, 0.2, \ldots, 0.9]$, which impacts the ICU readmission rates for proactive transfers, as well as the mortality and readmission rates for external arrivals (d_E and r_E) and the readmission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

Figure B.5: 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.

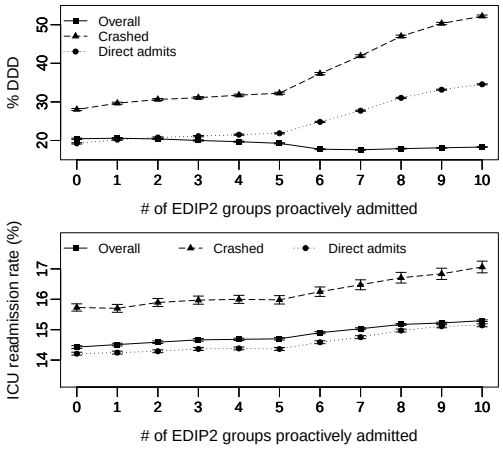


Proactive transfer during the whole day: We next consider the case where proactive transfers can occur during any EDIP2 decision epoch (instead of just the night-time one). Here we assume that our empirical estimates can be generalized to the whole day. These results are summarized in Figure B.6. While 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 B.6: Proactive transfers can occur throughout the day. $\mathbf{N} = 15$ ICU beds. $\Lambda = 14.2$ patients/day.



(a) Mortality and LOS versus ICU occupancy.



(b) Demand-driven discharge and readmission rates

Appendix C

Appendix to Chapter 4

C.1. Supplemental Information on Missing Data Imputation

C.1.1. SAPS3 for ICU visits

Table C.1: Regression table for SAPS3, N=59,125, $R^2 = 0.48$

	Estimate	Std. Error	t value	P-val
LAPS2	0.08	0.00	95.27	0.000
COPS2	0.03	0.00	34.17	0.000
Age	0.28	0.00	114.28	0.000
Sex	0.28	0.08	3.73	0.000
Non-ED, Surgical	-5.13	0.17	-31.02	0.000
ED, Medical	0.74	0.13	5.61	0.000
Non-ED, Medical	2.10	0.19	10.91	0.000
Transport in	8.26	0.21	39.20	0.000

C.1.2. EDIP2 for ward / TCU stay

Figure C.1: QQ plot of residuals from regressing SAPS3 on all other risk covariates

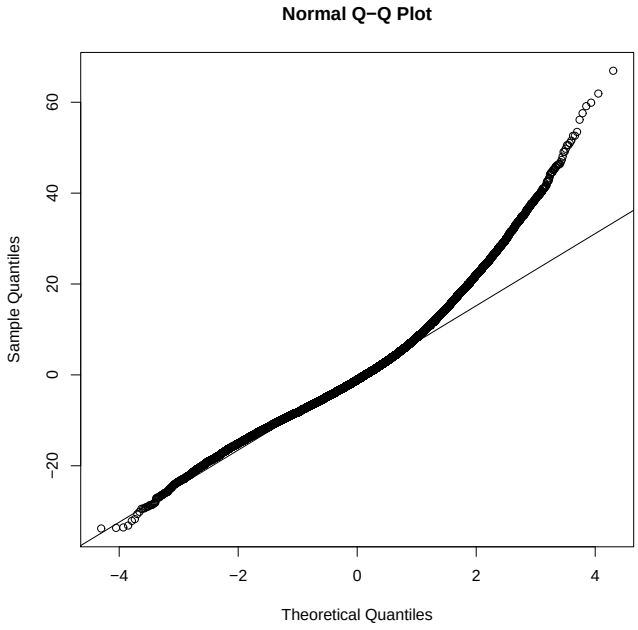


Table C.2: Comparison of hospitalizations with and without EDIP2 scores

	Without EDIP2 ($N = 1,939$)				With EDIP2 ($N = 295,553$)			
	Min	Mean	Median	Max	Min	Mean	Median	Max
LAPS2	0.00	78.89	66.00	288.00	0.00	62.13	54.00	310.00
COPS2	0.00	43.41	23.00	261.00	0.00	37.96	20.00	306.00
CHMR (%)	0.00	8.16	1.14	92.27	0.00	2.90	0.83	97.58
Age	18.00	65.59	68.00	113.00	18.00	64.73	67.00	109.00
Female (%)		48.53				54.42		
Flu season (%)		28.06				29.72		
Transport in (%)		2.42				1.95		
ICU count	0.00	0.38	0.00	5.00	0.00	0.17	0.00	23.00
In-hospital mortality (%)		21.20				2.34		
30-day mortality (%)		29.50				6.32		
LOS (hrs)	0.02	52.85	27.85	1260.00	0.02	86.33	61.13	13030.00

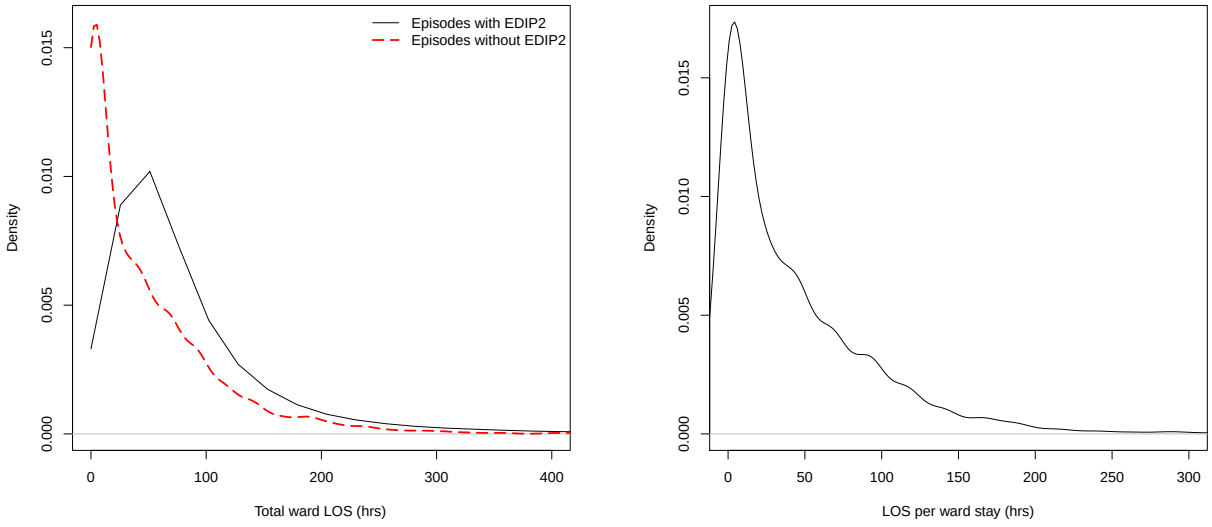
Table C.3: Comparison of hospitalizations without EDIP2 scores due to inconsistent records or quick ward discharge

	Inconsistent records ($N = 1,897$)				Quick ward discharge ($N = 42$)			
	Min	Mean	Median	Max	Min	Mean	Median	Max
LAPS2	0.00	78.09	65.00	288.00	21.00	115.33	118.00	227.00
COPS2	0.00	43.32	23.00	261.00	6.00	47.40	21.00	174.00
CHMR	0.00	7.93	1.10	92.27	0.03	18.36	9.98	78.24
Age	18.00	65.53	68.00	113.00	26.00	68.24	74.50	99.00
Female (%)		48.39				54.76		
Flu season (%)		27.78				40.48		
Transport in (%)		2.42				2.38		
ICU count	0.00	0.37	0.00	5.00	0.00	0.60	0.50	2.00
In-hospital mortality (%)		20.24				64.29		
30-day mortality (%)		28.68				66.67		

Table C.4: Comparison of ward LOS (for each ward admission) without EDIP2 scores due to inconsistent records or quick ward discharge

	Inconsistent records ($N = 2,285$)				Quick ward discharge ($N = 45$)			
	Min	Mean	Median	Max	Min	Mean	Median	Max
LOS (hrs)	0.02	44.79	24.77	847.00	0.02	2.99	3.23	5.60

Figure C.2: Distributions of LOS (hrs) in the ward



(a) Total ward LOS per hospitalization

(b) Ward LOS per ward admission with EDIP2 scores

C.2. Supplemental Information on Reduced Form Regressions

Table C.5: Control variables used in reduced form regressions on *ICU Admit*

Category	Variables
Patient characteristics	Age, Sex, COPS2, LAPS2, CHMR
ICU status	# and proportion of ED/non-ED Medical/Surgical patients in the ICU Average, min, max CHMR, COPS2, LAPS2, SAPS3, Age % Female, % of patients transported into KPNC
ICU admissions	# of ED/non-ED Medical/Surgical patients admitted in the previous hour Average[1] CHMR, COPS2, LAPS2, SAPS3, Age % Female, % of patients transported into KPNC
ICU discharges	# of ED/non-ED Medical/Surgical patients discharged in the previous hour
Ward status	# and proportion of ED/non-ED Medical/Surgical patients in the ward Average, min, max CHMR, COPS2, LAPS2, EDIP2, Age % Female, % of patients transported into KPNC
ED status	# of ED/non-ED Medical/Surgical patients needing decisions in the ED Average, min, max CHMR, COPS2, LAPS2, Age % Female, % of patients transported into KPNC
OR/PAR status	# and proportion of ED/non-ED Medical/Surgical patients in the OR/PAR Average, min, max CHMR, COPS2, LAPS2, Age % Female, % of patients transported into KPNC
Seasonal factors	Decision day-of-week, month, nurse shift
Fixed effects	21 hospital IDs, 38 disease categories

[1] Very few patients recently admitted, mostly 0 or 1, so not including distribution of severity scores.

Table C.6: Control variables used in reduced form regressions on patient outcomes

Variable	Description
Age	Patient age at time of hospital admission; cut into the following bins: 0–39, 40–64, 65–74, 75–84, and above 85 (114 being the max in sample)
Gender	Female = 1 and male = 0
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];
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;
Diagnosis	Primary diagnosis, grouped into 38 broad disease categories (e.g. pneumonia); categorical variables
Hospital ID	21 hospital IDs; categorical variables
Month/Day/Hour	Month/day-of-week/hour of hospital admission; categorical variables

COPS2 is omitted because it is defaulted to 10 if missing, so it may not be accurate.

Table C.7: Logistic regression of ICU admission, showing significant variables only, $N = 279,691$

	Estimate	Std. Error	z value	Pr($>|z|$)
(Intercept)	-1.228	0.449	-2.74	0.006
# ED Medical in ICU	-0.062	0.004	-14.21	0.000
# ED Surgical in ICU	-0.043	0.010	-4.23	0.000
# non-ED Surgical in ICU	-0.037	0.009	-4.11	0.000
% ED Medical in ICU	0.492	0.137	3.58	0.000
% ED Surgical in ICU	0.352	0.171	2.06	0.039
Max LAPS2 in ICU	-0.001	0.000	-2.66	0.008
Min Age in ICU	0.002	0.001	2.47	0.013
Max Age in ICU	-0.003	0.002	-1.96	0.050
% Transport-in in ICU	0.268	0.118	2.28	0.023
# admitted ED Medical pre 1hr	0.131	0.044	2.97	0.003
# admitted non-ED Surgical pre 1hr	-0.114	0.052	-2.18	0.029
Avg LAPS2 admitted pre 1hr	-0.001	0.001	-2.00	0.046
# discharged ED Medical pre 1hr	0.179	0.013	13.77	0.000
# discharged ED Surgical pre 1hr	0.094	0.036	2.61	0.009
# discharged non-ED Medical pre 1hr	0.211	0.035	6.00	0.000
# discharged non-ED Surgical pre 1hr	0.146	0.026	5.71	0.000
# ED Medical in ward	-0.006	0.001	-4.08	0.000
# ED Surgical in ward	-0.021	0.004	-5.78	0.000
# non-ED Medical in ward	-0.012	0.006	-2.10	0.036
% ED Medical in ward	1.131	0.327	3.46	0.001
% ED Surgical in ward	2.096	0.431	4.87	0.000
% non-ED Medical in ward	1.185	0.578	2.05	0.041
Avg CHMR in ward	-3.284	1.520	-2.16	0.031
Avg EDIP2 in ward	-59.614	7.772	-7.67	0.000
Min EDIP2 in ward	83.677	42.903	1.95	0.051
Max EDIP2 in ward	0.703	0.153	4.60	0.000
# ED Medical in ED	-0.152	0.012	-13.17	0.000
Max CHMR in ED	-1.366	0.293	-4.66	0.000
Max LAPS2 in ED	0.004	0.001	3.41	0.001
# non-ED Medical in OR	-0.041	0.013	-3.19	0.001
# non-ED Surgical in OR	-0.012	0.005	-2.15	0.032
Avg Age in OR	-0.007	0.003	-2.12	0.034
Max Age in OR	0.004	0.002	2.32	0.020
LAPS2	0.026	0.001	49.74	0.000
Age	-0.025	0.001	-34.19	0.000
COPS2	-0.001	0.000	-4.26	0.000
CHMR	-1.213	0.187	-6.49	0.000
Sex	-0.114	0.026	-4.41	0.000
Sunday	0.068	0.033	2.09	0.036
February	-0.087	0.044	-1.96	0.050
March	-0.153	0.043	-3.54	0.000
October	0.097	0.045	2.13	0.033
Nurse shift morning	-0.088	0.023	-3.74	0.000
Nurse shift night	0.212	0.022	9.62	0.000

Table C.8: Multinomial logit model on the three decision choices, showing significant variables only, $N = 279,691$

	Admit to the ICU			Wait in the ED		
	Coef.	Std. Err.	Pr(> |z|)	Coef.	Std. Err.	Pr(> |z|)
(Intercept)	-1.887	0.004	0.000	-4.264	0.004	0.000
# ED Medical in ICU	-0.041	0.004	0.000	0.049	0.002	0.000
# ED Surgical in ICU	-0.031	0.007	0.000	0.040	0.005	0.000
# non-ED Medical in ICU	-0.001	0.008	0.891	0.043	0.005	0.000
# non-ED Surgical in ICU	-0.019	0.006	0.002	0.049	0.003	0.000
% ED Medical in ICU	0.412	0.035	0.000	-0.106	0.034	0.002
% ED Surgical in ICU	0.349	0.019	0.000	-0.039	0.032	0.223
% non-ED Medical in ICU	-0.292	0.012	0.000	-0.205	0.026	0.000
Avg CHMR in ICU	-0.308	0.008	0.000	-0.525	0.004	0.000
Min CHMR in ICU	0.657	0.001	0.000	-1.388	0.000	0.000
Max CHMR in ICU	0.051	0.048	0.291	0.210	0.027	0.000
Max COPS2 in ICU	0.000	0.000	0.776	-0.001	0.000	0.000
Max LAPS2 in ICU	-0.001	0.000	0.017	0.000	0.000	0.529
Max SAPS3 in ICU	0.001	0.001	0.327	0.002	0.001	0.003
Min Age in ICU	0.002	0.001	0.063	-0.002	0.001	0.001
Max Age in ICU	-0.002	0.002	0.214	0.003	0.001	0.001
% Transport-in in ICU	0.246	0.015	0.000	-0.208	0.044	0.000
# admitted ED Medical pre 1hr	0.130	0.026	0.000	0.006	0.014	0.677
# admitted non-ED Surgical pre 1hr	-0.143	0.033	0.000	-0.041	0.017	0.016
Avg CHMR admitted pre 1hr	-0.284	0.003	0.000	-0.024	0.002	0.000
Avg LAPS2 admitted pre 1hr	-0.001	0.000	0.010	0.000	0.000	0.786
% Transport-in admitted pre 1hr	-0.053	0.014	0.000	0.094	0.046	0.040
# discharged ED Medical pre 1hr	0.159	0.010	0.000	-0.028	0.006	0.000
# discharged ED Surgical pre 1hr	0.060	0.026	0.022	-0.066	0.015	0.000
# discharged non-ED Medical pre 1hr	0.215	0.026	0.000	0.034	0.015	0.026
# discharged non-ED Surgical pre 1hr	0.135	0.019	0.000	-0.011	0.010	0.296
# ED Medical in ward	0.011	0.001	0.000	0.044	0.000	0.000
# ED Surgical in ward	0.000	0.002	0.809	0.051	0.001	0.000
# non-ED Medical in ward	-0.012	0.003	0.000	0.022	0.001	0.000
# non-ED Surgical in ward	0.010	0.002	0.000	0.019	0.001	0.000
% ED Medical in ward	0.108	0.009	0.000	-1.936	0.010	0.000
% ED Surgical in ward	0.777	0.004	0.000	-2.884	0.007	0.000
% non-ED Medical in ward	1.564	0.003	0.000	0.286	0.002	0.000
Avg CHMR in ward	-2.941	0.002	0.000	0.125	0.001	0.000
Min CHMR in ward	52.619	0.000	0.000	-7.790	0.000	0.000
Max CHMR in ward	0.108	0.058	0.063	0.142	0.032	0.000
Avg LAPS2 in ward	0.007	0.002	0.000	0.011	0.001	0.000
Min LAPS2 in ward	-0.001	0.002	0.542	-0.002	0.001	0.028
Max LAPS2 in ward	0.001	0.000	0.112	0.001	0.000	0.000
Max Age in ward	-0.005	0.002	0.016	0.000	0.001	0.707
% Female in ward	0.141	0.006	0.000	-0.094	0.009	0.000
% Transport-in in ward	-0.394	0.002	0.000	-1.141	0.002	0.000
Avg EDIP2 in ward	-94.466	0.000	0.000	-84.387	0.001	0.000
Min EDIP2 in ward	63.850	0.000	0.000	20.298	0.000	0.000
Max EDIP2 in ward	1.027	0.010	0.000	0.759	0.053	0.000
# ED Medical in ED	0.014	0.012	0.240	0.252	0.005	0.000
Avg CHMR in ED	0.425	0.016	0.000	1.869	0.016	0.000
Min CHMR in ED	0.009	0.012	0.466	-0.774	0.009	0.000
Max CHMR in ED	-0.980	0.023	0.000	-0.577	0.032	0.000
Avg COPS2 in ED	0.002	0.002	0.138	0.004	0.001	0.000
Min COPS2 in ED	-0.002	0.001	0.021	-0.004	0.000	0.000
Avg LAPS2 in ED	-0.001	0.002	0.644	0.003	0.001	0.006
Min LAPS2 in ED	-0.001	0.001	0.492	-0.002	0.001	0.000
Max LAPS2 in ED	0.002	0.001	0.076	-0.001	0.000	0.018
Max Age in ED	-0.005	0.002	0.048	0.003	0.001	0.010
% Female in ED	-0.015	0.033	0.646	0.052	0.018	0.004
# ED Medical in OR	0.008	0.010	0.422	0.055	0.006	0.000
# non-ED Medical in OR	-0.009	0.012	0.446	0.074	0.007	0.000

# non-ED Surgical in OR	-0.005	0.005	0.376	0.009	0.003	0.001
% ED Medical in OR	-0.079	0.043	0.065	-0.137	0.029	0.000
% ED Surgical in OR	-0.058	0.043	0.172	-0.058	0.028	0.039
% non-ED Medical in OR	-0.104	0.048	0.028	-0.336	0.035	0.000
Avg CHMR in OR	-1.248	0.000	0.000	-0.069	0.001	0.000
Min CHMR in OR	0.001	0.000	0.002	-0.641	0.000	0.000
Max CHMR in OR	0.458	0.001	0.000	0.252	0.001	0.000
Avg LAPS2 in OR	0.002	0.002	0.211	0.002	0.001	0.009
Min LAPS2 in OR	-0.001	0.001	0.625	-0.001	0.001	0.082
Max LAPS2 in OR	-0.001	0.001	0.012	0.000	0.000	0.135
Avg Age in OR	-0.005	0.003	0.141	0.005	0.002	0.011
Min Age in OR	0.001	0.002	0.494	-0.002	0.001	0.034
% Female in OR	-0.056	0.029	0.056	-0.044	0.017	0.008
% Transport-in in OR	-0.237	0.003	0.000	-0.230	0.002	0.000
LAPS2	0.026	0.000	0.000	0.000	0.000	0.023
Age	-0.028	0.001	0.000	-0.007	0.000	0.000
COPS2	-0.001	0.000	0.001	0.001	0.000	0.002
CHMR	0.342	0.021	0.000	2.609	0.018	0.000
Sex	-0.118	0.026	0.000	-0.008	0.014	0.539
Monday	-0.038	0.032	0.238	-0.139	0.019	0.000
Sunday	0.006	0.033	0.865	-0.110	0.020	0.000
Thursday	-0.042	0.032	0.193	-0.044	0.018	0.014
Tuesday	0.000	0.032	0.999	-0.083	0.018	0.000
Wednesday	-0.022	0.032	0.504	-0.061	0.018	0.001
February	-0.071	0.027	0.010	0.071	0.022	0.001
March	-0.072	0.026	0.006	0.117	0.021	0.000
April	-0.047	0.027	0.079	0.047	0.023	0.041
May	0.002	0.026	0.953	-0.055	0.023	0.017
August	-0.012	0.026	0.660	-0.052	0.024	0.032
October	0.029	0.026	0.265	-0.171	0.024	0.000
November	-0.026	0.034	0.452	-0.129	0.027	0.000
Nurse shift morning	0.240	0.024	0.000	0.675	0.013	0.000
Nurse shift night	0.299	0.023	0.000	0.199	0.013	0.000

Table C.9: Summary of regression results on patients with LAPS2 $\geq$ 110. ICUBusy = 1 if the ICU occupancy $\geq$ the 95th percentile of each hospital's ICU occupancy distribution. First-stage F-statistic for the IVs = 28.4. Results for categorical variables are omitted for conciseness of presentation.

	(1) Mortality	(2) LOS	(3) Readmit (1m)	(4) Readmit (2w)
First-stage regression on ICU admission decision				
ICUBusy	-0.290***	-0.361***	-0.285***	-0.285***
	(0.038)	(0.050)	(0.042)	(0.042)
PctSevere	-0.189*	-0.163	-0.187	-0.187
	(0.0960)	(0.128)	(0.106)	(0.107)
RecentAdmission	0.023	-0.079	-0.078	-0.075
	(0.299)	(0.392)	(0.332)	(0.332)
RecentDischarge	0.993***	1.428***	1.012***	1.013***
	(0.131)	(0.173)	(0.143)	(0.144)
Female	-0.013	-0.029	-0.022	-0.022
	(0.018)	(0.025)	(0.020)	(0.020)
LAPS2	0.025***	0.034***	0.024***	0.024***
	(0.001)	(0.001)	(0.001)	(0.001)
CHMR	-1.846***	-2.508***	-1.457***	-1.457***
	(0.142)	(0.184)	(0.173)	(0.173)
Age[40,65)	-0.364***	-0.511***	-0.408***	-0.408***
	(0.059)	(0.084)	(0.063)	(0.063)
Age[65,75)	-0.608***	-0.845***	-0.684***	-0.685***
	(0.060)	(0.085)	(0.063)	(0.063)
Age[75,85)	-0.849***	-1.181***	-0.945***	-0.945***
	(0.060)	(0.085)	(0.063)	(0.063)
Age[85,114]	-1.187***	-1.637***	-1.283***	-1.283***
	(0.061)	(0.086)	(0.065)	(0.065)
Second-stage regression on outcomes				
ICUAdmit	-0.163	-0.347***	-0.071	-0.037
	(0.115)	(0.017)	(0.165)	(0.167)
Female	0.017	-0.009	-0.070***	-0.051*
	(0.020)	(0.010)	(0.019)	(0.021)
LAPS2	0.002*	0.010***	0.003*	0.004*
	(0.001)	(0.000)	(0.001)	(0.001)
CHMR	3.142***	-1.396***	-0.284	-0.454*
	(0.160)	(0.075)	(0.173)	(0.189)
Age[40,65)	0.023	0.068	0.147*	0.032
	(0.080)	(0.036)	(0.070)	(0.075)
Age[65,75)	0.027	0.046	0.169*	0.0331
	(0.083)	(0.036)	(0.078)	(0.082)
Age[75,85)	-0.045	-0.033	0.132	0.031
	(0.085)	(0.036)	(0.084)	(0.088)
Age[85,114]	-0.039	-0.159***	-0.035	-0.127
	(0.090)	(0.037)	(0.094)	(0.098)
AcgOccVisited	-0.976***	0.118*	0.087	0.032
	(0.119)	(0.056)	(0.113)	(0.125)
N	26,346	25,881	22,235	22,235

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C.10: Summary of regression results on all patients. ICUBusy = 1 if the ICU occupancy $\geq$ the 95th percentile of each hospital's ICU occupancy distribution. First-stage F-statistic for the IVs = 97.3. Results for categorical variables are omitted for conciseness of presentation.

	(1) Mortality	(2) LOS	(3) Readmit (1m)	(4) Readmit (2w)
First-stage regression on ICU admission decision				
ICUBusy	-0.257***	-0.338***	-0.251***	-0.251***
	(0.020)	(0.026)	(0.020)	(0.020)
PctSevere	-0.217***	-0.292***	-0.209***	-0.210***
	(0.049)	(0.066)	(0.051)	(0.051)
RecentAdmission	0.410**	0.388*	0.394*	0.392*
	(0.147)	(0.197)	(0.153)	(0.153)
RecentDischarge	0.822***	1.047***	0.829***	0.832***
	(0.064)	(0.087)	(0.067)	(0.067)
Female	-0.070***	-0.088***	-0.073***	-0.073***
	(0.009)	(0.013)	(0.009)	(0.009)
LAPS2	0.014***	0.018***	0.013***	0.013***
	(0.000)	(0.000)	(0.000)	(0.000)
DEATH_HAT3	-0.049	0.159	0.245*	0.245*
	(0.078)	(0.102)	(0.096)	(0.096)
Age[40,65)	-0.238***	-0.349***	-0.240***	-0.240***
	(0.016)	(0.024)	(0.017)	(0.017)
Age[65,75)	-0.436***	-0.627***	-0.446***	-0.447***
	(0.018)	(0.026)	(0.018)	(0.018)
Age[75,85)	-0.645***	-0.902***	-0.662***	-0.662***
	(0.018)	(0.026)	(0.019)	(0.019)
Age[85,114]	-0.893***	-1.222***	-0.904***	-0.905***
	(0.021)	(0.028)	(0.021)	(0.021)
Second-stage regression on outcomes				
ICUAdmit	-0.193**	-0.337***	-0.128*	-0.050
	(0.061)	(0.007)	(0.058)	(0.066)
Female	-0.023	0.002	-0.023**	-0.025**
	(0.013)	(0.004)	(0.008)	(0.009)
LAPS2	0.009***	0.007***	0.005***	0.004***
	(0.000)	(0.000)	(0.000)	(0.000)
CHMR	2.583***	-0.601***	-0.318***	-0.343***
	(0.089)	(0.038)	(0.086)	(0.095)
Age[40,65)	0.225***	0.081***	0.108***	0.047**
	(0.040)	(0.008)	(0.016)	(0.017)
Age[65,75)	0.312***	0.098***	0.173***	0.087***
	(0.042)	(0.009)	(0.017)	(0.019)
Age[75,85)	0.312***	0.058***	0.149***	0.089***
	(0.042)	(0.009)	(0.018)	(0.020)
Age[85,114]	0.360***	-0.031***	0.076***	0.023
	(0.044)	(0.009)	(0.020)	(0.023)
AvgOccVisited	-1.007***	-0.056*	0.013	0.028
	(0.074)	(0.023)	(0.044)	(0.050)
N	170,007	168,395	163,115	163,115

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C.11: Summary of regression results on patients with LAPS2 $\geq$ 110, without IVs. Results for categorical variables are omitted for conciseness of presentation.

	(1) Mortality	(2) LOS	(3) Readmit (1m)	(4) Readmit (2w)
ICUAdmit	0.158***	0.292***	-0.009	-0.007
	(0.025)	(0.012)	(0.024)	(0.026)
Female	0.018	-0.011	-0.070***	-0.051*
	(0.020)	(0.010)	(0.019)	(0.021)
LAPS2	-0.000	0.005***	0.002**	0.003***
	(0.001)	(0.000)	(0.001)	(0.001)
CHMR	3.344***	-1.020***	-0.259	-0.442*
	(0.139)	(0.081)	(0.161)	(0.178)
Age[40,65)	0.060	0.143***	0.156*	0.036
	(0.080)	(0.038)	(0.067)	(0.071)
Age[65,75)	0.090	0.167***	0.183**	0.040
	(0.080)	(0.038)	(0.068)	(0.072)
Age[75,85)	0.043	0.128***	0.151*	0.039
	(0.080)	(0.038)	(0.067)	(0.072)
Age[85,114]	0.080	0.055	-0.011	-0.116
	(0.081)	(0.039)	(0.069)	(0.074)
AvgOccVisited	-0.975***	0.095	0.088	0.032
	(0.120)	(0.059)	(0.113)	(0.125)
N	26,346	25,881	22,235	22,235

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C.12: Summary of regression results on all patients, without IVs. Results for categorical variables are omitted for conciseness of presentation.

	(1) Mortality	(2) LOS	(3) Readmit (1m)	(4) Readmit (2w)
ICUAdmit	0.210***	0.262***	-0.006	0.008
	(0.018)	(0.007)	(0.013)	(0.014)
Female	-0.017	0.005	-0.021**	-0.025**
	(0.013)	(0.004)	(0.008)	(0.009)
LAPS2	0.007***	0.005***	0.005***	0.004***
	(0.000)	(0.000)	(0.000)	(0.000)
CHMR	2.567***	-0.750***	-0.361***	-0.363***
	(0.090)	(0.045)	(0.084)	(0.093)
Age[40,65)	0.253***	0.107***	0.114***	0.049**
	(0.041)	(0.009)	(0.015)	(0.017)
Age[65,75)	0.364***	0.148***	0.183***	0.092***
	(0.041)	(0.010)	(0.016)	(0.018)
Age[75,85)	0.387***	0.126***	0.164***	0.096***
	(0.041)	(0.010)	(0.016)	(0.018)
Age[85,114]	0.461***	0.058***	0.096***	0.033
	(0.042)	(0.010)	(0.018)	(0.020)
AvgOccVisited	-1.012***	-0.094***	0.013	0.028
	(0.075)	(0.024)	(0.044)	(0.050)
N	170,007	168,395	163,115	163,115

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$