2022

Essays On The Intersection Of Healthcare Operations And Economics

Jong Myeong Lim
University of Pennsylvania

Essays On The Intersection Of Healthcare Operations And Economics

Abstract

The essays in this dissertation wrestle with unique challenges presented by multiple, interacting entities within the healthcare industry. The essay, "Searching for the Best Yardstick: Cost of Quality Improvements in the U.S. Hospital Industry," takes the perspective of the regulator in improving incentive programs designed to induce hospitals to invest in quality. The key challenge in evaluating potential changes to such programs is to understand the underlying incentives that hospitals have in responding to the new incentives. Using structural estimation methods, the parameters of each hospital's decision-making process are estimated. The counterfactual analyses quantify the effects of recalibrating the Hospital Value-based Purchasing Program. The essay, "The Spillover Effects of Capacity Pooling in Hospitals," focuses on the unintended effects of off-service placement, a common capacity pooling strategy. Building on previous studies that document negative first-order effects on patients who are placed off service themselves, the spillover effects onto patients who are placed on service are analyzed. The instrumental variables approach reveals that there is a significant causal impact of off-service placement on patients who are placed on service. The essay, "Should We Worry About Moral Hazard? Estimation of the Slutsky Equation Using Indemnity Health Insurance Contracts," uncovers the differential response of consumers to different designs of health insurance. While previous studies have convincingly shown that ex-post moral hazard in health care does exist, there has been a lack of empirical evidence on the degree in which such moral hazard is welfare-reducing. Using a novel setting, the analysis provides evidence that moral hazard can lead to a significant welfare loss.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Health Care Management & Economics

First Advisor

Hummy Song

Second Advisor

Ken Moon

Subject Categories

Economics | Health and Medical Administration

ESSAYS ON THE INTERSECTION OF HEALTHCARE OPERATIONS AND ECONOMICS

Jong Myeong Lim

A DISSERTATION

in

Health Care Management & Economics

For the Graduate Group in Managerial Science and Applied Economics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2022

Supervisor of Dissertation

Hummy Song

Assistant Professor of Operations, Information and Decisions

Co-Supervisor of Dissertation

Ken Moon

Assistant Professor of Operations, Information and Decisions

Graduate Group Chairperson

Nancy Zhang, Professor of Statistics and Data Science

Dissertation Committee

Sergei Savin, Associate Professor of Operations, Information and Decisions
Claudio Lucarelli, Associate Professor of Health Care Management

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to the members of my committee, Professors Claudio Lucarelli, Ken Moon, Sergei Savin, and Hummy Song. It has been an honor to be your student, and I hope to follow your footsteps throughout my academic journey. I would like to especially thank Professor Hummy Song for introducing me to the field of operations management. I would also like to thank the faculty and staff members in the Healthcare Management Department and the Operations, Information and Decisions Department for all of their help and support throughout my time here at Wharton.

I also thank all of my fellow PhD students and friends I have met in Philadelphia for cherished times working through problem sets, discussing research, and sometimes just having fun going out and playing tennis.

I am extremely grateful to my parents, for without their love and sacrifice, I would not be where I am today. I would also like to thank my family in Incheon for their love and blessing. Above all, I give my deepest love, gratitude, and appreciation to my wife and my best friend, Eunae. I cannot thank you enough for your unrelenting support throughout this whole process. I would not have been able to do this without you.

ABSTRACT

ESSAYS ON THE INTERSECTION OF HEALTHCARE OPERATIONS AND ECONOMICS

Jong Myeong Lim

Hummy Song

Ken Moon

The essays in this dissertation wrestle with unique challenges presented by multiple, interacting entities within the healthcare industry. The essay, "Searching for the Best Yardstick: Cost of Quality Improvements in the U.S. Hospital Industry," takes the perspective of the regulator in improving incentive programs designed to induce hospitals to invest in quality. The key challenge in evaluating potential changes to such programs is to understand the underlying incentives that hospitals have in responding to the new incentives. Using structural estimation methods, the parameters of each hospital's decision-making process are estimated. The counterfactual analyses quantify the effects of recalibrating the Hospital Value-based Purchasing Program. The essay, "The Spillover Effects of Capacity Pooling in Hospitals," focuses on the unintended effects of off-service placement, a common capacity pooling strategy. Building on previous studies that document negative first-order effects on patients who are placed off service themselves, the spillover effects onto patients who are placed on service are analyzed. The instrumental variables approach reveals that there is a significant causal impact of off-service placement on patients who are placed on service. The essay, "Should We Worry About Moral Hazard? Estimation of the Slutsky Equation Using Indemnity Health Insurance Contracts," uncovers the differential response of consumers to different designs of health insurance. While previous studies have convincingly shown that *ex-post* moral hazard in health care does exist, there has been a lack of empirical evidence on the degree in which such moral hazard is welfare-reducing. Using a novel setting, the analysis provides evidence that moral hazard can lead to a significant welfare loss.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1

Introduction

One of the defining characteristics of the healthcare industry is the existence of multiple entities with unique goals and incentives. The success of a healthcare system, therefore, relies not only on managing the unique challenges that each entity faces but also on analyzing the interactions among the different entities. This dissertation focuses on three main perspectives within the healthcare system: regulator, healthcare organizations, and patients.

First, the regulator designs the nature of healthcare markets, for instance, the market for healthcare where patients seek medical care and healthcare organizations. The role of the regulator has become even more salient as the industry is shifting towards "value-based" care, where the goal is to focus more resources on high-value care while curtailing the growth of costs. Designing appropriate payment schemes and analyzing the impact of various incentive policies have become important issues. Second, various types of healthcare organizations seek to provide direct and indirect medical care to consumers. The healthcare organizations are faced with unique challenges in managing various aspects of healthcare delivery, including staffing, capacity management, and quality improvements. The decisions made by healthcare organizations, often in response to changes in incentive structures, involve trade-offs where the impacts on patients and workflows are not immediately obvious and can be unintended. Lastly, patients are the ultimate users of the healthcare system. Any organization-level or policy-level change will have an impact on patient behavior. Therefore, it is important to understand the impact that the interaction between patient population and healthcare organizations as well as payors have on health care systems.

Chapter 2 focuses on analyzing policy levers designed to incentivize quality improvements in hospitals. Policy levers are especially important in the hospital market because without any external incentives, financial gains from quality improvements are realized only when patient volume increases as a result of higher quality relative to competitors. However, there is a limited patient response to quality because hospitals often operate as local monopolies or oligopolies and because an increasing number of insurance plans limit in-network hospitals. This unique structure of the hospital market makes the *laissez-faire* approach of relying on patient choice an insufficient strategy for incentivizing hospitals to improve their quality of care. The essay, "Searching for the Best Yardstick: Cost of Quality Improvements in the U.S. Hospital Industry," takes the perspective of the policymakers to analyze and improve incentive programs that are designed to induce investments in quality of care.

Using data on Medicare's Value-Based Purchasing (VBP) Program, the essay first seeks to understand how organizations respond to financial incentives. Using structural estimation methods, parameters that govern hospitals' investment decisions are recovered. Specifically, a dynamic equilibrium model that involves investment decisions and a nested quality ladder is employed and estimated. Then, based on the estimated system, the counterfactual analyses explore the benefits, on the one hand, of modifying the overall size of the incentives and, on the other hand, of implementing a more focused program tailored to hospital type. The essay finds that increasing the size of the incentives from 2% to 4% would have resulted in an additional quality investment of US$1.2B from 2011 to 2018, leading to a 3.3% reduction in the average rate of central-line associated bloodstream infections (CLABSI). Furthermore, applying the incentives to the tailored hospital peer groups, even without changing the size of the incentives, can lead to an average reduction of 1.4% in the rate of CLABSI among groups of hospitals associated with the highest costs of quality investment. The results suggest that hospitals with higher costs of quality investment, and hence often of lower quality, are in fact under-incentivized in the current scheme, and the system would be better off if hospitals of similar types competed within a smaller peer group.

Chapter 3 turns to analyzing the impact of organization-level decisions on productivity and efficiency. In particular, the essay, "The Spillover Effects of Capacity Pooling in Hospitals," focuses on a capacity pooling strategy, often referred to as off-service placement, that addresses the day-to-day mismatch between the supply and demand of hospital beds. The key issue is that demand has large variance, in terms of both the number of arrivals and the types of patients, while supply of hospital beds is fixed, at least in the short run. This leads to situations where a patient must be placed in a bed that is not allocated to the specialty service the patient is admitted to, creating an off-service placement. Building on prior work that documents the negative first-order effects of off-service placement on patients who are placed off service themselves, this essay reveals that off-service placement causes a substantial spillover effect onto patients who are placed on service. Through causal inference, the essay finds that a one percentage point increase in the average proportion of the service's patients who are placed off service during a patient's hospitalization leads to a 2.7% increase in length of stay. Through a series of counterfactual analyses, alternate routing policies that could meaningfully improve the efficiency of care are considered.

Chapter 4 analyzes patient behaviors under different types of insurance contracts. The phenomenon in which the level of healthcare consumption increases when consumers have health insurance plans that reduce the purchasing price of medical care has been the focal point of a large stream of literature in health economics. While there has been convincing evidence documenting the existence of such "ex-post" moral hazard, the question of how much moral hazard is in fact welfare-reducing has been less studied. The key idea is that changes in consumer behavior is only welfare-reducing if such changes are driven by the substitution effect, in other words, purely driven by distortions in the price of healthcare. On the other hand, changes in consumer behavior caused by the income effect, or driven by changes in the purchasing power caused by lower prices in healthcare, are not welfare-reducing. Most of the health economics literature has assumed that most of the distortions in behavior due to moral hazard is welfare-reducing, while a few has argued that a substantial portion of the changes in consumer behavior is not welfare-reducing.

The challenge in tackling this debate is the lack of an appropriate setting where lump-sum transfer insurance contracts exist. The essay, "Should We Worry About Moral Hazard? Estimation of the Slutsky Equation Using Indemnity Health Insurance Contracts," exploits the private insurance market in South Korea where fixed indemnity contracts are popular. Since a lump-sum transfer does not alter the price of healthcare, changes in the behavior of patients with fixed indemnity contracts can be used to estimate the extent to which changes in consumer behavior is driven by non-welfare-reducing income effect. Subsequently, the changes in the behavior of patients with price-changing insurance contract in the same market can be used to identify the total magnitude of moral hazard. This results in the ability to indirectly estimate the portion of moral hazard that is in fact welfare-reducing. Using a nationwide representative sample, each component of the Slutsky equation is estimated, and the essay finds that around 80% of observable moral hazard is welfare-reducing and confirms the long-held belief.

Chapter 5 concludes, summarizes the findings, and lays out the directions for future research.

CHAPTER 2

Searching for the Best Yardstick: Cost of Quality Improvements in the U.S. Hospital Industry

2.1. Introduction

Yardstick competition is a policy tool deployed by regulators to engage firms in induced market competition beyond their local market boundaries. It can be especially useful when managing the incentive structure of firms that have local monopoly or oligopoly power, as often is the case in the health care industry. In its original formulation by Shleifer (1985), yardstick competition focused on extracting cost reduction efforts from monopolists by setting the firm's price, or compensation level, to be equal to the average cost of production among firms in the same industry. Tying the price to an industry cost benchmark creates an effective competition among monopolists operating in separate local markets. The driving force behind yardstick competition is the fact that the cost reduction efforts of a single firm not only result in extra profits for that firm, thereby incentivizing its cost-saving efforts, but also lead to lower prices for all other firms in the industry, ultimately forcing decreases in both costs and prices. Thus, individual firms' competitive efforts for cost reduction globally—and as Shleifer (1985) shows, efficiently—decrease both costs and prices across the covered markets.

Medicare as well as many large insurance firms in the U.S. have been utilizing the principles of such cost-based yardstick competition for inpatient care. Hospitals are compensated using a fixed-price payment scheme based on the given patient's "diagnostic-related groups" (DRGs), intended to create incentives for hospitals to improve efficiency and reduce the cost of care to maximize profits. The fixed-price system, otherwise known as prospective payment system, closely follows the original formulation of the yardstick competition by setting the level of payments in each DRG based on the expected utilization of resources needed to provide appropriate care for an average patient within that DRG. The idea of cost-based yardstick competition has been widely accepted and applied in the health care setting, and the general trend has been to expand the scope of the induced competition to include outpatient care as well as post-acute care within the framework of the "bundled payment" model.

However, these programs leave a gap in incentivizing quality of delivered care. In addition to the cost of care, the quality of care has always been considered as one of the most crucial characteristics of any health care system. In an ideal world, patients would be able to assess and compare the quality of care and choose hospitals based on the quality of care they deliver. The first barrier in regulating quality competition among care providers is that information about the quality of a given hospital is often difficult for patients to acquire and interpret. Medicare has partially addressed this issue by incentivizing the disclosure of quality-of-care data, organizing the data in a public website, and using a simple "star rating" system for hospitals. However, even if the quality of hospital care is known, hospitals operating as local monopolies or oligopolies may not respond strongly to patient choice behavior. As the hospital industry becomes increasingly concentrated owing to recent waves of mergers and acquisitions as well as tightening regulations that increase the barriers to entry, the ability of patients to actively select hospitals on the basis of quality of care may become more limited.

In this regard, using quality-based yardstick incentives presents a promising policy solution. In general, a quality-based yardstick competition program should collect relevant quality data from all firms in the industry and institute transfer payments that reward firms for higher quality services while effectively penalizing their lower-quality peers. Along these lines, Savva et al. (2019) propose modifying yardstick competition to induce hospitals to reduce patient wait times. Under their proposed scheme, hospital fixed-price reimbursements are complemented by payments that depend, for each hospital, on the difference between the average patient wait time at that hospital and the industry-average patient wait time. In this paper, we focus on Medicare's Hospital Value-Based Purchasing (VBP) Program, which implements multidimensional quality-based yardstick competition pursuant to the Affordable Care Act of 2010. All hospitals in the U.S. with inpatient volume from Medicare patients are subject to this program, and 2% of all Medicare inpatient reimbursements are withheld and redistributed across hospitals based on their performance on a common set of selected quality measures. This program allows us to empirically investigate an influential nation-wide policy implementing dynamic quality competition under yardstick incentives.

Conceptually, the quality-based yardstick competition mirrors its cost-based counterpart in that hospitals' current cost or quality levels can be observed, but the hospitals' respective abilities to reduce costs or improve quality are private information. Indeed, Shleifer (1985)'s yardstick regime shares similarities with the mechanism design literature, which focuses on inducing efficient actions under private information. However, our setting departs from Shleifer (1985) in one critical aspect. In the original cost-based setting, the profit-maximizing firm fully internalizes the gains from its own cost reductions. This leads to efficient cost reductions in the absence of any detailed knowledge, from the policymaker's perspective, of individual firms' efforts associated with cost reductions. In contrast, under the VBP Program, hospitals do not automatically capture the entire economic benefit from their own quality improvements. Therefore, calibrating the yardstick incentives based on informed estimates of the costs of quality investment is an important element in the design of the quality-based yardstick competition.

In the setting we analyze, hospitals decide whether to invest in quality based on their assessment of the yardstick incentive payments they expect to receive and the costs they incur. Clearly, if the size of yardstick incentives is too small compared to the cost of quality investment, hospitals will not respond to the yardstick competition. Furthermore, because the costs of quality investment are likely to differ across hospitals, it is important to understand the distribution of investment costs to assess the impact of the VBP incentives on the entire hospital industry. In our analysis, we adopt the perspective of government regulators, such as the Centers for Medicare and Medicaid Services (CMS), and analyze the outcomes of the current version of the VBP Program with the goal of estimating the unobserved costs of quality investment for all hospitals subject to the program. Accomplishing this goal allows us to focus on the following research questions:

1. How are the quality-of-care outcomes affected by the size of the yardstick incentives?
2. What are the potential effects of tailoring yardstick incentives to hospital types?

An important feature of the outcomes of quality investment is that they cannot be analyzed in a static manner: improvements in quality achieved in a particular time period will affect the quality of care at the same hospital in future time periods. Therefore, we adopt a dynamic approach to modeling hospital decisions and use a dynamic equilibrium model to describe the evolution of the entire hospital industry. In our model, each hospital engages in repeated rounds of the yardstick competition against all other hospitals and determines its best-response investment policy that maximizes the hospital's total discounted payoffs. We employ a novel application of structural estimation methods by clustering hospitals into a "quality ladder" (Pakes and McGuire, 1994) using a multidimensional notion of care quality based on measures used by the VBP Program and applying a hidden Markov chain framework to describe transitions within the ladder.

As a result, we recover key parameters that govern hospitals' decisions to invest in quality of care, including financial and non-financial investment costs, as well as parameters that shape stochastic investment outcomes. Furthermore, as a byproduct of the estimation process, we recover for every hospital the information on the quality level to which the hospital belonged throughout the implementation of the VBP Program. To answer our research questions, we use our estimates to perform counterfactual analyses and compare the efficacy of the current program against alternative designs of the quality-based yardstick competition. In particular, we find that increasing the size of the incentives from 2% to 4% would have resulted in an additional quality investment of US$1.2B from 2011 to 2018, which can lead to a 3.3% reduction in the average rate of central line-associated bloodstream infections (CLABSI) and a 1% increase in the average proportion of patients who rate their hospitals 9 or 10 on a 10-point scale. We also find that applying yardstick incentives to the tailored hospital peer groups, even without changing the size of the incentives, can lead to an average reduction of 1.4% in the rate of CLABSI among groups of hospitals associated with the highest costs of quality investment.

The rest of the paper is structured as follows. In Section 2.2, we review the related literature. In Section 2.3, we provide a detailed description of the VBP Program and the data we use. In Section 2.4, we introduce the structural model that serves as the basis for our estimation. The results of the estimation are presented in Section 2.5. In Section 2.6, we evaluate the alternative designs for the quality-based yardstick competition using counterfactual analyses. We summarize our findings in Section 2.7.

2.2. Related Literature

Yardstick competition was first introduced and analyzed by Shleifer (1985), which led to a stream of literature focusing on theoretical development (Dalen, 1998; Tangerås, 2002; Lefouili, 2015) as well as on examining applications across many industries including energy and electricity (Kuosmanen and Johnson, 2020), water (Sawkins, 1995; Cowan, 1997; Marques, 2006), railway (Mizutani, 1997), motor-vehicle inspections (Ylvinger, 1998), and port infrastructure (Estache et al., 2002). The industry that gained the most momentum in terms of both application and research has been the health care industry, especially the prospective payment system implemented by Medicare (Ellis and McGuire, 1986; Dranove, 1987; Dada and White, 1999).

More closely related to this work are studies that expand the original notion of cost-based yardstick competition to quality-based yardstick competition. Savva et al. (2019) and Tangeras (2009) explore modifying the cost-based yardstick competition in the health care industry to include a component that contracts on a quality indicator. The Hospital VBP Program analyzed in this paper is an example of a quality-based yardstick competition, and it has gained attention, especially in the medical literature. However, the existing literature on the impact of the VBP Program shows mixed findings. Ryan et al. (2017) and Lee et al. (2020) find that hospitals positively responded to the program by improving quality, whereas

Hong et al. (2020) find that there has not been any discernible impact on care processes or patient outcomes.

A similar program that targets readmission rates, the Hospital Readmission Reduction Program (HRRP), has gained traction in the operations management literature. Zhang et al. (2016) find that a significant number of hospitals are not incentivized by HRRP because the cost associated with reducing readmission is too high relative to the penalty imposed by the program. They also suggest localizing the benchmarking process, which we investigate in this paper by analyzing the impact of tailoring yardstick-based incentives to peer groups of similar hospitals. Arifoğlu et al. (2021) extend Zhang et al. (2016) by proposing an alternative payment scheme to overcome the shortcomings of the current design of HRRP. Chen and Savva (2018), Bastani et al. (2019), and Batt et al. (2020) explore unintended consequences and "spillover" effects of the program.

More broadly, this paper contributes to the operations management literature on incentive alignment in health care settings. In particular, payment policies, including bundled payment, have been shown to be effective in managing the incentives of different players in the health care industry, including providers (Aswani et al., 2019; Adida et al., 2017; Rajagopalan and Tong, 2021), hospitals (Andritsos and Tang, 2018; Guo et al., 2019), and pharmaceutical firms (So and Tang, 2000). Performance-based incentives, such as pay-for-performance systems, represent a more direct form of contracts that can be used to manage incentives. Lee and Zenios (2012) examine the pay-for-compliance system used for dialysis treatments, Jiang et al. (2012) analyze the impact of performance-based contracts on the allocation of outpatient care, and Adida and Bravo (2019) explore penalty contracts for referrals and outsourced care. This paper also adds to the stream of empirical studies focusing on quality-of-care in the hospital setting (KC and Terwiesch, 2011; Wang et al., 2019; Kuntz et al., 2019; Bartel et al., 2020).

Methodologically, this paper builds on a stream of literature on estimating dynamic choice models initiated by Rust (1987), which was extended to include various settings and to improve computational efficiency (Pesendorfer and Schmidt-Dengler, 2003; Bajari et al., 2007). We build on the quality ladder structure first proposed by Ericson and Pakes (1995) and Pakes and McGuire (1994), who model investment decisions in a Markov-perfect Nash equilibrium setting. Xu et al. (2018) find that the quality ladder model is well suited for analyzing industry response to regime changes. A key departure of this paper from the aforementioned work is that most of the previous studies assume that actions are observed, whereas in our setting actions are unobserved, and we must infer the investment decisions of hospitals as well as any cost associated with investment. Furthermore, we apply the structural estimation methods in a novel setting to investigate firm-level responses to a yardstick competition regime and examine the efficacies of alternative designs of yardstick competition.

2.3. Hospital VBP Program and Data

The Hospital VBP Program withholds 2% of all Medicare inpatient payments and redistributes them to hospitals as incentive payments based on the hospitals' performance on selected quality measures. Because any general acute care hospital with inpatient volume from Medicare patients is subject to the program, the VBP Program creates an induced yardstick competition that effectively spans the entire U.S. hospital industry.

An important feature of the program is that all of the quality measures it uses are also included in the Hospital Inpatient Quality Reporting Program, which incentivizes hospitals to disclose and report quality data to regulators as well as consumers. In our study, we use the data archives recording hospital performance in quality measures under the VBP Program from its first implementation in 2011 to the latest available data in 2018. The program has been using a total of 36 measures spanning multiple dimensions of quality, including clinical care process, outcomes, safety, patient satisfaction, and cost reductions.

The VBP Program's payments are awarded in the form of percentage point changes to Medicare reimbursement rates for inpatients in the fiscal year following the performance assessment. The rates are designated based on each hospital's composite score, which combines its performance across the tracked quality measures, and under a budget balancing constraint. We next describe the VBP Program in detail. At the end of the section, we provide descriptions of the datasets that provide us with operational and financial data.

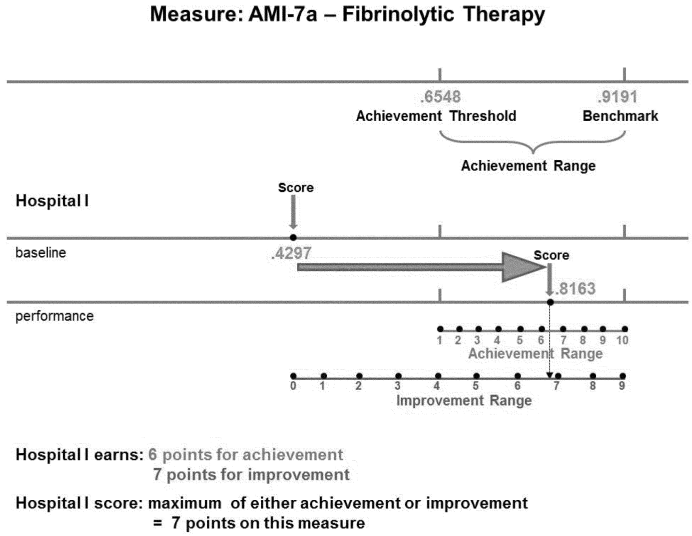
2.3.1. Quality Measures and Scores

Over the study period, between 20 and 26 quality measures were selected annually (and 36 over the entire period) for inclusion in the VBP Program's composite performance scores for hospitals. Measures used over this time period include, for example, the rates of hospital-acquired infections, the mortality and complication rates of specific patient groups, patient satisfaction survey results, and Medicare spending per beneficiary as a cost reduction proxy.

Timing of assessment. In converting a quality metric into a program score, two time periods are relevant. First, a hospital's current performance is tracked during the "performance period," which usually follows the current calendar year. In contrast, the program seeks to provide hospitals with a clear objective before each performance period begins, allowing each hospital to know how its performance during the upcoming calendar year will be evaluated. Therefore, the performance of a hospital during the performance period is compared with the past performances of other hospitals as well as the hospital's own historical performance during the "baseline period," which usually follows the calendar year two years prior to the performance period. The only exception to this design is the measure that tracks cost reductions; under this measure, the performance period doubles as the baseline period so that the respective dollar figures are compared directly without the use of any time discounting. Lastly, a hospital's scores from the performance period feed into the VBP Program's incentive payments applied in the following fiscal year.

Point-based score assessment. Let H denote the set of hospitals subject to the VBP Program. Each hospital $h \in H$ reports its measured performance in the jth quality metric tracked by the program as $q_{ht}^{(j)}$ in year t. To illustrate the process of assigning a numerical point score corresponding to the hospital's reported performance, Figure 2.1 provides an example using AMI-7a that tracks the percentage of patients with acute myocardial infarction (AMI) who received fibrinolytic therapy within 30 minutes of hospital arrival. In this

Figure 2.1: Example of hospital earning points under achievement or improvement



Note. Federal register. Vol. 76, No. 88. Friday, May 6, 2011. Department of Health and Human Services. Centers for Medicare & Medicaid Services.

example, 81.63% of AMI patients received therapy within 30 minutes at hospital h, i.e., $q_{ht}^{(j)} = 81.63\%$.

First, the hospital's performance is compared with the performances of peer hospitals, resulting in "achievement" points, $S_{ht}^{\mathsf{Achieve},j}$. To assign achievement points, an evaluation lower bound ("achievement threshold") $LB_{t-2}^{(j)} = 65.48\%$ is established at the median performance of all hospitals, and an evaluation upper bound ("benchmark") $UB_{t-2}^{(j)} = 91.91\%$ is established by the average performance of the top 10% of all hospitals. The "achievement range," which spans from the achievement threshold to the benchmark, is partitioned into equally sized achievement-point subranges from 1 to 10. The measured position of $q_{ht}^{(j)}$ is rounded to the nearest integer as its achievement points $S_{ht}^{\mathsf{Achieve},j}$. Performance below (above) the achievement range would round to 0 (10) points. In this case, the score rounds to 6 achievement points. Mathematically,

$$S_{ht}^{\mathsf{Achieve},j} = \begin{cases} 10, & \text{if } q_{ht}^{(j)} \geq UB_{t-2}^{(j)}, \\ 0, & \text{if } q_{ht}^{(j)} < LB_{t-2}^{(j)}, \\ \lfloor 9 \times \frac{q_{ht}^{(j)} - LB_{t-2}^{(j)}}{UB_{t-2}^{(j)} - LB_{t-2}^{(j)}} + 0.5 \rceil, & \text{otherwise}, \end{cases} \tag{2.1}$$

where $\lfloor\ \rceil$ is the rounding operator that rounds the bracketed value up to the nearest integer.

Second, the hospital's performance is compared with the hospital's own historical performance, resulting in "improvement" points, $S_{ht}^{\mathsf{Improve},j}$. Because the same hospital previously provided timely fibrinolytic therapy to 42.97% of AMI patients during the baseline period, an "improvement range" is constructed from the past metric, from $q_{h,t-2}^{(j)} = 42.97\%$ as the lower bound up to $UB_{t-2}^{(j)} = 91.91\%$. The improvement range is similarly partitioned into equi-distant point subranges from 0 to 9. Once again, $q_{ht}^{(j)}$'s position is rounded to the nearest integer—in this example, resulting in 7 improvement points.

The score $S_{ht}^{(j)}$ is the greater of the hospital's achievement points, $S_{ht}^{\mathsf{Achieve},j}$, and its improvement points, $S_{ht}^{\mathsf{Improve},j}$. Therefore, the hospital's final AMI-7a score

$$S_{ht}^{(j)} = \max\left(S_{ht}^{\mathsf{Achieve},j}, S_{ht}^{\mathsf{Improve},j}\right) = 7.$$

Scores for all measures are determined in the same manner, and hospitals receive points in each measure, $S_{ht}^{(j)} \in \{0, \ldots, 10\}$.

Quality metric domains. Each quality metric $q_{ht}^{(j)}$, and hence its associated score $S_{ht}^{(j)}$, belongs to one of five "domains": Clinical care process, Clinical care outcomes, Safety, Patient experience, and Efficiency. The VBP composite score's coverage of domains has varied by program year, ranging from two domains (Clinical care process and Patient experience) in its first performance year to four domains in 2020 (Clinical care outcomes, Safety, Patient experience, and Efficiency). The list of measures and domains used in VBP from fiscal year 2013 to fiscal year 2020 are presented in Tables A.1 and A.2 of Appendix A.1, respectively.

2.3.2. Composite Scores and Incentive Payments

A hospital's points in quality measures are first aggregated to the domain level and finally combined into a single composite score ranging from 0 to 100 based on pre-specified weights. Table A.3 of Appendix A.1 lists the domains and their weights used to calculate the composite scores. Let $\{S_{ht} \in [0, 100] : h \in H\}$ denote the hospitals' composite scores achieved under such weights in the performance year t.

Incentive payments. We characterize the resulting incentive payments. The incentive payments are applied to Medicare inpatient payments in the fiscal year following the end of the performance period. For example, the performance during the 2018 calendar year results in incentive payments during the 2020 fiscal year, which begins in October 2019 and ends in September 2020. During the nine-month gap between the end of the performance period and the beginning of the following fiscal year, the CMS collects and processes performance data from all hospitals and resolves any potential discrepancies in the data. Accordingly, let $R_{ht'}$ denote Medicare inpatient reimbursements charged by hospital h prior to any program adjustments or withholding in the fiscal year following t (we denote the fiscal year by t'). Finally, let $\bar{S}_t$ denote the reimbursement-weighted industry average composite score for performance year t:

$$\bar{S}_t := \frac{\sum_{h \in H} S_{ht} \times R_{ht'}}{\sum_{h \in H} R_{ht'}}. \tag{2.2}$$

Under the VBP Program, 2% of Medicare inpatient reimbursements are first withheld, supplying a total budget of $2\% \times \sum_{h \in H} R_{ht'}$ to be redistributed as incentives. A linear exchange function is calibrated with rate r_t such that a hospital with composite score S_{ht} will receive an incentive payment percentage of $r_t \times S_{ht}\%$ on its Medicare inpatient reimbursements in the fiscal year following the end of the performance period. The VBP Program is budget-neutral, implying that for budget balance, $\sum_{h \in H}(r_t \times S_{ht}\% - 2\%) \times R_{ht'} = 0$.

Therefore, with some algebra, the VBP Program delivers hospital h a net benefit of

$$2\% \times R_{ht'} \times \left(\frac{S_{ht}}{\bar{S}_t} - 1 \right) \tag{2.3}$$

in the next fiscal year t'. The benefit is positive if hospital h's score S_{ht} exceeds the reimbursement-weighted industry average $\bar{S}_t$ and is negative if it is less. No hospital can lose more than 2% of reimbursements under the VBP Program.

2.3.3. Financial and Operational Data

In addition to the data on hospital quality performance, we require operational and financial data on the hospitals to understand how the hospitals respond to the yardstick incentives. In this work, we combine the American Hospital Association (AHA) survey data, financial data from the Healthcare Cost Report Information System (HCRIS), and the Inpatient Utilization and Payment Public Use File (PUF). AHA survey data are collected through annual surveys on all member hospitals as well as any other voluntary hospitals. Being a national organization, the AHA is able to collect data that comprises of nearly all hospitals in the U.S. with detailed characteristics, including operating location, inpatient bed size, number of total inpatient and outpatient cases, as well as any special designations, such as being a teaching hospital. The HCRIS data include information on the revenues and costs of operation. Although the data include a more detailed breakdown of revenues and costs, we rely on total inpatient and outpatient revenues and total operating costs given that accounting policies may differ across hospitals. Furthermore, any costs associated with investment are likely grouped differently across hospitals. In most cases, they will be indistinguishable from the baseline cost of operation, essentially meaning that they are unobserved. Lastly, we use Inpatient PUF data, which include average payments that Medicare made to hospitals as well as the number of cases in each DRG. [1]

Using these three datasets, we compute the average revenue per case that hospitals receive

[1] Inpatient PUF data are accessible at https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Inpatient. AHA survey data and HCRIS data were accessed through Wharton Research Data Services.

from Medicare inpatient cases, non-Medicare inpatient cases, and outpatient cases. To compute the average revenue per case for Medicare inpatient and non-Medicare inpatient cases, we take the total inpatient revenues and first separate this value into Medicare and non-Medicare portions using the Inpatient PUF data. Then, we divide the total revenues by the number of cases in each type. Similarly, average revenue per case for an outpatient visit is calculated by dividing the total outpatient revenues by the total number of outpatient cases. In addition, we use the reported figures in the AHA survey and the HCRIS data to estimate, for each hospital, the number of cases for each patient type and the annual total cost of operation.

2.4. Model

In this section, we introduce the major components of our model that describe hospitals' best-response investment policies and the resulting equilibrium in the presence of yardstick incentives. The equilibrium best-response investment policies are used to estimate the model parameters that best maximize the likelihood of the observed hospital performance in quality measures and the observed financial data. We essentially combine two models. First, we introduce a quality ladder model that captures the dynamic evolution of hospital quality, in which hospitals are positioned at discrete quality levels that can be improved by investing to earn stochastic quality jumps. Thus, the transitions between quality levels are modeled as a choice-dependent Markov chain. Second, we model the hospitals' rewards under the VBP Program. Hospitals weigh the expected benefits of quality improvement in the form of future VBP incentive payments against the costs associated with investment in quality. The key feature of the model, and of yardstick competition, is that a hospital's expected rewards are determined not only by its own quality alone but also by the performance of peer hospitals in the quality metrics. The goal of the combined model is to capture how hospitals respond to yardstick competition through dynamic investment policies.

2.4.1. Quality Ladder

Hospitals. We consider a system of H hospitals, denoted $h = 1, \ldots, H$, that make investment decisions at times $t = 1, 2, \ldots, \infty$. We denote the fiscal years in which incentive payment percentages are applied as $t' = 1, 2, \ldots, \infty$ such that performance at time t results in incentive payments at time t'. The state space consists of five quality types, or ladder levels, denoted $\omega_{ht} \in \Omega$, that summarize the overall quality of the hospital h at time t. We choose five levels following the CMS's Five-Star Quality Rating System (Adelman, 2020), and although the ladder structure is largely an arbitrary simplification that provide us with tractability, estimating the emission distributions of actual performances in quality measures will allow us to transition back into characterizing the system in a tangible manner.

Quality measures. The performance of hospital h in quality measures $q_{ht}^{(j)}$ is indexed by $j \in J := \{1, \ldots, 36\}$. We use $\mathcal{J}_t \subset J$ to denote the subset of quality metrics tracked by the yardstick competition at time t. Given a current quality ladder level $\omega_{ht} \in \Omega$, hospital h emits an array of stochastic quality metrics $\vec{q}_{ht} := \left(q_{ht}^{(1)}, \ldots, q_{ht}^{(J)}\right) \in \mathbb{R}^J$. Performance in each quality measure $q_{ht}^{(j)}$ follows a distribution $G_\omega^{(j)}$ that depends on the hospital's quality level ω_{ht}. Based on considerations of tractability and fit, we specify $G_\omega^{(j)}$ to be a rectified normal distribution parameterized by its mean of $\mu_\omega^{(j)}$ and standard deviation of $\sigma_\omega^{(j)}$. We use rectified normal distributions because for most of the measures there are natural bounds in which performance is tracked, e.g., the upper bound of a quality measure that tracks the number of infections would be zero, and a measure that tracks the proportion of patients receiving appropriate care would be bounded between zero and one. Consistent with the notion of a vertical quality ladder, we impose monotonicity, in the sense of stochastic dominance, on each quality metric to increase up the quality ladder levels.

2.4.2. Transitions on the Quality Ladder

Quality investment. Hospitals individually transition between levels on the quality ladder through investment, denoted x_{ht}. First, we simplify investment to be a binary action, i.e., a hospital either invests or does not invest by comparing expected payoffs conditional on each action. We assume that investment incurs both financial and non-financial costs per patient, denoted c_h^1 and c_h^2, respectively, for hospital h. Financial costs include additional labor and capital needed to improve quality, such as instituting a team of analysts or hiring additional medical staff, and non-financial costs include institutional hurdles, such as the cost of devising new contracts or implementing new internal policies and initiatives that may not affect the financials of the hospital but end up being too costly for the hospital to invest in quality. Hospitals also incur baseline operating costs per patient, denoted c_h^0, which are independent of the investment decision. We assume that both investment costs and baseline operating costs per patient are heterogeneous across hospitals but are time invariant.

We assume that c_h^0 and c_h^1 are independently drawn across hospitals from a common bivariate lognormal distribution, i.e., $\log c_h^0$ and $\log c_h^1$ are bivariate normally distributed. We denote the parameters of the distribution of $(\log c_h^0, \log c_h^1)$ by $(\mu_{c^0}, \sigma_{c^0}^2, \mu_{c^1}, \sigma_{c^1}^2, \rho_c)$. Similarly, c_h^2 is independently drawn across hospitals from a common lognormal distribution, and we assume that $\log c_h^2$ is normally distributed with mean of μ_{c^2} and variance of $\sigma_{c^2}^2$. For brevity, we define $\boldsymbol{c}_h = (c_h^0, c_h^1, c_h^2)$, $\mu_{\boldsymbol{c}} = (\mu_{c^0}, \mu_{c^1}, \mu_{c^2})$, and $\sigma_{\boldsymbol{c}} = (\sigma_{c^0}, \sigma_{c^1}, \sigma_{c^2})$.

Transitions. The initial distribution of quality levels is denoted F_0. Investment, x_{ht}, affects the transition probabilities, denoted F_x, between quality levels such that investment in quality at time t increases the probability of the hospital moving up to a higher quality level and decreases the probability of moving down to a lower quality level at time $t+1$. Therefore, F_x essentially is a set of two 5×5 transition matrices given the investment decision with $F_{x=investment}$ stochastically dominating $F_{x=no\ investment}$.

We make two assumptions for computational and identification purposes. First, we assume that hospitals have a positive probability of moving up to a higher quality level only if they invest in quality, i.e., quality improvements can only occur if hospitals actively invest in quality. Second, we assume that hospitals only have positive probabilities to transition to the adjacent quality levels. This assumption implies that changes in the quality of a hospital occur gradually, e.g., a hospital with mediocre quality cannot suddenly become one of the top quality hospitals even if investment is made. If the current level is either the lowest or the highest, hospitals cannot move outside the five quality ladder levels. These assumptions provide computational benefits by significantly reducing the number of transition parameters we must estimate.

2.4.3. Program Incentives

Calculating incentive rewards. Under the VBP Program, a hospital's realized performance signals $\vec{q}_{ht}$ are evaluated against industry performance benchmarks to yield its composite score S_{ht} based on achievement points and improvement points. By applying the incentives disbursal formula (2.3) to the body of scores $\{S_{ht} : h \in H\}$, we obtain the incentive payments the program delivers to each hospital.

For reasons of tractability, our model simplifies the composite scores in the following way. Under the VBP convention, a hospital's scores in each metric $j \in \mathcal{J}_t$,

$$S_{ht}^{(j)} = \max\left(S_{ht}^{\mathsf{Achieve},j}, S_{ht}^{\mathsf{Improve},j}\right),$$

are aggregated to the domain level and then averaged under pre-determined weights to arrive at the hospital's composite score S_{ht}. Under our simplification, we ignore the improvement points $S_{ht}^{\mathsf{Improve},j}$—instead, we handle each metric j's achievement points $S_{ht}^{\mathsf{Achieve},j}$ as though they were the metric scores $S_{ht}^{(j)}$. We then follow the VBP convention in aggregating to the domain level and calculating the weighted average $\widetilde{S}_{ht}$ as the hospital's alternative composite score. By construction, $\widetilde{S}_{ht}$ is biased downward in comparison to the true composite score S_{ht}; hence, we correct for the bias by regressing S_{ht} on $\widetilde{S}_{ht}$:

$$S_{ht} = \alpha + \beta \times \widetilde{S}_{ht} + \epsilon_{ht}, \tag{2.4}$$

and then using the estimated coefficients $\widehat{\alpha}$ and $\widehat{\beta}$ to project the bias-corrected composite scores $\widehat{S}_{ht} = \widehat{\alpha} + \widehat{\beta} \times \widetilde{S}_{ht}$. In short, we construct the composite scores $\widetilde{S}_{ht}$ using the achievement points instead of the maximum of the achievement and improvement points in each area j

Table 2.1: Regression of observed composite scores on simplified composite scores only based on achievement points

	(1)
Simplified composite score calculated using achievement points	0.859^{***}
	(0.00366)
Constant	10.36^{***}
	(0.108)
Observations	11278
R^2	0.875

Note. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

and then bias-correct to arrive at $\widehat{S}_{ht}$. The bias-correcting regression is presented in Table 2.1.

Through this simplification, we abstract away from tracking the past performance in all measures $\mathcal{J}_t$ for each individual hospital. Conceptually, we therefore focus our model on the incentives generated by industry benchmarking, which is the central mechanism of interest in yardstick competition. Empirically, we verify that $\widehat{S}_{ht}$ accurately approximates the original composite score S_{ht}, capturing 87.5% of its variation. Thus, the modeled incentive payments are derived by applying the incentives disbursal formula (2.3) to the body of scores $\left\{\widehat{S}_{ht} : h \in H\right\}$.

2.4.4. Expected payoffs.

Hospital h's expected payoff at time t is

$$E[\pi(\omega_{ht}, x_{ht})] = \underbrace{(1 - a_t)p_{ht}^M \lambda_{ht}^M + p_{ht}^{MP} \lambda_{ht}^M + p_{ht}^C \lambda_{ht}^C + p_{ht}^O \lambda_{ht}^O}_{\text{Baseline revenues}} + \underbrace{\beta^2 E[b_{ht'}|\omega_{ht}] p_{ht'}^M \lambda_{ht'}^M}_{\text{Incentive payments}}$$
$$- \underbrace{(c_h^0 + (c_h^1 + c_h^2)\mathbf{1}(x_{ht} = \mathit{investment}))\lambda_{ht}^{adj}}_{\text{Total costs}}. \quad (2.5)$$

Baseline revenues. We assume that hospitals receive three different types of patient flows: Medicare inpatient, non-Medicare inpatient, and outpatient. The throughput for each category of patients is denoted λ_{ht}^M, λ_{ht}^C, and λ_{ht}^O, respectively, and they generate marginal revenues denoted p_{ht}^M, p_{ht}^C, and p_{ht}^O, respectively. We separate out out-of-pocket payments for

Medicare inpatient cases, denoted p_{ht}^{MP}, as these are not subject to payment withholdings for the VBP Program. Note that the marginal revenues are indexed by both h and t, and the hospital and time heterogeneity in reimbursement rates comes from differences in case mix, geographic and market-level factors, and hospital-specific factors, such as being a teaching hospital. Lastly, a_t is the percentage of Medicare inpatient payments that are withheld for the VBP Program.

Incentive payments. $b_{ht'}$ is the incentive payment percentage determined by the quality level ω_{ht}. It is applied to revenues from Medicare inpatient cases during the following fiscal year, denoted t'. We assume a common discount factor of $\beta = 0.97$, and given the 21-month lag between the beginning of the performance period, i.e., when investment decisions are made, and the following fiscal year, we simply assume that incentive payments are discounted by β^2.

Total costs. We introduce λ_{ht}^{adj} to denote adjusted number of patients, which is a commonly used measure to represent the overall workload of a hospital. It essentially is an inpatient-equivalent volume of total cases, and it is calculated by taking the total number of inpatient cases then adding the number of outpatient cases weighted by the proportion of revenue from inpatient versus outpatient sources. In essence, the adjusted number of patients provides us with a single measure of throughput on which we apply marginal costs. c_h^0 is the baseline operating costs per patient, c_h^1 and c_h^2 are the costs of investment, and x_{ht} is hospital h's investment decision at time t.

2.4.5. Dynamic Investment Policies

Bellman solution. Given the expected payoffs conditional on quality levels and investment decisions at time t, hospitals solve an infinite-horizon Markov decision process for the optimal investment policy by maximizing the expected payoffs. The approximation of the expected incentive payments that restricts the dependency to the current quality position allows us to impose a Markovian structure where the current quality and the current investment decision provide sufficient information on the dynamics of the trajectory of quality levels. Therefore, the optimal policy of Markov strategy can be recovered by recursively solving for the Bellman equation

$$V_{ht}(\omega_{ht}) = \max_{x_{ht}} E[\pi(\omega_{ht}, x_{ht})] + \beta \sum_{\Omega} P(\omega_{h,t+1}|\omega_{ht}, x_{ht}) V_{h,t+1}(\omega_{h,t+1}). \tag{2.6}$$

Informational assumptions. We make the following assumptions about the model and the information that is available to the hospitals. We assume that the trajectory of number of patients served is deterministic and known. Our rationale is that it is reasonable to assume that patient demand is not significantly influenced by the quality scores given the fact that one of the purposes of a yardstick competition is to impose an induced competition to mitigate monopolistic behaviors. We empirically test this assumption by regressing hospital throughput on composite scores using a fixed-effects framework. The results are presented in Appendix A.2, and we do not find any distinguishable pattern that may suggest that hospital throughput is positively correlated with performances in quality measures that we observe through the VBP Program. Similarly, we assume that the reimbursement levels are deterministic and known. In other words, we do not consider the possibility that hospitals have increased bargaining powers as a result of increases in quality scores.

Structural estimation. The key challenge in the estimation process is that we do not directly observe hospitals' quality levels, their investment decisions, and the breakdown of total operating costs into the baseline operating costs and costs associated with quality investment. Using the expectation-maximization algorithm and a hidden Markov chain framework, we iteratively find the maximum likelihood estimates of the parameters of the common distribution of operating and investment costs $(\boldsymbol{\mu_c}, \boldsymbol{\sigma_c}, \rho_c)$, the parameters of the emission distribution of performances in quality measures $(\mu^j_\omega, \sigma^j_\omega)$, the transition matrices conditional on investment decision F_x, and the initial distribution of quality levels F_0. As a byproduct of the estimation process, we also recover the quality levels each hospital belonged throughout the implementation of the VBP Program as well as each hospital's baseline operating costs and investment costs. We provide details of the estimation process in Appendix A.3.

2.5. Estimation Results

In this section, we report the results of the structural estimation. We first present the estimates, for each hospital, of the baseline operating costs and investment costs. Given the estimated cost values, we evaluate goodness-of-fit by comparing the point estimates of the total costs with the observed total costs for each hospital. Furthermore, we analyze how the cost of quality investment is correlated with any observed hospital characteristics, such as hospital size and type. Next, we present parameters related to the quality levels of the hospitals: the distribution of hospitals over the quality ladder, transition matrices conditional on investment decisions, and emission parameters of performances in the quality measures used in the VBP Program. While the estimated parameters and cost values are interesting in their own right, they also serve as the building blocks of the counterfactual studies that we describe in Section 2.6, where we evaluate alternative designs of the yardstick competition.

2.5.1. Parameters and Estimated Values Related to Operating and Investment Costs

Estimated parameters. We summarize the estimates of the parameters in Table 2.2. The estimated parameters of the common log-normal distribution of hospitals' baseline costs per adjusted patient are 9.34 and 0.30. In other words, the logged values of the baseline costs per adjusted patient for all hospitals follow a normal distribution with mean of 9.34 and standard deviation of 0.30, which translates to an average of $11,908 per patient. Similarly, the parameters of the log-normal distribution of the financial costs of quality investment are 0.78 and 2.36, which translates to an average of $35 per patient. The parameters of the log-normal distribution of the non-financial costs of quality investment are -4.55 and 0.03, and the estimated correlation factor between the baseline costs per adjusted patient and financial costs of investment is 0.28.

Empirical distribution. In addition to estimating the parameters of the common distributions, our estimation process allows us to recover the cost values of each hospital with

Table 2.2: Estimated cost parameters and empirical distribution

	Estimated parameters				*Empirical distribution*	
	(1)		(2)		(3)	
	Per patient, logged		Per patient, in $		Annual, in $	
Baseline operating costs	9.34	(0.30)	11,908	(3,654)	314,140,000	(352,350,000)
Financial costs of investment	0.78	(2.36)	35.33	(571.2)	281,680	(462,670)
Non-financial costs of investment	-4.55	(0.03)	0.011	(0.00032)	258	(205)

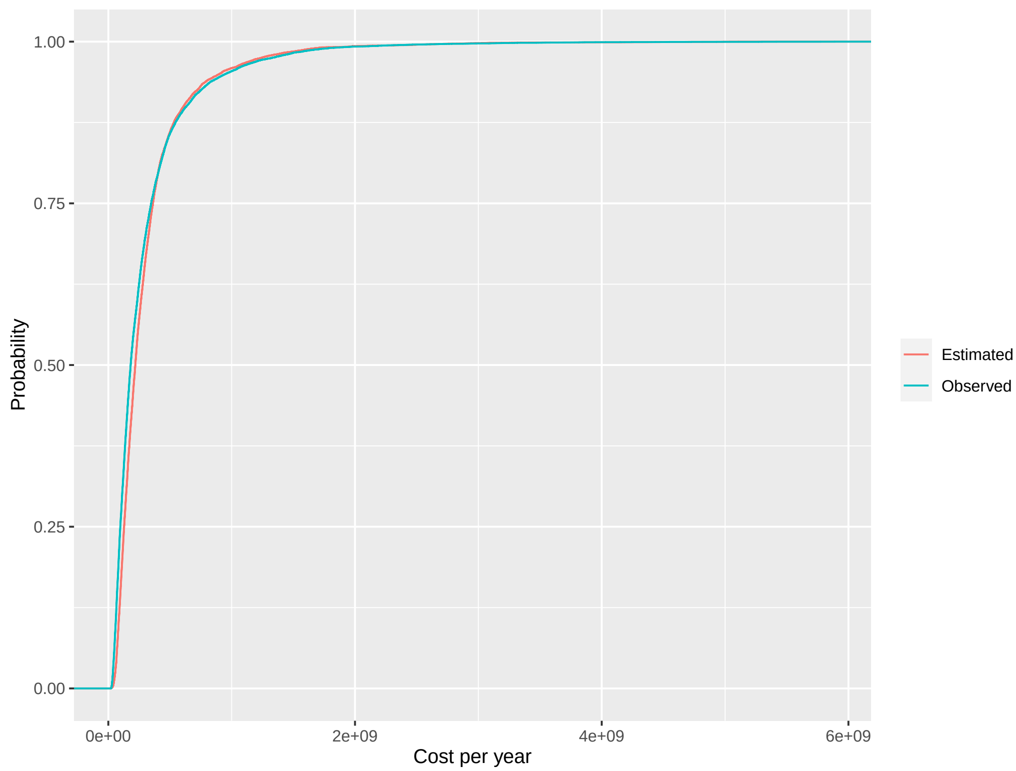
Note. Standard deviations in parentheses. (1) presents the estimated parameters. (2) presents the arithmetic mean and standard deviation given by the log-normal distribution. (3) presents the mean and standard deviation of the empirical distribution.

the highest likelihood. In other words, while we assume that the cost values of each hospital are drawn from the distribution with parameters described above, we can pinpoint which draw can most likely explain the observed data. We refer to the set of cost values with highest likelihood for each hospital as the "empirical distribution" of cost values. We find that the mean and standard deviation of the estimated annual baseline costs are $314,140,000 and $352,350,000, respectively. The mean and standard deviation of the financial costs of investment per year are $281,680 and $462,670, respectively, and the mean and standard deviation of the non-financial costs of investment per year are $258 and $205, respectively.

As expected from the estimated cost parameters, we find that the cost values are long-tailed, in line with an often-cited characteristic of the U.S. hospital industry (National Center for Health Statistics, 2017). Furthermore, both financial and non-financial costs associated with investment in quality exhibit large variance across hospitals, suggesting that the same yardstick incentives applied to the entire hospital industry may not be the most effective method for inducing competition among hospitals. Lastly, we find that non-financial costs of investment are dominated by financial costs, which suggests that hospitals do not face institutional burdens of investing in quality that are not passed through as financial costs.

Goodness-of-fit. Using the estimated cost values for each hospital, we test goodness-of-fit of the estimated cost values by comparing the estimated total costs that each hospital incurred during a given year, including any financial costs of investment, with the observed total costs that each hospital incurred during the same year. We use the estimated ladder-

Figure 2.2: Cumulative distribution functions of estimated and observed total costs



based quality trajectory and the optimal investment policy to identify whether a given hospital made investments toward quality improvement. We discuss the estimated quality trajectory in more detail in Section 2.5.2. We present the cumulative distribution functions of the estimated total costs compared with the observed total costs in Figure 2.2. Although the estimation slightly underestimates the proportion of hospitals in the lower end of the cost spectrum, we find that our estimation process is able to predict the observed costs with rather high accuracy.

Cost of investment by hospital characteristics. Next, we further delve into investment costs and perform analyses to identify any pattern present in the distribution of investment costs. To do so, we regress the cost of investment per adjusted patient on various hospital characteristics. We present the results in Table 2.3. First, we look into whether the size of a hospital has any correlation with the cost of investment. *A priori* it is unclear whether a larger hospital would need to incur higher or lower cost in order to improve its

Table 2.3: Hospital characteristics and cost of quality investment

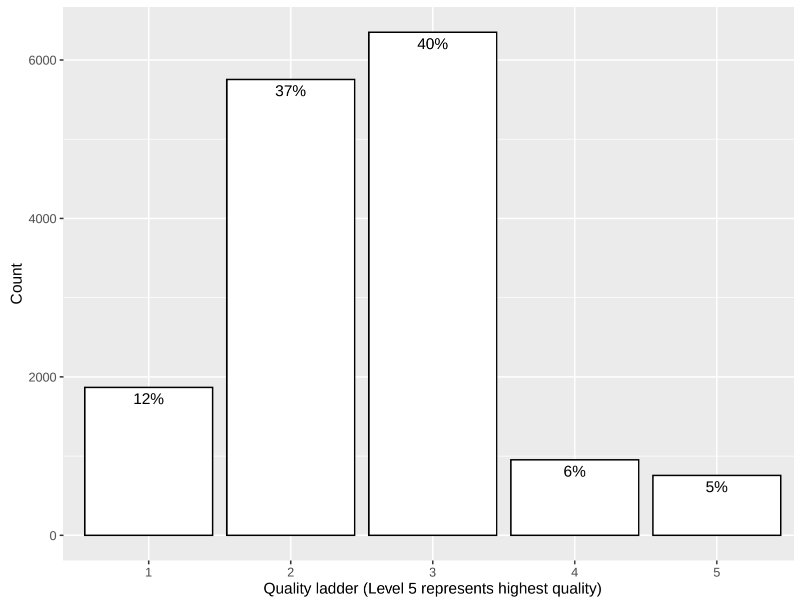
	(1) Investment cost, in $	
Bed size type		
49 or less beds	(.)	(.)
50–99 beds	-0.479	(0.837)
100–199 beds	1.397	(0.846)
200–299 beds	4.317***	(0.951)
300–399 beds	4.938***	(1.011)
400–499 beds	7.089***	(1.490)
500 or more beds	7.056***	(1.278)
Teaching status		
Not a teaching hospital	(.)	(.)
Teaching hospital	4.604***	(1.180)
CBSA type		
Metro	(.)	(.)
Micro	-1.904***	(0.438)
Rural	-1.915*	(0.769)
Control type		
For-profit	(.)	(.)
Government	-5.184***	(0.754)
Nonprofit	-3.642***	(0.609)
Constant	9.872***	(0.944)
Observations	1960	

Note. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. (.) indicates variable dropped due to collinearity.

quality. On the one hand, larger hospitals may have lower costs of investment per patient due to economies of scale. In other words, the cost of inputs necessary for quality improvement, such as a policy initiative or an additional team of analysts, could be spread among a bigger patient base, reducing the cost of investment per patient. On the other hand, since larger hospitals are more likely to be associated with a more complex internal structure, such hospitals may need to incur higher costs to improve their quality. Using the estimated cost values, we find that hospitals with a larger number of inpatient beds are associated with higher cost of investment. In particular, we find that medium hospitals with 200–299 beds are associated with higher cost of investment per adjusted patient by $4 compared to small hospitals. Hospitals with 400 or more beds are associated with higher cost of investment by $7 compared to small hospitals and $3 compared to medium hospitals.

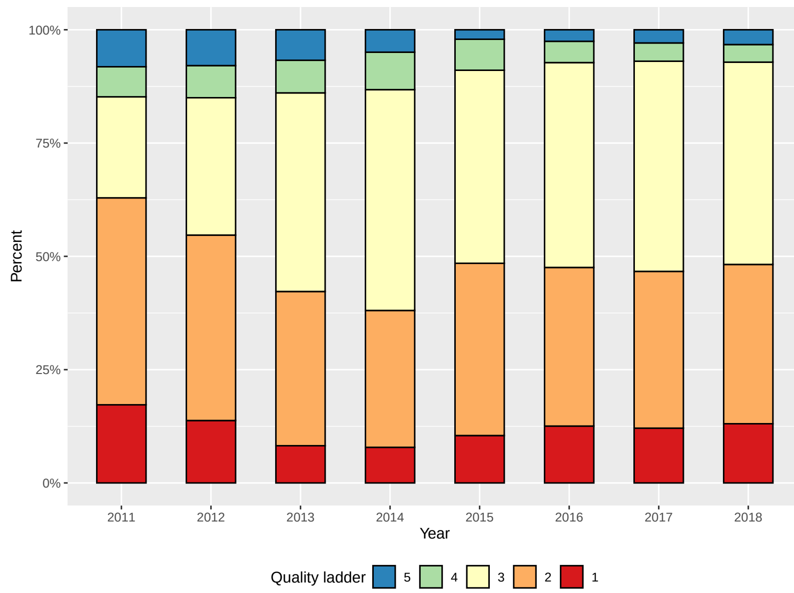
Aligned with the idea that hospitals with more complex internal structures must incur

Figure 2.3: Estimated distribution of hospitals within the quality ladder from 2011 to 2018



additional cost of investment, we find that teaching hospitals tend to have higher cost of investment compared to non-teaching hospitals. We also find that the location in which a hospital operates has an effect on the cost of quality investment. We test this using the core-based statistical area (CBSA) in which each hospital belongs. Areas under CBSA are grouped into three categories: metropolitan, micropolitan, and rural. We find that hospitals that operate in metropolitan areas incur higher cost of investment compared to hospitals that operate in micropolitan and rural areas. Finally, we find that for-profit hospitals tend to have the highest investment costs and government hospitals tend to have the lowest investment costs, with nonprofit hospitals in the middle. We use the result of this exercise as motivation for the counterfactual analyses, discussed in Section 2.6, where we test the efficacy of yardstick competitions with tailored peer groups in which hospitals only compete with other hospitals of the same type. The rationale is that since different types of hospitals have different underlying distribution of investment costs, the yardstick competition should be more inducive to quality improvement when the competition occurs among hospitals with a similar level of investment costs.

Figure 2.4: Changes in the distribution within the quality ladder



2.5.2. Quality Ladder

Similar to the cost of investment, we also recover the most likely path of quality levels for each hospital. We present the distribution of hospitals in each quality level in Figure 2.3. We find that a large portion of hospitals (40%) are in the middle level. 37% of hospitals are in the second bottom level, indicating that the majority of the hospitals fall somewhere in the middle or the lower level. 12% of hospitals are in the lowest level, and 5% and 6% of hospitals are in the highest and the second highest level, respectively. This finding is in line with the intuition on the hospital industry that the majority of the hospitals have medium quality and there are hospitals with lower quality and a small set of hospitals with exceptional quality.

Next, we examine the changes in the distribution within the quality ladder from the beginning of the implementation of the VBP Program until 2018. Figure 2.4 highlights the dynamic nature of quality improvements and how hospitals responded to the yardstick competition. First, we find that the distribution of hospitals in each quality level rapidly changed from 2011 to 2014. Beginning in 2015, the distribution settles down, suggesting that hospitals first actively responded to the yardstick incentives until an equilibrium was reached in 2015. Second, the percentage of hospitals in the top two quality ladder levels actually decreased over the implementation of the VBP Program. This finding is in line with the intuition behind the yardstick competition in which a hospital does not necessarily have to be the absolute best hospital, but it just has to be above enough hospitals to be rewarded incentive payments. In contrast, we find that hospitals in the second lowest level were most incentivized by the program, which again intuitively makes sense because a hospital needs to beat the median performance of all peer hospitals in order to start earning incentive payments. The hospitals in the lowest level seem to respond at first to the yardstick competition and improve their quality. However, as they discover that the distribution of quality also increased, primarily driven by the second lowest group, the hospitals eventually lose interest, and we find that the percentage of hospitals in the lowest group rapidly drops for the first three years or so and then gradually increases again. Overall, the yardstick competition resulted in more hospitals converging toward the middle, resulting in decreases in the percentage of hospitals in both the higher and lower quality groups.

In terms of mobility within the quality ladder, we find that only 16% of all hospitals stayed in the same quality level over the entire study period. In total, 63% of all hospitals moved up a quality level at least once, and 58% moved down a quality level at least once. In terms of range of movement between quality levels, we find that 69% of all hospitals stayed within a single level or a nearest level, 24% of hospitals were part of three quality levels during the study period, and the remaining 7% were part of more than four quality levels. This suggests that hospitals were able to improve their quality, or let their quality degrade, as measured by the quality measures included in the Hospital VBP Program. Furthermore, this again provides evidence that a large number of hospitals did in fact respond to the yardstick incentives created by the Hospital VBP Program.

2.5.3. Emission Distributions of Performances in Quality Measures

Next, we present the parameters of the distributions of performances for hospitals in each quality level. Table 2.4 presents the means and standard deviations for each quality measure used in the Hospital VBP Program by quality levels. The estimated parameters show that the distributions of performance scores across quality levels for most measures are significantly different from each other, suggesting that the hospitals can be considered clustered by the quality ladder structure as we have assumed. Furthermore, the differences in the distributions across quality levels imply that hospitals have ample incentive to invest in quality, i.e., to move up a quality ladder to receive higher performance scores, which in turn are translated into higher incentive payments.

Table 2.4: Emission distributions of performance scores by quality ladder

Measure ID	Quality level 1 (lowest)		Quality level 2		Quality level 3		Quality level 4		Quality level 5 (highest)	
AMI-7a	0.000	(0.000)	0.754	(0.784)	0.754	(0.780)	1.000	(0.045)	1.000	(0.000)
AMI-8a	0.905	(0.132)	0.960	(0.097)	0.984	(0.061)	0.984	(0.091)	1.000	(0.000)
COMP-HIP-KNEE	-0.026	(0.007)	-0.026	(0.005)	-0.026	(0.005)	-0.026	(0.007)	-0.024	(0.004)
HAI-1	-0.726	(0.869)	-0.517	(0.566)	-0.458	(0.589)	-0.271	(1.364)	-0.271	(0.614)
HAI-2	-0.720	(0.823)	-0.720	(0.692)	-0.720	(0.784)	-0.360	(1.779)	-0.360	(0.896)
HAI-3	-0.841	(1.061)	-0.834	(0.827)	-0.834	(0.804)	-0.709	(1.499)	-0.489	(0.796)
HAI-4	-0.738	(1.148)	-0.738	(1.162)	-0.738	(1.014)	-0.738	(0.888)	-0.738	(0.953)
HAI-5	-1.090	(1.124)	-0.879	(0.810)	-0.818	(0.821)	-0.818	(1.567)	-0.518	(1.160)
HAI-6	-0.820	(0.462)	-0.820	(0.377)	-0.793	(0.426)	-0.793	(0.923)	-0.793	(0.500)
HCAHPS-CD	0.744	(0.033)	0.780	(0.025)	0.814	(0.025)	0.834	(0.034)	0.850	(0.025)
HCAHPS-CM	0.570	(0.038)	0.612	(0.027)	0.653	(0.029)	0.673	(0.043)	0.702	(0.033)
HCAHPS-CN	0.719	(0.034)	0.768	(0.021)	0.809	(0.022)	0.819	(0.028)	0.850	(0.020)
HCAHPS-CQ	0.585	(0.049)	0.623	(0.043)	0.676	(0.043)	0.708	(0.043)	0.732	(0.043)
HCAHPS-CT	0.432	(0.038)	0.490	(0.029)	0.537	(0.031)	0.537	(0.048)	0.595	(0.033)
HCAHPS-DI	0.825	(0.033)	0.861	(0.023)	0.882	(0.021)	0.882	(0.029)	0.903	(0.018)
HCAHPS-OVR	0.594	(0.052)	0.679	(0.040)	0.739	(0.043)	0.739	(0.060)	0.801	(0.041)
HCAHPS-PM	0.636	(0.037)	0.681	(0.022)	0.719	(0.024)	0.719	(0.040)	0.755	(0.027)
HCAHPS-RS	0.567	(0.051)	0.621	(0.038)	0.680	(0.042)	0.717	(0.050)	0.748	(0.045)
HF-1	0.892	(0.138)	0.935	(0.081)	0.960	(0.040)	0.960	(0.144)	1.014	(0.058)
IMM-2	0.913	(0.103)	0.943	(0.063)	0.961	(0.035)	0.961	(0.073)	0.992	(0.013)
MORT-30-AMI	0.859	(0.012)	0.859	(0.013)	0.859	(0.013)	0.859	(0.008)	0.859	(0.010)
MORT-30-HF	0.880	(0.013)	0.880	(0.015)	0.880	(0.015)	0.880	(0.011)	0.881	(0.011)
MORT-30-PN	0.886	(0.015)	0.886	(0.016)	0.886	(0.016)	0.886	(0.015)	0.889	(0.013)
MSPB-1	-1.016	(0.076)	-1.001	(0.060)	-0.987	(0.059)	-0.966	(0.077)	-0.958	(0.063)
PC-01	-0.001	(0.055)	-0.001	(0.050)	-0.001	(0.050)	0.000	(0.000)	0.000	(0.000)
PN-3b	0.964	(0.042)	0.979	(0.027)	0.984	(0.017)	0.984	(0.047)	1.009	(0.023)
PN-6	0.941	(0.061)	0.960	(0.036)	0.960	(0.027)	0.960	(0.074)	1.012	(0.037)
PSI-90	-0.599	(0.164)	-0.571	(0.146)	-0.538	(0.137)	-0.525	(0.101)	-0.506	(0.104)
SCIP-Card-2	0.927	(0.081)	0.971	(0.029)	0.991	(0.014)	1.008	(0.158)	1.026	(0.046)
SCIP-Inf-1	0.974	(0.036)	0.987	(0.014)	0.995	(0.007)	1.005	(0.102)	1.005	(0.018)
SCIP-Inf-2	0.969	(0.036)	0.987	(0.012)	0.994	(0.006)	1.000	(0.088)	1.004	(0.016)
SCIP-Inf-3	0.949	(0.045)	0.976	(0.017)	0.984	(0.010)	0.984	(0.079)	1.000	(0.020)
SCIP-Inf-4	0.935	(0.056)	0.958	(0.040)	0.978	(0.026)	0.980	(0.081)	0.997	(0.032)
SCIP-Inf-9	0.923	(0.091)	0.961	(0.035)	0.990	(0.014)	1.008	(0.140)	1.008	(0.023)
SCIP-VTE-1	0.960	(0.039)	0.983	(0.013)	0.996	(0.006)	0.998	(0.154)	1.007	(0.015)
SCIP-VTE-2	0.961	(0.049)	0.981	(0.021)	0.996	(0.010)	1.003	(0.137)	1.003	(0.015)

Note. Standard deviations in parentheses.

Note that the distributional assumption that the performance scores follow rectified normal distributions implies that some of the estimated means will be above the upper bound of a measure. E.g., quality level 5's mean for HF-1, which measures the proportion of heart failure patients given appropriate discharge instructions, is 1.014. This implies that the hospitals in the highest quality level on average have latent quality with a mean above the upper bound, and the latent performances are censored by the upper bound. The quality measures with negative mean indicate that the performance in the measure is scored in a way that a lower absolute value of the score indicates better performance. We simply multiply negative one to these measures so that a higher number, whether negative or positive, indicates better performance. For example, the highest quality level's expected score on HAI-1, which measures the rate of CLABSI acquired within the hospital, is 0.271, whereas the expected score on the same measure for the lowest quality level is 0.726. We therefore both report and treat them in our analyses as negative numbers.

2.5.4. Transition Matrices

Lastly, we present the estimated investment-dependent transition matrices. As discussed in Section 2.4.2, we assume that the quality of a hospital follows an investment-dependent Markov process where the state consists of five quality levels. Table 2.5 presents the transition matrices governing the Markov process. The results show that hospitals in the lower spot on the ladder tend to have smaller probability of the quality decreasing when investment is not made. For instance, the probability of moving down from the second level (0.12) is much lower than the probability of moving down from the fourth level (0.45) or the top level (0.39). In contrast, investment is more likely to result in improvement in quality when hospitals are in a lower spot on the ladder. For instance, the probability of moving up from the first level (0.48) is much higher than the probability of moving up from the fourth level to reach the top level (0.21). This result is coherent with the intuition that improving quality becomes increasingly difficult for hospitals already with high quality. In other words, from the perspective of a hospital with relatively low quality, investment in quality may result in a much more certain improvement because the hospitals can choose to tackle "low-hanging

Table 2.5: Investment-dependent transition matrices

No investment	1	2	3	4	5
1	1.00	0	0	0	0
2	0.12	0.88	0	0	0
3	0	0.17	0.83	0	0
4	0	0	0.45	0.55	0
5	0	0	0	0.39	0.61

Investment	1	2	3	4	5
1	0.52	0.48	0	0	0
2	0.07	0.65	0.28	0	0
3	0	0.12	0.78	0.10	0
4	0	0	0.40	0.40	0.21
5	0	0	0	0.28	0.72

Note. Each box represents the probability of transitioning from the quality level on the left to that on the top. Level 5 represents the group with the highest quality.

fruits" and implement policies to improve the quality of care to a certain level. However, once those options are exhausted, future investment in quality will need to be more subtle and strategic.

2.6. Implications for Alternative Designs of the Yardstick Competition

The structural estimation method employed in this work allows us to not only recover unobserved parameters around cost of quality improvement and outcomes of investment but also perform counterfactual analyses to simulate different designs of the yardstick competition. The benefit of performing counterfactual analyses is that we can evaluate the current policy against alternative policies and ultimately have the ability to prescribe a policy based on the goal of the firms or government agencies implementing yardstick competitions.

First, we examine the responses of the hospitals to changes in the size of the yardstick incentives. The current program redistributes 2% of all Medicare payments, and we simulate the responses of the hospitals under a range of schemes that redistributes 1% of payments all the way up to 4% of payments in 0.25% increment. We find that the industry overall invests more heavily in quality improvements as the size of the yardstick incentives increases. However, the marginal gain from investment in terms of quality improvements decreases as the total cost of investment that the entire industry incurs increases.

Second, we simulate tailored policies where hospitals compete for yardstick incentives against other hospitals of similar type. Motivated by the finding that the distribution of cost of investment varies by a significant amount across hospitals with different sizes, locations, and types, we simulate two yardstick competitions with tailored peer groups: (1) *size-based* competition where hospitals compete with other hospitals with similar bed size and (2) *type-based* competition where hospitals compete with other hospitals that operate in a similar setting.

Counterfactual equilibrium Before we present the results of the counterfactual studies, we describe the method in which we obtain counterfactual results. Obtaining the equilibrium under a counterfactual policy requires that the investment decisions, and subsequently the trajectory of quality levels and performances in quality measures, are self-consistent with the environment in which each hospital competes. In other words, while we assume that hospitals take into consideration not each and every other hospital participating in the yardstick competition but rather the distribution of all other hospitals, the performances in quality measures and the composite scores that each hospital receives must be able to regenerate the industry-level distribution.

We compute the counterfactual equilibrium through an iterative process, where we first obtain optimal dynamic investment policies given observed distributions of performances in quality measures and composite scores. Using the initial quality levels identified through the estimation process, we forward simulate the trajectory within the quality ladder for subsequent periods using the dynamic investment policies and investment-dependent transition matrices. We then draw performances in each quality measure for each hospital using the estimated emission distributions given the sample path of quality levels for the hospital. In turn, we use the performances in quality measures to compute the composite scores that each hospital receives. In the next iteration, we update the industry-level distributions of performances in quality measures and composite scores by combining all hospitals' draws of performances in quality measures and composite scores. We repeat the first steps of obtaining dynamic investment policies, forward simulating the trajectory within the quality ladder, and drawing performances in quality measures and composite scores. We iterate this process until the distributions of performances in quality measures and composite scores converge to stationary points, i.e., updated distributions no longer deviate from previous iterations. We discuss the details of the convergence criteria in Appendix A.4.

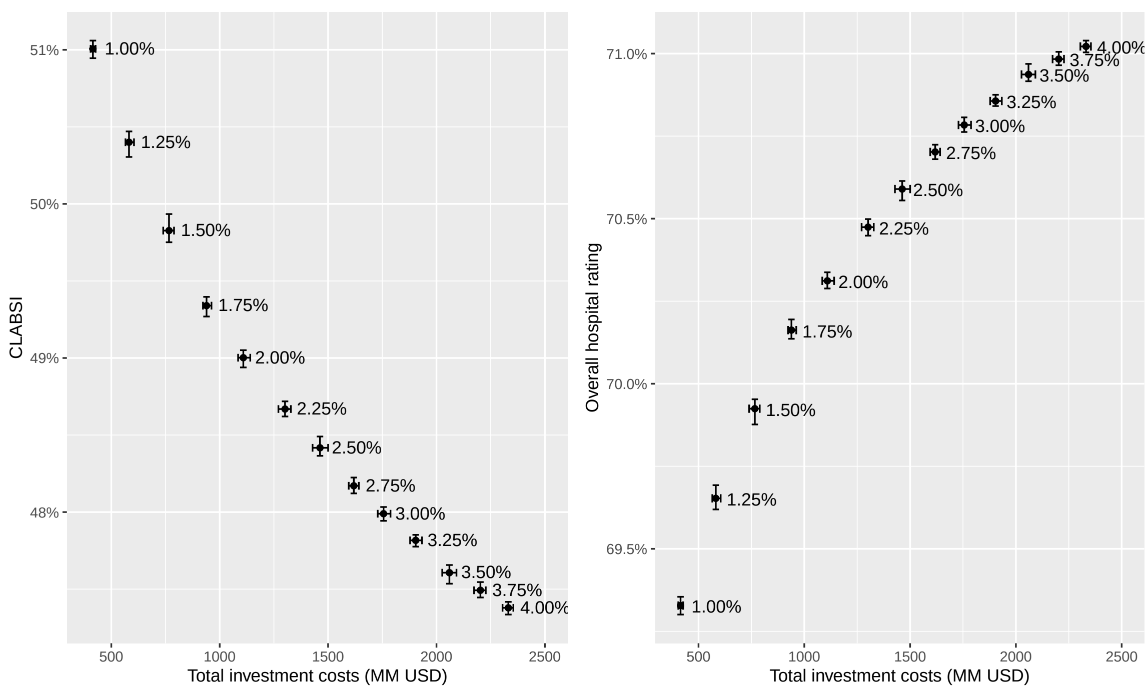
2.6.1. Changing the Size of Yardstick Incentives

One of the most salient aspects of the yardstick competition from the perspective of the hospitals that are subject to the program is the size of the incentive payments. For the case of the VBP Program, the size of the incentive payments is directly related to the percentage of Medicare payments withheld and redistributed. In the first set of counterfactual analyses, we examine the effect of changes in the percentage of Medicare payments that are withheld. Conceptually, hospitals overall would respond to higher incentive payments by investing in quality more frequently. On the other hand, if the size of the incentive payments is too small, then more and more hospitals will find the incentive payments to be not worth the cost of investment. However, the degree to which hospitals respond to different sizes of yardstick incentives is not clear and must be empirically analyzed.

To examine the responses of hospitals to changes in the size of yardstick incentives, we vary the percentage of Medicare payment withholdings from 1% to 4% in increments of 0.25%. Noting that the current program started off withholding 1% of payments and then gradually increased the percentage to 2% over a five-year period, we follow the same suit and increase the payment percentage in a linear manner from 1% to the counterfactual target percentage over a five-year period. The exact payment percentages withheld used in the counterfactual analyses are presented in Table A.4 of Appendix A.1.

To illustrate the effect of changing the size of yardstick incentives, we present in Figure 2.5 the impact on two selected quality measures that represent two different dimensions of quality: CLABSI and hospital overall rating. First, CLABSI is one of the most common, yet potentially lethal, hospital-acquired infections where bacteria or viruses enter the bloodstream through the central venous catheter and cause an infection (Haddadin et al., 2021). It is estimated that as many as 28,000 patients die from CLABSI annually in the U.S. alone

Figure 2.5: Counterfactual results of changes in Medicare payment withholding percentage



Note. The error bars represent 95% confidence intervals for both performance in selected quality measures and total investment costs incurred from 2011 to 2018.

(AHRQ, 2021). We find that increasing the percentage of payments that are withheld for the program from 2% to 3% results in an increase of $646M in total investment in quality from 2011 to 2018, which can lead to a 2.1% decrease in the average rate of infection. Increasing the size of yardstick incentives even further to 4% can result in an average decrease of 3.3% in the rate of infection.

Second, hospital overall rating represents patient satisfaction, and it is measured by the proportion of patients who gave the hospital a 9 or 10 on a patient satisfaction survey question asking the patient to rate the hospital on a 10-point scale. We find that increasing the size of yardstick incentives to 4% can result in a 1% increase in the average proportion of patients who would rate their hospital as 9 or 10. On the other hand, reducing the size of yardstick incentives to 1% can lead to a 1.4% decrease in the average proportion of patients who are satisfied with their hospital.

Overall, we confirm that hospitals respond to changes in the size of yardstick incentives by changing their investment decisions, and this analysis highlights the importance of setting the appropriate size of yardstick incentives. Furthermore, it can be used to prescribe a specific level of incentive payments given a desired average level of quality.

2.6.2. Tailored Peer Groups

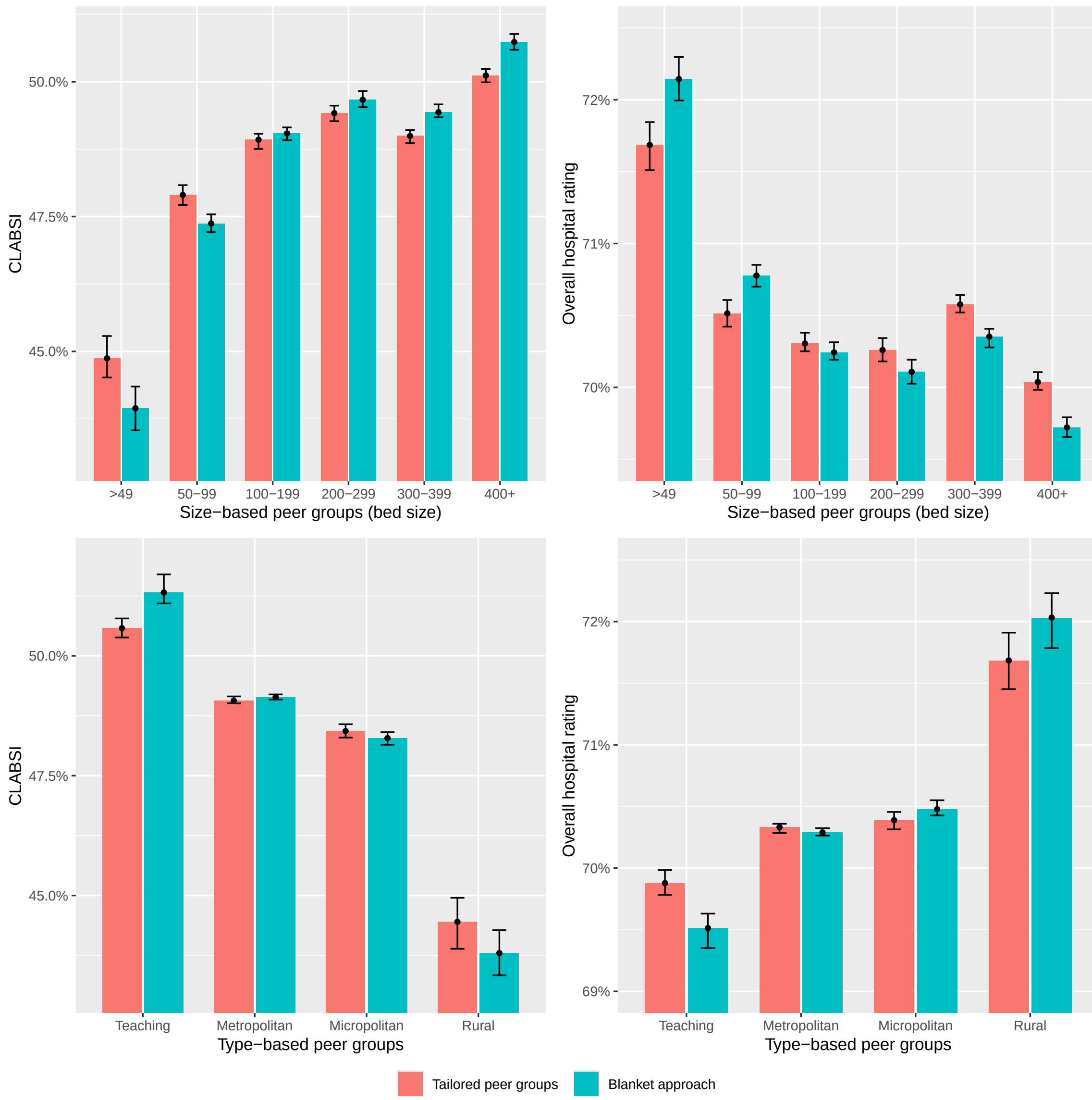
The next set of counterfactual analyses is motivated by our finding that the costs associated with investment in quality significantly vary across different types of hospitals. Shleifer (1985) posits that modifying the yardstick competition by adjusting the compensation based on the expected cost of effort can result in optimal outcomes when firms have heterogeneous costs of effort. The intuition is that hospitals with higher cost of investment will find incentive payments to be insufficient as hospitals with lower cost push up the distribution of quality. In contrast, hospitals with lower cost of investment will find quality improvements to be even more rewarding as the incentive payments are in part made higher due to the lack of investment from hospitals with higher costs. Therefore, tailoring the peer groups in which hospitals compete for yardstick incentives can be an effective solution if the goal of the policymaker is to ensure a certain level of quality over the entire industry, i.e., reduce the variation in quality, or induce quality investments from hospitals with higher cost of investment.

One potential implementation of such modification is to group hospitals by characteristics that are known to be associated with cost of investment, making hospitals compete with peer hospitals with similar distribution of cost of investment. We examine two alternative schemes based on our finding that the cost of investment has statistically significant association with the size of the hospital, measured by the number of inpatient beds, and operating location, grouped by metropolitan, micropolitan, and rural areas, as well as being a teaching hospital. Under each of these alternative schemes, the overall size of the yardstick incentives is kept at the "baseline" level of 2%.

First, we analyze the effect of a yardstick competition with *size-based peer groups* in which hospitals are categorized based on the size of the hospital. In particular, we divide the hospitals into six categories: less than 49 beds, 50–99 beds, 100–199 beds, 200–299 beds, 300–399 beds, and 400 or more beds. We choose these categories following how Medicare classifies hospitals by bed size. As presented in Section 2.5.1, we find that larger hospitals tend to incur higher cost of investment, likely driven by their more complex organizational structures. Second, we analyze the effect of *type-based peer groups* in which hospitals that operate in a similar setting, in terms of location and teaching designation, are grouped together. The location is categorized by metropolitan, micropolitan, and rural areas, and we separate teaching hospitals as an individual group because we find that teaching hospitals incur higher cost of investment even after the operating location is controlled for. This is in line with the idea that even if two hospitals operate in the same metropolitan area, for example, an academic medical center will be vastly different from a community hospital.

Again, we highlight the counterfactual results using two quality measures: CLABSI and hospital overall rating. We report our findings in Figure 2.6, where we compare the outcomes of tailored yardstick competitions against the "blanket approach" where all hospitals

Figure 2.6: Counterfactual results of yardstick competition with tailored peer groups



Note. The error bars represent 95% confidence intervals of the average performance within each group.

non-discriminatorily compete for the same yardstick incentives. We find that both tailored yardstick competitions are successful in inducing quality improvements from groups of hospitals associated with higher cost of investment. In particular, we find that hospitals with 400 or more beds, the group associated with highest cost of investment, having the hospitals compete for yardstick incentives within the group of largest hospitals results in an average reduction of 1.2% in the rate of infection as well as an increase of 0.45% in the proportion of patients who rate their hospital as 9 or 10. We find similar patterns for the tailored yardstick competition with type-based peer groups that categorize hospitals based on location and teaching designation. We find that for teaching hospitals, the group associated with higher cost of investment, tailored peer groups lead to a reduction of 1.4% in the rate of infection and an improvement of 0.5% in the proportion of patients who are satisfied with their hospital. In contrast, we find that in both tailored yardstick competitions, the average quality of the groups with lower cost of investment actually declines by a small margin, suggesting that hospitals with lower cost of investment are over-incentivized and hospitals with higher cost of investment are under-incentivized under the blanket approach.

Our finding not only supports the intuition on how different hospitals with different cost of investment respond to yardstick incentives but also highlights the importance of understanding the underlying distribution of costs associated with quality investment. The two sets of counterfactual analyses lend themselves to the possibility of combining the two approaches we explored. For instance, since we observe that larger hospitals and teaching hospitals have the highest cost of investment, and therefore generally lower performance in quality measures, policymakers can combine the two approaches and vary the size of yardstick incentives in each peer group for a more effective competition.

2.7. Discussion and Conclusion

In this work, we employ a novel application of structural estimation methods to empirically investigate the VBP Program, a quality-based yardstick competition that involves almost the entire hospital industry in the U.S. We model hospitals' responses to the quality-based yardstick competition as a dynamic equilibrium where hospitals engage in quality investments by weighing the costs of investment against the expected benefits from incentive rewards that take into account both the individual hospital's quality as well as the distribution of peer hospitals' quality. Using a hidden Markov chain framework to model changes in quality of care, we estimate structural parameters that allow us to analyze how hospitals respond to the yardstick competition.

In our analysis, we adopt the perspective of the payer for health care services and treat the current implementation of the VBP Program as a diagnostic tool that reveals the hidden costs of quality investments within the U.S. hospital system. Our initial goal, the estimation of financial and non-financial costs associated with quality investment for individual hospitals comprising this system, serves as a bridgehead for the main task of designing future yardstick incentive schemes. Our analysis focuses on two alternatives to the current implementation of the yardstick incentives. Under the first alternative, the peer yardstick group encompasses all hospitals, as is done under the current VBP Program, but the overall size of the yardstick incentives is varied. Under the second alternative, the size of the yardstick incentives is kept at the current 2% level, but we design tailored peer groups that only include hospitals with similar type or size. We find that increasing the size of the incentives from 2% to 4% can lead to a 3.3% reduction in the average rate of CLABSI and a 1% increase in the average proportion of patients who rate their hospital 9 or 10 on a 10-point scale. We also find that designing tailored peer groups, even without changing the size of the incentives, can lead to an average reduction of 1.4% in the rate of CLABSI among groups of hospitals associated with the highest costs of quality investment. Our analysis is designed to enable payers to perform quantitative assessments of the impact of alternative yardstick programs on the achieved quality of patient care.

Our work has significant managerial and policy implications. Most directly, the underlying structure and parameters estimated in this work can be used to inform policymakers on how to better design yardstick competitions. Furthermore, large insurance firms in the U.S. have also been incorporating both cost-based and quality-based yardstick incentives into their provider reimbursement policies and, thus, can utilize our findings to better understand how effective these policies can be. Although we focus our analysis on the hospital industry in the U.S., our work has further implications for other industries with potential to use yardstick incentives to motivate behavioral changes. For instance, online gig-economy platforms can use yardstick incentives to induce competition among service providers that reaches outside of a particular local market or service area. Yardstick competition can be utilized in other settings in the health care industry as well, such as for large "home-based" health care agencies with offices in multiple locations or different physician groups within a large hospital system.

Finally, there are limitations in this work that also outline avenues for future research. First, our assumption that investment in quality is a binary action may not hold in reality. Although the framework that we propose can be easily modified to include multiple levels of investment decisions, the estimation of an increased set of parameters would require much more detailed data collected over a longer observation period. Second, although we do not find evidence that the change in performance on the set of quality measures selected for the VBP Program leads to changes in hospital throughput, the quality of care may be positively correlated with patient demand for care in settings where there exists a significant level of local competition among hospitals. An examination of the impact of geographic competition on the performance of yardstick incentives represents an interesting direction for future work.

CHAPTER 3

The Spillover Effects of Capacity Pooling in Hospitals

3.1. Introduction

Hospitals often face significant variability in demand, both in terms of the number of patients needing care and the type of care needed by each patient. While such variability exists on the demand side, hospitals operate with a fixed number of beds, not only across the entire hospital but also within each specialty service (e.g., Cardiology, General Surgery, etc.), which is allocated a fixed number of beds. The mismatch between the number and type of patient arrivals and the capacity of hospital beds presents significant challenges in matching the supply with the demand.

One of the strategies employed by hospitals and many other industries that face similar problems is capacity pooling. This practice allows the hospital to utilize underused capacity in a less busy service when other services are at or near full capacity. Pooling the capacity of hospital beds results in the placement of patients of a focal service in a bed that is located in a unit that has been designated for another service; this is called "off-service placement" (Dong et al., 2020; Stylianou et al., 2017; Song et al., 2020). While off-service placement allows for a more efficient use of beds, recent empirical work has found that off-service placement has negative consequences when it comes to the care provided for the patients who are placed off service. Stylianou et al. (2017) find that these patients have longer lengths of stay on average. Song et al. (2020) obtain similar results using an instrumental variable approach to estimate the causal effect of off-service placement, finding that off-service patients experience longer lengths of stay and a higher likelihood of readmission within 30 days.

An important question that remains is whether the practice of off-service placement has any downstream effects on the rest of the patient population that is not impacted by the practice first-hand. In other words, are there any spillover effects of off-service placement that hospital administrators should be aware of? The existing literature has not yet explored this possibility of off-service placement having a broader spillover effect on patients who have not been placed off service themselves but do belong to the same service that has some of its patients placed off service. *A priori*, the answer to this question is not obvious. While it is possible that the consequences of off-service placement are limited to the negative first-order effects previously documented, the broader impact on the workflow of the physicians who are caring for all patients on their service—regardless of their placement—could lead to substantial negative spillover effects whereby patients who are placed on service are also negatively impacted. Understanding these spillover effects is important to hospital administrators because they have significant implications for managing hospital capacity, especially given many hospitals operate at very high levels of utilization.

In this paper, we quantify the spillover effects of off-service placement as they are experienced by on-service patients of the same specialty service. From the perspective of an on-service patient, the extent to which patients who belong to the same service are placed off service will change constantly over the course of her hospitalization due to the inbound and outbound movements of other patients. To fully capture the spillover effect of off-service placement while simultaneously accounting for its time-varying nature, we operationalize two key components of off-service placement experienced by a given on-service patient: the *level* of off-service placement (i.e., the overall degree to which patients belonging to the service are placed off service) and the *volatility* of off-service placement (i.e., the frequency and the magnitude of the changes in the degree to which patients belonging to the service are placed off service).

Using an instrumental variable approach, we find that off-service placement has substantial negative spillover effects on the efficiency and quality of care received by on-service patients. First, patients placed on service tend to experience longer lengths of stay when the average level of off-service placement for the service is high. Specifically, a one standard deviation increase in the level of off-service placement during a patient's hospitalization is associated with a 29% increase in length of stay. When the volatility of off-service placement is high, patients experience not only longer lengths of stay but also a higher likelihood of readmission to the hospital within 30 days and a higher likelihood of clinical trigger activation. In this case, a one standard deviation increase in the volatility of off-service placement during a patient's hospitalization is associated with a 13% increase in length of stay.

Using the point estimates from our empirical analyses, we conduct a series of counterfactual analyses to show the expected performance of several alternate routing policies that may be able to reduce the negative impact of off-service placement while retaining the benefits of capacity pooling. We find that limiting the practice of reserving on-service beds in anticipation of future demand can lead to significant reductions in the overall level of off-service placement, which in turn is expected to result in better patient outcomes. A policy of boarding patients for an extra hour when an on-service bed is expected to become available soon and another policy of prioritizing early discharges may also be effective in reducing the level of off-service placement; both of these policies are expected to lead to shorter lengths of stay and a decreased likelihood of clinical trigger activation.

A key contribution of this work lies in the quantification of the spillover effects of off-service placement. In doing so, we add to the recent stream of work analyzing the effects of off-service placement in health care delivery settings, especially with regards to challenges and unintended consequences when implementing capacity pooling strategies (Stylianou et al., 2017; Song et al., 2020; Dong et al., 2020; Kim et al., 2020). More broadly, this work also contributes to the literature on capacity management in health care delivery settings (Shi et al., 2016; Xie et al., 2020; Dai and Shi, 2020; Dong and Perry, 2020). The practical implications of our work are substantial, given many hospitals utilize off-service placement as a capacity pooling strategy. For hospital managers, our work further highlights the importance of better managing off-service placement since its impact is not only limited to those patients who are placed off service themselves but also extends to the patients who are placed in on-service beds, effectively impacting the entire population of hospitalized patients.

The rest of the paper is organized as follows. Section 3.2 discusses off-service placement in more detail, paying particular attention to the level and the volatility of off-service placement as it impacts on-service patients. The research setting and data are introduced in Section 3.3. Section 3.4 motivates our empirical strategy of using instrumental variables. We specify the empirical models and present the main results in section 3.5. Section 3.6 presents alternate specifications, including nonlinear analyses and additional robustness checks. The procedures and results of the counterfactual simulation studies are discussed in section 3.7. Section 3.8 concludes.

3.2. Off-Service Placement: First-Order Effects and Spillover Effects

In recent years, there has been a growing body of work that seeks to understand the effects of off-service placement on patient outcomes and system performance. In this section, we provide a brief overview of the existing literature and motivate why there is a need to better understand the potential spillover effects of this practice.

3.2.1. First-order Effects of Off-service Placement

To date, the research that studies this widespread practice of off-service placement can be characterized as focusing on its first-order effects. Some of this work has sought to address the question of how off-service placement impacts the efficiency and quality of care received by patients who themselves have been placed off service. Using an instrumental variables approach, Song et al. (2020) estimate that being placed off service is associated with a 23% increase in length of stay and 13% increase in the likelihood of 30-day hospital readmission for hospitalized medical/surgical inpatients. Similar effects are documented in studies that focus on specific specialty services, such as general medicine (Kohn et al., 2020a; Bai et al., 2018), pulmonary medicine (Kohn et al., 2020b), and cardiac medicine (Alameda and Suárez, 2009).

Meanwhile, others have focused on the decision-making process underlying the bed placements and its impact on system-wide performance. Shi et al. (2016) develop a stochastic network model that allows for off-service placement to show the effects of this practice on average waiting time performance. Dong et al. (2020) find that bed managers are more likely to place patients off service when the service is busy and the admission occurs during the overnight shift. They also illustrate that a more uniform routing policy could reduce the overall levels of off-service placement and improve system performance. In contrast, Dai and Shi (2019) treat the bed manager's decision-making process as a Markov decision process and solve for the optimal routing policy using approximate dynamic programming. Izady and Mohamed (2021) propose a routing policy in which a cluster of services have a designated flex unit to which patients can be admitted. They find that an optimal configuration of these clusters can lead to reductions in the cost of denied admissions.

3.2.2. Spillover Effects of Off-service Placement

Beyond the first-order effect, placing patients off service may have an important second-order effect as well, which we refer to as the spillover effect. There are two ways in which a spillover effect of off-service placement could impact on-service patients. First, there may be a spillover effect at the *service level*, wherein on-service patients belonging to a particular service may be impacted by the extent to which there are off-service patients who also belong to the same service. Second, there may be a spillover effect at the *unit level*, wherein on-service patients located in a particular unit may be impacted by the extent to which there are off-service patients who are located in the same unit. In this paper, we focus specifically on the *service-level* spillover effects for two inter-related reasons. First, we are interested in understanding how the practice of off-service placement impacts the work being carried out by physicians and nurses. Whereas unit-level spillover effects are expected to be impacted by the work being done by nurses, service-level spillover effects should be impacted by the work of both nurses and physicians. Second, prior work by Song et al. (2020) shows that the work of physicians may be more impacted by off-service placement than that of nurses, given their examination of potential mechanisms underlying the first-order effects of off-service placement. For brevity, we refer to service-level spillover effects simply as the "spillover effect" in the remainder of this paper.

Prior research shows convincingly that the efficiency and quality of care for patients is impacted by operational factors such as the system's load (Kc and Terwiesch, 2009; Kuntz et al., 2015; Berry Jaeker and Tucker, 2017), service level mismatch (Kim et al., 2015; Chan et al., 2019), and facility layout (Meng et al., 2020). The spillover effect of off-service placement, separately accounting for the previously documented operational factors such as system load, may be another important aspect that has been overlooked. It is also particularly difficult to operationalize because the extent to which the service is engaged in off-service placement is continuously changing over the course of a given patient's hospitalization. For the sake of exposition, consider for example Patient A who was admitted to the General Medicine service and placed in an on-service bed. When she was first admitted, say the General Medicine service had 100 patients under its care, of whom 20 were placed off service. As other patients are admitted to, discharged from, and transferred into and out of the General Medicine service, both the total number of patients who belong to the General Medicine service and the number of General Medicine patients who are placed off service continuously changes. Say halfway through her hospitalization, the service has 110 patients under its care with 30 placed off service. By the time she is ready to be discharged, the service has 90 patients with 12 placed off service. In this example, the extent to which the General Medicine service is engaged in off-service placement during Patient A's hospitalization evolves over time from 20% to 27% to 13%.

The example of Patient A highlights that there may be two relevant dimensions through which off-service placement may impact patients who are placed on service: the *level* of off-service placement as experienced by the on-service patient and the *volatility* of off-service placement as experienced by the on-service patient. While the former captures the extent to which the service engages in off-service placement on average, the latter captures the fluctuation in off-service placement that occurs during the course of the on-service patient's hospitalization. We hypothesize that each of these factors will have a negative impact on the efficiency and quality of care experienced by patients who are placed on service.

There are several potential mechanisms through which increases in the *level* of off-service placement might negatively impact on-service patients. First of all, physicians provide care for all patients who belong to the service, regardless of whether the patient is placed in an on-service bed or in an off-service bed. In contrast, nurses are only responsible for the care of patients who are located in their unit, regardless of the service to which the patient belongs. As a result, having patients placed in an off-service unit creates an *ad hoc* team comprised of the service's physicians and the unit's nurses, who do not have an established working relationship. Higher levels of off-service placement may increase the number of such *ad hoc* teams being formed, which are likely to result in increased coordination costs (Reagans et al., 2005; Dobson et al., 2009). Having more off-service patients may also create disruptions to the provider workflow, as the care coordination meetings and rounds for off-service patients require physicians to allocate additional time outside of their normal routine (Gesensway, 2010). Typically, this involves physically traveling to the unit where the off-service patient is located, which consumes additional time and further disrupts the physician's workflow. Although we do not focus on distance as a key metric in this paper, Meng et al. (2020) find that the distance between patient beds and the nurses' station significantly affects care patterns, such that nurses are more likely to batch tasks for patients who are located in rooms that are farther away. Similarly, we expect the provision of care to be different for patients who are placed on service versus off service, and that the care provided for off-service patients may in turn affect the way in which care is provided for on-service patients. Taken together, we hypothesize that increases in the level of off-service placement will lead to decreases in the efficiency and quality of care for patients who are placed on service.

There are also several ways in which the *volatility* of off-service placement, caused by large and frequent changes in off-service placement, may create disruptions to the process of delivering care to on-service patients. In the operations management literature, it has long been established that high levels of variability—a measure of volatility—are associated with higher cost and lower quality (Lee and Tang, 1998; de Treville and Antonakis, 2006; Sriram et al., 2015; Lee et al., 2004; Fisher and Raman, 1996). In the health care setting specifically, reducing variability in patient flow and patient types is associated with improvements in operational efficiency (Chand et al., 2009; Litvak et al., 2005; Soremekun et al., 2011). When it comes to off-service placement, we expect that having highly volatile levels of off-service placement across the service would be associated with disruptions to the care delivery process and higher demands on physicians' time, since admissions, transfers, and discharges involving off-service patients mechanically increase the volatility of off-service placement. Admissions into off-service beds create disruptions for on-service patients because physicians must now travel to and care for patients located elsewhere. When patients are located farther away rather than co-located, the additional travel increases the workload, further exacerbating the negative effects of increased workload on outcomes (Kc and Terwiesch, 2009; Aiken et al., 2002). Even when off-service patients are discharged, which would lower the level of off-service placement and the workload of the service's physicians, physicians must dedicate additional time and attention during these events (e.g., to disposition decisions and discharge planning), which can in turn cause disruptions for other patients who are on service. Such disruptions are exacerbated by the fact that otherwise-routine rounds and meetings with nursing teams (during which discharge planning is often discussed) are usually absent for off-service patients. Furthermore, movements involving off-service patients result in the formation and dissolution of *ad hoc* teams of physicians and nurses, which can also disrupt the care of other patients who are on service. Thus, we hypothesize that large and frequent changes, regardless of the direction, in the level of off-service placement will lead to decreases in the efficiency and quality of care for patients who are placed on service.

3.3. Research Setting and Data

3.3.1. Research Setting

We collaborated with a large academic medical center located in the northeastern region of the United States. As of 2016, this hospital had 473 medical/surgical inpatient beds, which were located across 17 units and allocated to eight services. Here, a *unit* refers to a physical location where there are a certain number of beds. In turn, each unit is designated to a particular *service*, which is a department comprised of a single clinical specialty or a related group of specialties that tend to have smaller volumes. Figure B.1 in Appendix B.1 shows which units are designated for which services in the study hospital. Because each unit belongs to a specific service, we are able to determine whether the patient occupying a bed in a particular unit has been placed on service (if she belongs to the service for which the unit has been designated) or off service (if she belongs to a service other than the one for which the unit has been designated). Thus, at the service level, the number of patients who are placed off service can vary over time, and the number of distinct off-service units across which these patients are placed can also vary over time.

In our study hospital, there are eight unique services: Cardiac Medicine, Cardiac Surgery, East Surgery, General Medicine, Neurology, Oncology Medicine, Transplant, and West Surgery.[2] It is important to note that physicians belong to a particular *service*, whereas nurses belong to a particular *unit*. Such organization is ubiquitous among hospitals as physicians often specialize in particular clinical specialties, while nurses are generally trained to provide nursing care for all types of medical/surgical inpatients.[3] In effect, this means that physicians are responsible for the care of all patients who belong to their service, regardless of whether the patient is placed on or off service. In contrast, nurses are responsible for the care of all patients who are located in their unit, regardless of the service to which the patient belongs. This has implications for the familiarity that accumulates between physicians and nurses; the most frequent teamwork and coordination happens between physicians of a given service and the nurses working on the unit that is designated to that particular service.

While physicians and nurses are responsible for the care of the patient, the decision to place the patient in a particular on- or off-service bed is made by the hospital's bed managers. Bed managers are centralized at the hospital level and serve the role of admission controllers. They are experienced nurses with clinical knowledge, and they coordinate with the admitting

[2] East and West Surgery are groups of relatively smaller surgical specialties.

[3] Exceptions include Hospitalists, who specialize in providing inpatient care rather than a specific type of condition or a body part, and Oncology nurses, who specialize in administering chemotherapy and other cancer treatments.

service and the various units across the hospital to assign the patient to a specific inpatient bed. The bed manager does not have any discretion over which service to admit the patient to; the service makes this determination based on the patient's medical conditions and clinical needs. The bed manager, however, does have the discretion to determine where to physically place the patient given the availability of beds in different units across the hospital. For example, if there is an incoming Cardiac Medicine patient but all on-service beds for Cardiac Medicine are occupied, the bed manager can assign this patient to an open bed in a unit that belongs to General Medicine; this creates an off-service placement for the Cardiac Medicine service. Note, when the bed manager decides to place a patient off service, only other medical/surgical beds within the same level of care are considered. In other words, this Cardiac Medicine patient would not be placed in a critical care bed (e.g., in a Cardiac Intensive Care unit) or a non-medical/surgical bed (e.g., in an Obstetrics unit).

3.3.2. Data and Analysis Sample

Our data consist of detailed patient and operational data from October 1, 2015 to September 30, 2016. By combining multiple proprietary data sources, we are able to accurately track the location of all patients at the bed-hour level over the entire study period. In addition, we have detailed information about each patient encounter, including demographic characteristics, primary diagnosis, complications, and comorbidities. A direct comparison of the patient's service and bed assignment allows us to determine whether the patient was placed on or off service. The granularity of our data also allows us to track, at the hour level, the number of on- and off-service patients belonging to a given service.

During the 12-month study period, there were a total of 52,476 patient encounters at this hospital. To define our analysis sample, we first exclude patient encounters without any time in a medical/surgical bed. Since our analysis focuses on how the on-service patient population is impacted by the spillover effects of off-service placement, we further restrict our sample to patients who were placed in an on-service bed. In addition, we exclude patients who were transferred from an on-service bed to an off-service bed, or *vice versa*, in order

Table 3.1: Summary statistics of analysis sample

	Mean	SD	Min	Max
Mean of proportion off service	0.16	0.12	0	0.49
SD of proportion off service	0.021	0.017	0	0.17
Length of stay (days)	4.16	3.33	0.70	23.8
Logged length of stay (days)	1.48	0.54	0.53	3.21
Hospital readmission (%)	18.9	39.1	0	100
Clinical trigger activation (%)	8.65	28.1	0	100
In-hospital mortality (%)	1.05	10.2	0	100
Age (years)	62.1	16.7	16.8	107.4
Female (%)	50.4	50.0	0	100
DRG cost weight	1.82	1.37	0.49	17.7
Complications or comorbidities (%)	24.6	43.1	0	100
Number of transfers	2.82	0.89	2	9
Unit-level utilization (%)	91.4	6.91	26.2	100
Service-level utilization (%)	91.2	5.68	50.2	99.9
ICU encountered (%)	14.9	35.6	0	100
Admitted on weekday (%)	84.9	35.8	0	100
Admission shift				
AM shift	12.1	32.6	0	100
PM shift	27.7	44.8	0	100
Overnight shift	60.2	48.9	0	100

Note. N = 14,793. SD, standard deviation.

to isolate the spillover effect of off-service placement from the direct effect of being placed off service. In other words, we restrict our analysis sample to those who were placed in an on-service bed during the entire duration of their hospital stay, which leaves us with a total of 14,793 patient encounters across 13,295 unique patients.

3.3.3. Outcome Measures

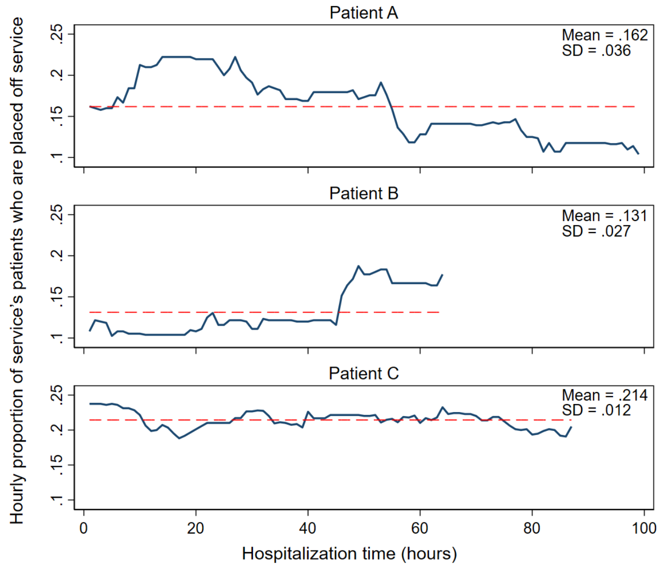
We consider four patient-level outcome measures that proxy for the efficiency and quality of care: (1) hospital length of stay, (2) hospital readmission, (3) clinical trigger activation, and (4) in-hospital mortality. We define length of stay as the time from each patient's first admission into a medical/surgical bed until the patient's discharge. For example, if a patient's first entry point to the hospital is the emergency department, we do not consider the patient to have begun her hospital stay until the patient is transferred out of the emergency department and into a medical/surgical bed. Our definition of length of stay is purposefully aligned with the goal of analyzing the impact of off-service placement and is designed to represent the efficiency of care a patient receives once the patient begins her stay in a medical/surgical bed. Because this measure is right-skewed, we log transform it to calculate the logged length of stay. Each of the remaining three measures—hospital readmission, clinical trigger activation, and in-hospital mortality—are operationalized as binary variables. Hospital readmission is defined as an all-cause readmission to the hospital within 30-days after discharge. A clinical trigger is a discretionary or a non-discretionary alarm notifying the physician and the nursing team of a potential deterioration in the patient's condition. In-hospital mortality is another measure that proxies for care quality and captures whether the patient died during the course of the hospitalization. We report summary statistics of each of these measures in Table 3.1.

3.3.4. Key Explanatory Variables

We have two key explanatory variables of interest: the *level* of off-service placement and the *volatility* of off-service placement as experienced by the focal on-service patient during her hospitalization. We measure each using hourly snapshots of the proportion of a service's patients who are placed off service, which is calculated by dividing the number of off-service patients by the total number of patients belonging to the focal service; this measures the extent to which a service's patients are placed off-service. Using these hourly snapshots, we then define the level of off-service placement by calculating the mean of the hourly snapshots over the duration of each on-service patient's hospital stay. We define the volatility of off-service placement by calculating the standard deviation of the hourly snapshots over the duration of each on-service patient's hospital stay. Table 3.1 reports summary statistics of each of these measures.

To illustrate these two key measures, Figure 3.1 depicts how the proportion of a service's patients who are placed off service evolved over the course of the hospitalization of 3 different patients who were placed on service. We observe that patient A was admitted when her service had 16% of its patients placed off service, which soon increased to 23% and gradually decreased to 12% by the time she was nearing discharge. In contrast, patient B was admitted when his service had 11% of its patients off service, which stayed relatively constant until

Figure 3.1: Hourly snapshots of proportion off service



Note. This figure shows, for three different patients who were placed on service, hourly snapshots of the proportion of their respective service's patients who were placed off service over the course of their hospitalization. The horizontal dashed line shows the mean of the hourly snapshots. The mean and standard deviation of the hourly snapshots are shown in the upper-right corner for each patient.

the last 10 hours of his hospitalization when it increased to 17%. For patient C, her service's proportion of patients who were placed off service remained relatively constant at around 21% throughout her hospitalization. As we can see from the summary statistics shown in each panel, the mean and standard deviation of these hourly snapshots vary significantly across the three patients.

3.3.5. Patient and Operational Characteristics

In our estimations, we account for several patient and operational characteristics that may also impact our outcome measures of interest. We control for patient demographic characteristics, including age and gender. We also account for three proxies of patient severity: DRG cost weight, an indicator for the presence of complications or comorbidities, and an indicator for whether the patient spent time in the intensive care unit (ICU). The DRG cost weight represents the relative level of resources needed to treat a patient with a certain diagnosis. The average DRG cost weight of all patients admitted to a given hospital is sometimes referred to as the Case Mix Index (CMI), and the average DRG cost weight in our analysis of 1.82 indicates that patients admitted to our study hospital have diagnoses that are on average 82% more costly to treat than the average Medicare patient. We also observe and account for the presence of complications or comorbidities as classified in the DRG. Another proxy of severity is whether the patient spent time in the ICU, which suggests the need for higher-intensity care.

We also account for several characteristics of each patient's hospitalization, including the number of intra-hospital transfers, the time of day when the patient was admitted, whether the patient was admitted on a weekday, and the type of service to which the patient was admitted. The average patient in our analysis sample had 2.82 transfers during her hospitalization, which includes the admission and discharge events, as these constitute transfer events into or out of a bed. We observe and account for heterogeneity in admission time and day by controlling for the shift during which the patient was admitted and whether the patient was admitted on a weekday. Controlling for the type of service to which the patient was admitted helps us address heterogeneity across patient types.

Finally, we account for operational factors that may impact a patient's stay. In particular, we account for the overall busyness of the unit and of the service, so that we can isolate these effects from the spillover effect that we are interested in identifying. Specifically, we control for the average hourly utilization level of the unit in which each patient was placed and the service to which the patient belonged. We construct the hourly unit-level utilization by dividing the number of beds that are unavailable for an incoming patient (i.e., occupied, reserved, or closed beds) by the number of total beds in the unit (i.e., all unavailable beds plus open beds). Similarly, we construct the hourly service-level utilization by dividing the number of unavailable beds by the total number of beds in all units that have been designated to the service. Given the prior research that shows that high levels of workload impose an inverted U-shaped effect on patient outcomes (Kuntz et al., 2015; Berry Jaeker and Tucker, 2017), we include the squared terms of both unit-level and service-level utilization measures in our analyses as well.

3.4. Empirical Strategy

To identify the spillover effects of off-service placement, our goal is to estimate the following:

$$y_i = \beta_0 + \beta_1 \cdot \textit{Level of off-service placement}_i + \beta_2 \cdot \textit{Volatility of off-service placement}_i + \rho \cdot \mathbf{X}_i + \epsilon_i \tag{3.1}$$

Here, y_i is the outcome measure of interest for patient i, $\mathbf{X}_i$ is a vector of control variables, and ϵ_i captures the error term. The spillover effect of off-service placement is captured by the coefficients on the level and volatility of off-service placement.

3.4.1. Endogeneity of the Level of Off-service Placement

From previous studies, we know that the decision to place patients in an on- versus off-service bed is a function of patient severity that is unobservable to the econometrician (Song et al., 2020). Although we focus in this paper on the population of *on*-service patients in order to estimate the spillover effects of off-service placement, the selection of patients into the on-service population still poses endogeneity concerns. For example, if the focal service is near full capacity, a relatively sicker patient arriving to the service is more likely to be placed on service than a relatively healthier patient. If we were to assume that services tend to have a higher proportion of off-service patients when they are near capacity, this will lead to a biased selection of sicker patients into the population of on-service patients. In turn, this would mean that sicker patients who are placed on service are more likely to begin their hospitalization when the service is carrying a higher proportion of off-service patients. In this case, we would expect β_1 to be an overestimate of the true spillover effect attributable to the level of off-service placement.

While we do indeed observe that the proportion of a service's patients who are placed off service is often low when the service's level of utilization is low, we also observe that, when the service's level of utilization is high, there is a large heterogeneity in the proportion of patients who are placed off service. In fact, the correlation coefficient between the proportion

Table 3.2: Expected placement decision when the service is near full capacity

	(i) Focal patient is of lower severity	**(ii) Focal patient is of higher severity**
(a) High level of off-service placement at service level	More likely to be placed *on* service	More likely to be placed *on* service
(b) Low level of off-service placement at service level	More likely to be placed *off* service	More likely to be placed *on* service

of patients off service at the time of admission and the service's level of utilization at one hour prior to admission is quite low ($r = 0.19$). This prompts us to separately examine cases when the level of off-service placement at the time of the focal patient's admission is high versus low, conditional on the service being near full capacity.

The difference in the likelihood of being placed off service for a sicker patient versus a healthier patient largely stems from the bed manager's decision to keep some on-service beds reserved in anticipation of the arrival of future patients with higher level of severity. Nevertheless, the ability to place a patient off service is also limited by the capacity constraints in units belonging to other services, both in terms of not having any available beds themselves or trying to keep some beds open in anticipation of patients arriving to those other services. Therefore, having many patients placed off service already (i.e., a high level of off-service placement at the service level) will make it harder for the focal service to place additional patients off service, forcing more of the incoming patients to be placed on service until all on-service beds are full. This is illustrated in row (a) of Table 3.2. On the other hand, if a service has a relatively small number of patients who are currently placed off service, this suggests that there is a greater chance of being able to place incoming patients off service. In this case, healthier patients will be more likely to be placed off service while sicker patients will be more likely to be placed on service, as we see in row (b) of Table 3.2.

To articulate the endogeneity concern, we reexamine the relationships summarized in Table 3.2 by columns as opposed to by rows. Overall, we see that lower severity (healthier) patients are more likely to be placed off service compared to higher severity (sicker) patients only when the service has a relatively small number of its patients already placed off service. This pattern would result in nonrandom assignment of the level of off-service placement where healthier patients who are placed in on-service beds experience a disproportionately higher level of off-service placement at the time of their admission (column (i)) compared to sicker patients (column (ii)). Since healthier patients tend to experience both a higher level of off-service placement and better outcomes, we would expect β_1 in Equation (3.1) to underestimate the true spillover effect of off-service placement in this case. Such routing behaviors, and the resulting endogeneity, is exhausted once all on-service beds are occupied, at which point all patients regardless of their level of severity will be routed to off-service beds until the entire hospital reaches full capacity.

This endogeneity concern hinges, in part, on the previously stated assumption that the bed manager's off-service placement decisions are limited by capacity constraints. We can test this assumption empirically by examining whether the number of off-service patients belonging to the service at the time of the focal patient's admission has a concave relationship with the focal patient's likelihood of being placed off service. If so, this would suggest that patients are less likely to be placed off service when the service already has a high level of off-service patients at the time of the focal patient's admission. Table 3.3 shows the results from regressing the probability of a focal patient being placed off service on the linear and squared terms of the number of off-service patients that the focal patient's service has at the time of her admission. The concave relationship illustrated by the coefficients suggests that off-service placements are indeed restricted by capacity constraints. We conduct a similar test by fitting a quadratic probability model that predicts the focal patient's probability of being placed off service using the number of off-service patients belonging to the relevant service at the time of her admission. Figure 3.2 shows these results, which again validates our assumption.

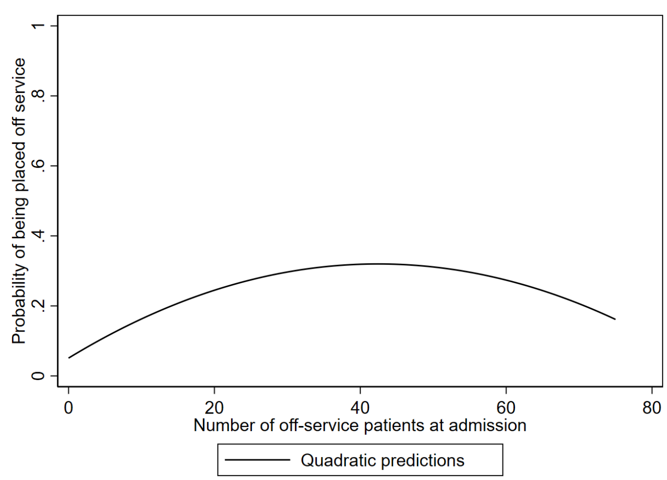
While the *level* of off-service placement may be endogenously determined by patient severity, the *volatility* of off-service placement, as experienced by patients who have been placed on service, does not suffer from the same endogeneity concern. This is because the volatility of the service's off-service placement during the patient's hospitalization captures the degree

Table 3.3: Probability of off-service placement based on the number of off-service patients at admission

	Off-service placement
Number of off-service patients at admission	0.0127***
	(22.98)
(Number of off-service patients at admission)^2	-0.000149***
	(-14.58)
Constant	0.0509***
	(10.61)
Observations	18351

Note. t-statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Figure 3.2: Quadratic probability model predicting the likelihood of off-service placement as a function of the number of off-service patients at admission



Note. This figure plots the resulting fit of a quadratic probability model that predicts a patient's probability of being placed off service using the number of off-service patients belonging to the focal service at the time of admission.

of disruption that occurs *after* the placement decision for the focal patient has already occurred. Events that drive the volatility of off-service placement are functions of overall demand shocks and the characteristics of other patients as opposed to the focal patient. Specifically, unobserved factors concerning the focal patient do not affect the bed managers' decisions to place other patients into on- versus off-service beds during the remainder of the focal patient's hospital stay. Furthermore, we control for hourly utilization at both the unit level and at the service level to separately account for demand shocks, such as an unexpected influx of patients, which could affect both the outcome variables and the volatility of off-service placement.

3.4.2. Instrumental Variables

We use an instrumental variable (IV) approach to address the endogeneity problem in the level of off-service placement. This allows us to recover a causal estimate of the spillover effect of the level of off-service placement. For an IV to be valid, it must satisfy two conditions: (1) it must be correlated with the endogenous variable (i.e., the mean of the service's proportion of off-service patients), and (2) it must be uncorrelated with the structural error term, ϵ_i, which captures unobserved patient characteristics conditional on all of the control variables as well as the main explanatory variables.

For our analyses, we use two IVs: (1) the preadmission service-to-hospital utilization ratio and (2) the preadmission hospital utilization excluding the focal service. We measure each variable in the hour prior to the focal patient's time of admission. The two IVs reflect key information that bed managers use when making the placement decision for the focal patient; the one-hour lead accounts for the delays between when the patient is assigned to a particular bed and when the patient is physically transported and placed in that bed.

We expect each of these IVs to be correlated with the level of off-service placement in the following way: a service has a greater proportion of its patients placed off service when (a) it is busier relative to the overall hospital and when (b) the rest of the hospital (excluding the focal service) is less busy (i.e., there are off-service beds available to send patients to).

In other words, we expect the level of off-service placement to increase as the service-to-hospital-utilization ratio increases and as the rest of the hospital's utilization decreases. A simple regression, as we will see in section 3.5.1, illustrates that this indeed is the pattern that we observe in the data, allowing us to reject the null hypothesis that there is no relationship between the endogenous variables and each of the IVs. Specifically, we find that the level of off-service placement is positively associated with the service-to-hospital-utilization ratio and negatively associated with the rest of the hospital's utilization.

With regards to the exclusion condition, each of the IVs must be uncorrelated with unobserved factors captured by the structural error term, ϵ_i. Stated differently, each of the IVs should affect the outcome variable only through the endogenous variable, conditional on the control variables. We rely on the fact that both IVs are measured in the hour prior to admission, which should rule out any potential connection between the instruments and unobserved patient severity. For the exclusion restriction to be valid, it is also crucial that the IVs capture utilization at the hospital and the service level as opposed to at the unit level. While utilization at the unit level may be correlated with unobserved patient severity since the bed manager could select sicker patients to be admitted when the unit's utilization level increases (Dong et al., 2020), utilization at the hospital level and at the service level are determined exogenously, since the bed manager does not have the discretion to turn patients away or assign a patient to a different service. Thus, utilization at the hospital and service levels are not correlated with unobserved factors related to patient severity. Instead, they reflect the overall conditions of the hospital and of the relevant service at the time of the bed manager's decision making.

One mechanism through which the exclusion condition could be violated is if the preadmission utilization levels remain relatively stable throughout the patient's hospitalization. This would result in a strong positive correlation between the preadmission utilization levels and the average utilization levels that the patient experiences during her hospital stay. Prior research has documented that high levels of utilization have a negative effect on the efficiency and quality of care (Kc and Terwiesch, 2009; Aiken et al., 2002), and therefore it is important to separately account for the utilization level experienced by the patient during the course of her hospitalization. To separately account for this factor, we control for the average utilization level of both the unit and the service in all of our analyses.

3.5. Empirical Models and Estimation Results

3.5.1. Spillover Effects of Off-service Placement

Using the two IVs discussed in section 3.4.2, we estimate the causal spillover effect of off-service placement using a two-stage approach. For the continuous outcome measure—logged length of stay—we use a two-stage least squares (2SLS) model. For the binary outcome variables—hospital readmission, trigger activation, and in-hospital mortality—we use a two-stage residual inclusion (2SRI) model, otherwise known as the control function approach.

We begin by estimating the following first-stage equation at the patient level, which is common to both the 2SLS and 2SRI models:

$$\begin{aligned} &\textit{Mean of proportion off service}_i \\ &\qquad = \gamma_0 + \gamma_1 \cdot \textit{Service-to-hospital utilization ratio}_i \\ &\qquad + \gamma_2 \cdot \textit{Hospital utilization excluding focal service}_i \\ &\qquad + \gamma_3 \cdot \textit{SD of proportion off service}_i + \delta \cdot X_i + e_i \end{aligned} \tag{3.2}$$

For the 2SLS approach, we then estimate the following second-stage equation:

$$y_i = \beta_0 + \beta_1 \cdot \textit{Mean of } \widehat{\textit{proportion}} \textit{ off service}_i + \beta_2 \cdot \textit{SD of proportion off service}_i + \rho \cdot X_i + \epsilon_i \tag{3.3}$$

Including the predicted *Mean of proportion off service*$_i$ from the first-stage equation as an explanatory variable in the second-stage equation allows us to causally estimate the spillover effect by exploiting variation in the *Mean of proportion off service*$_i$ that is caused by the two IVs.

For the 2SRI approach, we estimate the following second-stage Probit model:

$$P(y_i = 1 | \textit{Mean of proportion off service}_i, \textit{SD of proportion off service}_i, X_i)$$
$$= \Phi(\beta_0 + \beta_1 \cdot \textit{Mean of proportion off service}_i + \beta_2 \cdot \textit{SD of proportion off service}_i + \rho \cdot X_i + \hat{e}_i) \tag{3.4}$$

Including the residuals from the first-stage equation, $\hat{e}_i$, allows us to consistently estimate the causal effect of the spillover effect stemming from off-service placement.

We report the results from these estimations in Table 3.4.[4] In column (1), the first-stage results show that the preadmission service-to-hospital utilization ratio is positively associated with the mean of the proportion of a service's patients who are placed off service ($\gamma_1 = 0.070$, $p < 0.001$). Specifically, a one standard deviation increase in this ratio is associated with a 0.5 percentage point increase in the mean of the service's proportion of off-service patients. On the other hand, the preadmission utilization of the rest of the hospital (excluding the focal service) is negatively associated with the mean of the proportion off service ($\gamma_2 = -0.125$, $p < 0.001$). Here, a one standard deviation increase in the rest of the hospital's utilization is associated with a 0.6 percentage point increase in the mean of the service's proportion of off-service patients. Both of these results are consistent with our expectation as discussed in section 3.4.2. In addition, the large Kleibergen-Paap F-statistic ($F = 71.96$) suggests that these two IVs are strong joint predictors of the endogenous regressor, $\textit{Mean of proportion off service}_i$.

In columns (2) to (5), we estimate the second stage equations. Our results show that there are substantial negative spillover effects of off-service placement. We find that a higher level of off-service placement that a service has during a patient's hospitalization is associated with a longer length of stay for the focal patient, and that greater volatility in the level of off-service placement is associated with both a longer length of stay and a higher likelihood of experiencing a clinical trigger activation. Specifically, the coefficients indicate that a 1

[4]See Appendix B.2 for full results tables including coefficients for all control variables.

percentage point increase in the mean of the proportion of patients placed off service leads to a 2.5% increase in length of stay and a 1 percentage point increase in its standard deviation leads to a 7.8% increase in length of stay. With regards to the likelihood of experiencing a clinical trigger activation, a 1 percentage point increase in the standard deviation of the proportion of patients placed off service is associated with a 1.45 percentage point increase in the likelihood of experiencing a clinical trigger activation.

One way to put into perspective the magnitude of these coefficients is to calculate the marginal change associated with a one standard deviation increase in the variable of interest. A back-of-the-envelope calculation indicates that a one standard deviation increase in the level of off-service placement is associated with a 28.8% increase in length of stay, whereas a one standard deviation increase in the volatility of off-service placement is associated with a 12.9% increase in length of stay and a 2.6 percentage point increase in the likelihood of experiencing a clinical trigger activation. Given that 8.6% of all on-service patients experienced at least one clinical trigger activation during their hospitalization, a 2.4 percentage point increase can be interpreted as a 29.5% increase at the mean.

3.5.2. Potential Mechanisms Underlying the Spillover Effects

Challenges in Coordinating with Nursing Teams Caring for Off-service Patients.

A potential mechanism that could underlie the negative spillover effects of off-service placement is the challenges in coordination between physicians and nursing teams that arise when patients are placed off service. To examine this possibility, we conduct additional analyses in which we substitute, in the first- and second-stage equations, the proportion of patients placed off service with the number of units across which off-service patients have been placed. If coordination challenges were to explain our findings, then we would expect to continue to see these negative spillover effects when accounting for the number of separate units across which a service has placed off-service patients, i.e., the number of nursing teams with which the service's physicians must coordinate. Since physicians routinely conduct rounds only on the units that are designated to the service, they typically do not have regular working

Table 3.4: Spillover effect of off-service placement, operationalized using the service's proportion of patients placed off service

	(1) Mean of proportion off service	(2) Logged length of stay	(3) Hospital readmission	(4) Trigger activation	(5) In-hospital mortality
Preadmission service-to-hospital utilization ratio	0.0699***				
	(0.0105)				
Preadmission hospital utilization excluding focal service	-0.125***				
	(0.0157)				
Mean of proportion off service		2.484***	-0.348	-1.263	-0.870
		(0.567)	(1.384)	(1.812)	(4.775)
SD of proportion off service	1.358***	7.753***	2.694	10.47***	-0.812
	(0.0559)	(0.867)	(2.012)	(2.499)	(7.293)
Controls	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes
Model	2SLS	2SLS	2SRI	2SRI	2SRI
	1st stage	2nd stage	2nd stage	2nd stage	2nd stage
Observations	14793	14793	14793	14793	14793

Note. Standard errors (in parentheses) are heteroskedasticity robust for continuous outcome variables and bootstrapped for binary outcome variables. Controls not shown include age, sex, DRG cost weight, complications and comorbidities, number of transfers, ICU encountered, unit-level utilization, (unit-level utilization)^2, service-level utilization, (service-level utilization)^2, service type, admission shift, and weekday admission. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

relationships with nursing teams based in other (off-service) units. As a result, we would expect that having to coordinate care with a greater number (i.e., higher level) of or with a frequently changing number (i.e., higher volatility) of nursing teams in off-service units will decrease the efficiency and quality of care for patients who are placed on service.

The results of these estimations are shown in Table 3.5. We find that having off-service patients placed across a greater number of units, on average, is associated with longer lengths of stay for on-service patients. When the number of units across which off-service patients are placed is more volatile, on-service patients experience longer lengths of stay, higher likelihood of readmission to the hospital, and higher likelihood of clinical trigger activation. Specifically, we find that having off-service patients in one additional unit is associated with a 4.4% increase in length of stay for on-service patients who belong to the same service. In addition, a one unit increase in the standard deviation of the number of units with off-service patients is associated with a 31.9% increase in length of stay, a 2.7 percentage point

Table 3.5: Spillover effect of off-service placement, operationalized using the number of units across which the service has patients placed off service

	(1) Mean of units with off-service patients	(2) Logged length of stay	(3) Hospital readmission	(4) Trigger activation	(5) In-hospital mortality
Preadmission service-to-hospital utilization ratio	4.558*** (0.323)				
Preadmission hospital utilization excluding focal service	-4.051*** (0.537)				
Mean of units with off-service patients		0.0441*** (0.0108)	0.00658 (0.0353)	-0.0188 (0.0357)	-0.0134 (0.1000)
SD of units with off-service patients	0.923*** (0.0535)	0.319*** (0.0178)	0.107* (0.0431)	0.276*** (0.0380)	0.0557 (0.117)
Controls	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes
Model	2SLS 1st stage	2SLS 2nd stage	2SRI 2nd stage	2SRI 2nd stage	2SRI 2nd stage
Observations	14793	14793	14793	14793	14793

Note. Standard errors (in parentheses) are heteroskedasticity robust for continuous outcome variables and bootstrapped for binary outcome variables. Controls not shown include age, sex, DRG cost weight, complications and comorbidities, number of transfers, ICU encountered, unit-level utilization, (unit-level utilization)^2, service-level utilization, (service-level utilization)^2, service type, admission shift, and weekday admission. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

increase in the likelihood of readmission, and a 4.1 percentage point increase in the likelihood of clinical trigger activation. A similar back-of-the-envelope calculation as discussed earlier shows that a one standard deviation increase in the average number of units with off-service patients is associated with a 15.6% increase in length of stay for on-service patients, and that a one standard deviation increase in the standard deviation of the number of units with off-service patients is associated with a 15.3% increase in length of stay, a 6.9% increase in the likelihood of readmission, and a 22.5% increase in the likelihood of clinical trigger activation.

Challenges During High-workload Periods Associated with Patients Placed Off Service.

To examine a second mechanism, we focus on the finding that it is not only the *level* but also the *volatility* associated with off-service placement that seems to result in the negative spillover effects. Volatility in the proportion of a service's patients who are placed off service is generated by admissions into and discharges out of off-service units. Furthermore, compared to the interim period of a patient's hospitalization, the periods when a patient is being admitted and discharged tend to demand the most time and attention from the physician. In the next set of analyses, we examine whether challenges associated with these high-workload periods due to movements into and out of off-service units might be another mechanism underlying our main findings.

To test our hypothesis that the higher workload experienced while managing the admissions and discharges of off-service patients contributes to the negative spillover effects, we construct a new explanatory variable: the count of total movements into and out of off-service units incurred by patients belonging to the same service during the focal patient's hospitalization. Since the average rate of admissions and discharges varies significantly across different services, we demean the count of movements at the service level. Then, we substitute the volatility measure in Equations (3.2), (3.3), and (3.4) with this demeaned count of movements in off-service units. The results of estimating this new set of second-stage equations, shown in Table 3.6, illustrate that the negative spillover effects of the volatility of off-service placement continue to hold with this alternate specification. Increases in the number of movements into and out of off-service units during a given on-service patient's hospitalization are associated with a longer length of stay, a higher likelihood of readmission, and a higher likelihood of clinical trigger activation.

Of course, it is possible that movements into and out of on-service units also contribute to periods of higher levels of workload and create disruptions for other patients who are placed on service. However, we would expect that the potential impact of movements into and out of off-service units would have a larger spillover effect than movements into and out of on-service units. This is because providing care for patients who are placed off service creates additional disruptions beyond providing care for on-service patients, given the need to take extra time away from routine rounds and to coordinate care with nursing teams whom the physician may not have well-established working relationships. To examine this possibility, we test whether the spillover effect of movements into and out of *off*-service units is statistically significantly greater than the spillover effect of movements into and out of

Table 3.6: Spillover effect of off-service placement, operationalized using the number of movements into and out of off-service units

	(1) Logged length of stay	(2) Hospital readmission	(3) Trigger activation	(4) In-hospital mortality
Mean of proportion off service	2.593***	-0.216	-1.100	-1.086
	(0.470)	(1.679)	(2.132)	(4.124)
Demeaned count of movements	0.00521***	0.000862***	0.00433***	0.000729
into and out of off-service units	(0.000132)	(0.000233)	(0.000279)	(0.000552)
Controls	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Model	2SLS	2SRI	2SRI	2SRI
	2nd stage	2nd stage	2nd stage	2nd stage
Observations	14793	14793	14793	14793

Note. Standard errors (in parentheses) are heteroskedasticity robust for continuous outcome variables and bootstrapped for binary outcome variables. Controls not shown include age, sex, DRG cost weight, complications and comorbidities, number of transfers, ICU encountered, unit-level utilization, (unit-level utilization)^2, service-level utilization, (service-level utilization)^2, service type, admission shift, and weekday admission. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

on-service units.

A straightforward approach to compare these two effects would be to include both types of movements (vis-à-vis off-service units and on-service units) in a single regression analysis, and then compare the coefficients using a statistical test. However, there exists a high level of correlation ($r = 0.91$) between these two measures, which results in multicollinearity. To avoid this issue, we estimate a seemingly unrelated regression with two models, with each model accounting for either the demeaned count of movements into and out of off-service units or the same for on-service units. In Table 3.7, we see that the spillover effects stemming from movements into and out of off-service units have a greater impact on the length of stay than those stemming from movements into and out of on-service units. Performing a Wald test confirms that the difference between these two coefficients is statistically significant at the 0.001 level.

3.6. Alternate Specifications

To assess the robustness of our findings and further evaluate model fit, we consider additional specifications.

Table 3.7: Seemingly unrelated regression on spillover effects of movements into and out of on-service units versus off-service units

	(1) Logged length of stay	(2) Logged length of stay
Mean of proportion off service	2.777***	2.884***
	(0.503)	(0.454)
Demeaned count of movements into and out of off-service units	0.00267***	
	(0.000132)	
Demeaned count of movements into and out of on-service units		0.00153***
		(0.0000719)
Controls	Yes	Yes
Month FE	Yes	Yes
Model	2SRI 2nd stage	2SRI 2nd stage
Observations	14793	14793

Note. Standard errors (in parentheses) are heteroskedasticity robust for continuous outcome variables and bootstrapped for binary outcome variables. Controls not shown include age, sex, DRG cost weight, complications and comorbidities, number of transfers, ICU encountered, unit-level utilization, (unit-level utilization)^2, service-level utilization, (service-level utilization)^2, service type, admission shift, and weekday admission. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

3.6.1. Nonlinear Analyses

We begin by considering nonlinear models to estimate the spillover effects. Nonlinear models may provide a more comprehensive picture of how off-service placement impacts the efficiency and quality of care for on-service patients. To allow for a high degree of flexibility in our model, we perform a semiparametric analysis for each of the outcome variables. To address the endogeneity concern regarding the level of off-service placement, we use the same 2SRI approach described in section 3.5. Then, we follow Robinson (1988) to estimate partially linear models where each of the explanatory variables of interest, $\textit{Mean of proportion off service}_i$ and $\textit{SD of proportion off service}_i$, is assumed to have a nonparametric functional form. In addition, we include the squared term of the linear explanatory variable to allow for as much flexibility as possible. In effect, we estimate the following two equations for each outcome variable:

$$y_i = \beta_0 + G(\textit{Mean of proportion off service}_i) + \beta_1 \cdot \textit{SD of proportion off service}_i + \beta_2 \cdot \textit{SD of proportion off service}_i^2 + \rho \cdot X_i + \hat{e}_i + \epsilon_i \quad (3.5)$$

$$y_i = \beta_0 + G(SD\ of\ proportion\ off\ service_i) + \beta_1 \cdot Mean\ of\ proportion\ off\ service_i$$
$$+ \beta_2 \cdot Mean\ of\ proportion\ off\ service_i^2 + \rho \cdot X_i + \hat{e}_i + \epsilon_i \quad (3.6)$$

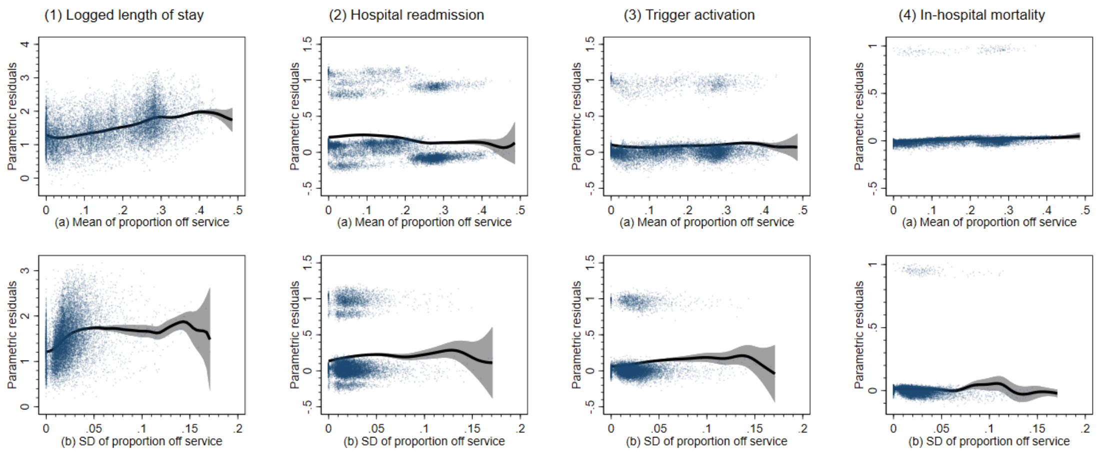
The nonparametric function, G, is estimated using a kernel regression, and $\hat{e}_i$ is estimated by the residuals from the first-stage equation. We present these estimation results visually in Figure 3.3.

The results of the semiparametric analysis suggest that the spillover effects are generally linear with a few notable exceptions. For the spillover effects stemming from the *level* of off-service placement, the semiparametric plots show that when the proportion of a service's patients who are placed off service is very low (less than 5%), we do not find evidence of a spillover effect leading to increases in the length of stay (panel 1a) or higher likelihoods of clinical trigger activation (panel 3a). We see a similar pattern with respect to the spillover effects from the volatility of off-service placement on length of stay. The marginal effect of the volatility of off-service placement on length of stay is relatively small when the standard deviation of the proportion of patients placed off service is low (panel 1b). The magnitude of the effect increases as the volatility of off-service placement increases, until it starts to decrease again once the volatility increases even further (panel 1b). This suggests that there may be an upper bound at which the volatility of off-service placement affects patient outcomes.

3.6.2. Maximum Positive Deviation and Maximum Negative Deviation

Next, we employ an alternate operationalization of the volatility of off-service placement. One possible critique of our current approach is that the standard deviation of the proportion of a service's patients who are placed off service may simply capture sporadic increases in the level of off-service placement. In other words, it is possible that the negative spillover effects that we identify using the standard deviation of proportion off service are purely driven by increases in off-service placement that are not captured by the mean of the proportion off service. To examine whether both positive and negative deviations in the proportion of off-

Figure 3.3: Semiparametric estimations of the spillover effect of off-service placement



Note. This set of figures presents the results of conducting semiparametric analyses using partially linear models where one explanatory variable is estimated nonparametrically (Robinson, 1988). Parametric residuals (blue dots) represent the variation in the outcome variables that are not explained by the parametric portion of the model. The fitted line is generated using a Gaussian kernel weighted local polynomial regression. The grey band around the fitted line shows the 95% confidence interval around the nonparametric fit.

service patients negatively impact the efficiency and quality of care, we perform an additional analysis wherein we substitute the standard deviation of the proportion off service with the maximum positive and maximum negative deviations, respectively, from the proportion of off-service patients at the time of admission. The results, reported in Table 3.8, show that the deviation in both the positive direction and in the negative direction have similar negative spillover effects on the length of stay and trigger activation likelihood for on-service patients.

3.7. Counterfactuals

So far, we have focused on identifying and quantifying the spillover effects of off-service placement. Our analyses suggest that, not only does off-service placement have a negative first-order effect on those patients who are placed off service, it also has a negative spillover effect on patients who are placed on service. In what follows, we conduct a series of counterfactual analyses to identify which, if any, alternate routing policies may be successful in mitigating the spillover effects of off-service placement and to estimate the magnitude of the potential gains from adopting each of the policies. These analyses will help hospital

Table 3.8: Spillover effect of off-service placement, operationalized using the maximum positive and negative deviations from the proportion off service at time of admission

	(1) Logged length of stay	(2) Hospital readmission	(3) Trigger activation	(4) In-hospital mortality
Mean of proportion off service	2.128***	-0.484	-1.852	-1.529
	(0.570)	(1.749)	(2.094)	(4.467)
Maximum positive deviation in proportion off service	3.530***	0.615	5.034***	1.996
	(0.354)	(1.074)	(1.298)	(2.832)
Maximum negative deviation in proportion off service	2.502***	0.941	4.001**	-2.020
	(0.343)	(0.967)	(1.216)	(2.711)
Controls	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Model	2SLS 2nd stage	2SRI 2nd stage	2SRI 2nd stage	2SRI 2nd stage
Observations	14793	14793	14793	14793

Note. Standard errors (in parentheses) are heteroskedasticity robust for continuous outcome variables and bootstrapped for binary outcome variables. Controls not shown include age, sex, DRG cost weight, complications and comorbidities, number of transfers, ICU encountered, unit-level utilization, (unit-level utilization)^2, service-level utilization, (service-level utilization)^2, service type, admission shift, and weekday admission. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

administrators determine ways in which they can continue to enjoy the benefits of capacity pooling while minimizing their negative spillover effects.

3.7.1. Setup of Counterfactual Analyses

We leverage the granularity of our data and perform simulations using the actual admissions, transfers, and discharges of patients in the observed data.[5] Similar to the approach used in Bertsimas and Pauphilet (2020), we assume that our data provide an accurate representation of the underlying data-generating process, which includes the patient flow, the severity of different types of patients, and the service rates for patients as a function of both observable and unobservable factors.

To begin the simulation, we take a snapshot of the hospital on the first day, and then we implement an alternate routing policy whenever patients move into and out of medical/surgical

[5]Another approach would be to construct a queueing model consisting of arrival rates and service rates that are functions of the spillover effects we estimated. While such an approach provides flexibility to simulate virtually any counterfactual policies, its accuracy is limited by the many necessary assumptions underlying the model. We bypass these concerns regarding the extent to which such assumptions would represent the real world by using a data-based approach.

units. Throughout the simulation, we carefully track the utilization of each unit and each service at the hour level. Because non-medical/surgical units (e.g., intensive care units, observation beds) are beyond the scope of this study, we route movements into and out of those units to reflect the observed movements. Once all patients are re-routed using an alternate routing policy, we then calculate the (counterfactual) proportion off service that each patient experienced in each hour of her hospital stay. These hourly snapshots are then used to calculate the mean and the standard deviation of the proportion of patients off service, in exactly the same way that they were calculated using the observed data for the empirical analyses.

Next, we calculate the predicted counterfactual outcomes for three measures: length of stay, likelihood of hospital readmission, and likelihood of trigger activation. Because our empirical analyses suggest that there are no spillover effects with respect to in-hospital mortality, we do not calculate a predicted counterfactual likelihood for that outcome measure. In calculating these predictions, we make a slight modification to our previous approach discussed in section 3.5. Specifically, we separately identify spillover effects for medical specialty services versus surgical specialty services in order to utilize more granular effect estimates for different patient types.[6]

As with the empirical analyses, we focus our counterfactual analyses on the spillover effects of off-service placement onto the population of patients who are placed on service. The first-order effects of off-service placement on the outcomes of patients who are themselves placed off service has been previously studied and are beyond the scope of this paper (see Song et al., 2020). As a result, we focus strictly on the counterfactual outcomes of patients who are placed on service. Of course, one of the goals of the alternate routing polices is to reduce the overall number of off-service patients, which in turn is expected to reduce the spillover effects of off-service placement. Given that the previous literature has found negative first-order effects of off-service placement on patients who have been placed off

[6]See Appendix B.3 for details on how we identify heterogeneous treatment effects by service type.

service, the counterfactual results from our simulations can be considered a lower bound of the total effect, as they account for only the reduced spillover effects and not the reduced first-order effects. These results are also conservative because we hold the universe of on-service patients fixed as those in the observed data and do not allow for the possibility of more admissions when patients in the system are discharged at a faster rate.

3.7.2. Alternate Routing Policies

On until 0 beds.

Bed managers in the hospital often keep some number of on-service beds reserved in anticipation of future admissions or transfers, especially when the number of open beds remaining on service is small and the arriving patient seems to have a relatively low level of severity. This first policy does not allow for on-service beds to be left unoccupied in anticipation of future demand. Instead, all arriving patients are placed in on-service beds until there are no more beds available. Once all on-service beds have been exhausted, patients are placed in the service's primary off-service unit; once that is full, they are placed in the service's secondary off-service unit. For each service, we define a primary and secondary off-service unit by using the observed data and identifying the two units where the greatest number of off-service patients were placed; these are shown in Figure B.2 of Appendix B.1. Restricting the number of units across which off-service patients are placed could help reduce coordination costs between physicians and nursing teams. In cases when a service has multiple on-service units, we prioritize placements into units that do not serve as a designated off-service unit for another service to ensure that as many patients as possible are placed on service. Once both primary and secondary off-service units have been exhausted, patients are placed in any medical/surgical unit with the greater number of available beds.

On until 0 beds + boarding.

One way to reduce the incidence of off-service placement is to allow for some additional boarding time prior to admission into a medical/surgical bed. While excessive time spent boarding is associated with undesirable outcomes (Chalfin et al., 2007; Mathews and Long, 2015; Rabin et al., 2012), if it helps avoid an off-service placement, the combined benefits may outweigh the losses. Since bed managers can observe which beds are expected to become available in the next few hours (e.g., due to an expected discharge or expected transfer), if they are given the discretion to delay the bed placement of an incoming patient when an on-service bed is expected to become available soon, it may be possible to place the patient on service and avoid an off-service placement. Based on discussions with clinical leaders at our study hospital and to err on the side of being too conservative, we allow for only one hour of additional boarding time.

On until 0 beds + earlier discharge.

Rather than allowing for additional boarding, which aims to reduce off-service placements by delaying admissions, another option would be to expedite discharges. In other words, hospitals could prioritize discharging patients earlier in the day (i.e., morning versus afternoon or evening) to facilitate admissions into on-service beds (Shi et al., 2016). Benson et al. (2006) find that 12% of all surgical patients in a UK hospital experienced delays in discharge despite being medically fit to leave. When simulating this policy, we assume that patients who were discharged between 11AM and 5PM could have been discharged at 11AM if physicians and nurses had reorganized their days to prioritize discharges; these times were also determined based on our discussions with clinical leaders at our study hospital. When discharges occur earlier in the day, patients arriving in the afternoon who would otherwise have been placed in an off-service bed could be placed on service.

On until 0 beds + hospital-wide flex units.

Rather than designating primary and secondary off-service units for each service, this policy designates two units as hospital-wide flex units. From the observed data, we identify the two units that received the highest number of off-service patients and designate these as the two hospital-wide flex units. These two units no longer serve as on-service units for a specific service; instead, they provide off-service care for off-service patients across all services. Under this policy, each incoming patient is placed on service as long as there is an available on-service bed. Once there are no more on-service beds available, patients are sent to the flex unit with the most available beds. Once both flex units become full, patients are then placed in any medical/surgical unit with the most available beds.

On until a few beds.

This policy seeks to mimic the behavior that was prohibited in the three polices above—reserving on-service beds in anticipation of future arrivals. Here, we allow patients to be placed off service when there are fewer than five on-service beds remaining. To simulate the bed managers' decisions, we use the following approach to calculate probabilities from the observed data, which will allow us to determine the likelihood of a given patient being placed off service. First, for each service, we calculate the proportion of patients who were placed off service given x open beds at the time of their arrival (where $x \leq 5$). For example, if there were 100 admissions to the General Surgery service when 1 General Surgery bed was open and 30 of these patients were placed off service, the probability of an incoming General Surgery patient being placed off service when exactly 1 on-service bed is available is 0.3. Then, using these probabilities, if a patient is assigned to be placed off service, we first route them to the primary off-service unit, and then to the secondary off-service unit. We show the set of computed probabilities for each service in Table B.1 of Appendix B.1.

On until a few beds + protected services.

Across the different services within the hospital, some tend to have higher levels of their patients placed off service than others. Specifically at our study hospital, the Cardiac Surgery, East Surgery, Oncology, and Transplant services are specialized services that try to minimize the incidence of placing their patients in off-service beds; this can be seen in Figure B.3 of Appendix B.1. In some cases, this is due to licensing restrictions. For example, administering chemotherapy requires special nursing training and licensing, so patients who are admitted for cancer treatment will always be placed on service in a unit designated for the Oncology service. As another alternate routing policy, we designate these four services as "protected services" and minimize the chances of their needing to place their own patients off service by restricting off-service patients from other services from flowing into units that are designated to them. In other words, on-service units for the protected services do not serve as designated off-service units for other services, and patients in the protected services are always placed on service as long as there is an open bed. For the non-protected services, we continue to implement the “on until a few beds” policy. This alternate routing policy is the strictest policy that we test in the sense that it explicitly reduces the level of capacity pooling in an attempt to reduce off-service placements and their associated spillover effects.

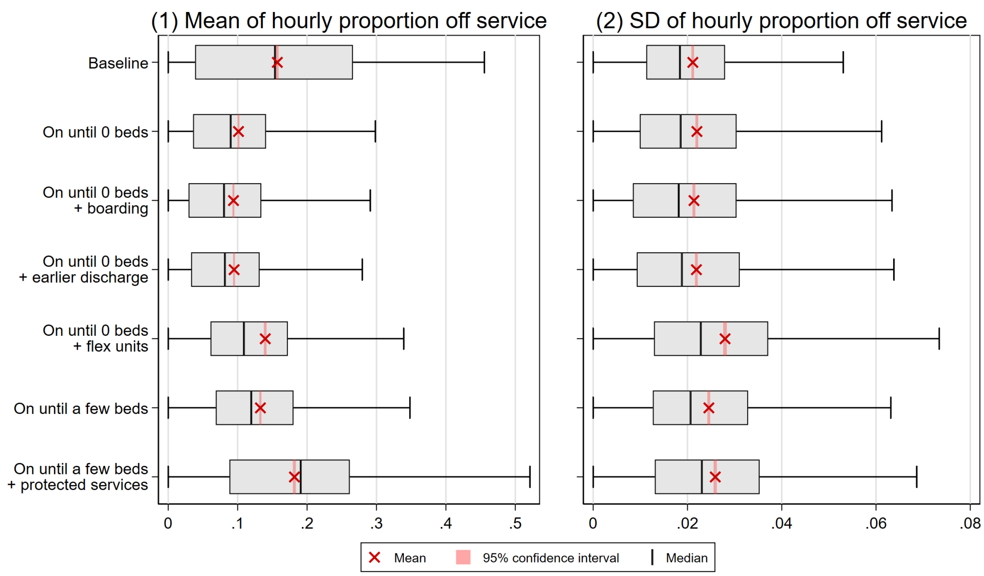
3.7.3. Simulation Results.

The counterfactual results derived from our simulations illustrate that alternate routing policies could indeed reduce the overall level of off-service placement and, in turn, result in reductions in the average patient length of stay. On the other hand, the volatility of off-service placement is much more challenging to address as it stems from the inherent variability in patient demand; as a result, most of the alternate routing policies we consider are not successful in meaningfully reducing the volatility of off-service placement. Nevertheless, we find meaningful impacts on the outcome measures of interest.

The boxplots in Figure 3.4 report the counterfactual mean (panel 1) and standard deviation (panel 2) of the proportion of patients who are placed off service. The top row labeled “Baseline” reflects the observed data. The subsequent rows represent each alternate routing policy described in section 3.7.2. Next, Figure 3.5 plots the sample mean and 95% confidence interval of average length of stay, predicted likelihood of hospital readmission, and predicted likelihood of clinical trigger activation.

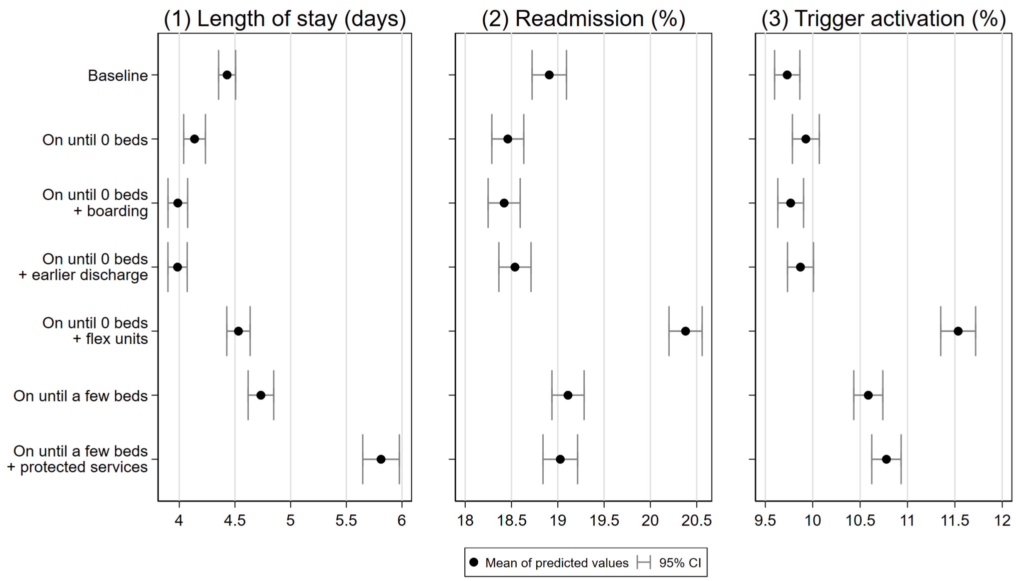
Based on the counterfactual outcomes from implementing the “on until 0 beds” policy, we see that the practice of reserving on-service beds in anticipation of a future arrival is one of the drivers that increases the overall level of off-service placement. Significant reductions in off-service placement can be achieved by restricting bed managers from reserving on-service beds and allowing them to place patients off service only if there is no available bed on service. That said, the observed behavior of reserving beds could be a result of the

Figure 3.4: Counterfactual mean and standard deviation of proportion of patients who are placed off service



Note. This figure reports the counterfactual mean and standard deviation of the proportion of patients who are placed off service. Each box plot shows, for the corresponding measure and routing policy, the sample mean and its 95% confidence interval, the median, the first and third quartile, and the minimum and maximum (excluding outliers).

Figure 3.5: Counterfactual length of stay, likelihood of hospital readmission, and likelihood of trigger activation



Note. This figure shows, for the corresponding measure and routing policy, the counterfactual mean and its 95% confidence interval.

bed manager having private information (unobservable to the researcher) about upcoming patient arrivals. Thus, the true gains from implementing a "on until 0 beds" policy may be more muted. Nevertheless, our results suggest that the practice of reserving beds should be limited and only be utilized when absolutely necessary.

We consider boarding and earlier discharge as an add-on to the "on until 0 beds" policy. We find that boarding a patient for an extra hour when an on-service bed is anticipated to become available may be an effective policy that could reduce both the overall level and the volatility of off-service placement. In practice, boarding patients who are transferred from other areas of the hospital (e.g., the emergency department) will demand additional resources from those areas; thus, the overall impact on the entire system must be considered. For patients who are being admitted directly, boarding the patient for a little more time until an on-service bed becomes available, as opposed to admitting the patient as soon as possible into an off-service bed, can be an effective solution. The benefits of additional boarding, however, must be weighed against potential clinical concerns in delaying care provision (Chalfin et al., 2007; Mathews and Long, 2015; Rabin et al., 2012). Early discharge also seems to reduce the overall level and volatility of off-service placement.

Finally, the outcomes from implementing the "on until 0 beds + hospital-wide flex units" and the "on until a few beds + protected services" policies provide insight into why careful planning of capacity pooling is crucial. While the two policies differ in their approach, neither policy is successful in reducing the overall level and volatility of off-service placement because each fails to sufficiently account for the capacity constraints of the units that are on the receiving end of off-service placements. Our simulations illustrate that cordoning off a set of protected services leads to situations where patients arriving to the other (non-protected) services experience high rates of off-service placement for sustained periods of time due to limited on-service capacity for prolonged periods. Similarly, with two hospital-wide flex units, patients on average experienced a large increase in the volatility of off-service placement because of the capacity constraints imposed on each of the two services that, in effect, lost one on-service unit. Furthermore, there were frequent periods of time when the two flex units did not provide enough capacity to place off-service patients across the eight services.

3.8. Discussion and Conclusions

In this paper, we investigate whether, and to what extent, there is a spillover effect of off-service placement that impacts the efficiency and quality of care for patients who are placed on service. This is a challenging question to address empirically for two reasons: (a) the service's level of off-service placement constantly changes over time, which means a given on-service patient may experience both high and low levels of off-service placement during her hospitalization, and (b) there is an endogeneity concern stemming from non-random assignment of patients into the on-service population. We address both of these challenges using detailed patient and operational data and an instrumental variable approach. To address the former, we characterize how a given on-service patient may be affected by off-service placement by operationalizing both the overall *level* and the *volatility* of off-service placement. We measure these two dimensions separately using bed-hour level data that tracks the location, service, and on- versus off-service status of each patient in the hospital. To address the latter, we identify two IVs that allow us to overcome the endogeneity concerns and estimate the causal spillover effects of off-service placement. We find that the on-service patient population experiences substantial negative spillover effects from off-service placement. Specifically, on-service patients experience longer lengths of stay when the service's average level of off-service placement is high during their hospitalization. When the volatility of off-service placement is high, they experience not only longer lengths of stay but also a higher likelihood of hospital readmission and a higher likelihood of clinical trigger activation.

Our findings have important managerial implications. For hospital administrators, this work further highlights the importance of better managing the practice of off-service placement, which is widespread among hospitals all around the world (e.g., Shi et al., 2016, Stylianou et al., 2017). Our findings illustrate that the effects of off-service placement reach beyond the patients who are placed off service and, instead, impact all patients throughout the hospital; this highlights the urgent need to re-examine the way in which hospitals leverage various capacity management strategies.

Our counterfactual results from the simulation studies provide insight into which other routing policies may be effective in reducing the incidence of off-service placement and improving outcomes for all patients, regardless of their placement location. Our findings suggest that hospitals should limit the practice of reserving on-service beds in anticipation of future arrivals of sicker patients, and that they should place off-service patients across fewer units in order to minimize coordination costs between physicians and nurses. Policies like earlier discharge initiatives, which do not target off-service placement specifically but are designed to improve overall efficiency, can also lead to meaningful improvements by allowing more patients to be placed on service. In cases when the bed manager has visibility into upcoming discharges, another possibility would be to board a patient a little longer when an on-service bed is expected to open up soon. Of course, each of these policies must be carefully considered by weighing the benefits of reducing off-service placement against the potential costs incurred by doing so (e.g., increased boarding time).

Our work opens up several avenues for further investigation of capacity pooling strategies and its implications for patient care. Although our findings are robust to several alternate specifications, our data come from a single hospital. Given the widespread use of off-service placement, we invite other researchers to study these effects of off-service placement in different settings to provide external validation to our findings. Methodologically, our study relies on the validity of the IVs used to address the endogeneity concern. Although we verify, to the best of our ability, the assumptions necessary to consider the instruments to be valid, a research setting that allows for random assignment of patients to on- versus off-service beds would be more robust. Future work could also extend our counterfactual analyses by moving beyond the first-order effect of the alternate routing policies. When considering the potential implications for throughput, the reduction in the counterfactual length of stay represents a lower bound of the potential gains, since policies that decrease patients' average length of stay would allow the hospital to admit additional patients and, thereby, increase throughput.

Hospitals continue to innovate to find ways to improve their efficiency and performance while operating with limited capacity. Understanding and addressing the various challenges surrounding the practice of placing patients off service will help hospital administrators implement better capacity management practices and improve the efficiency and quality of care.

CHAPTER 4

Should We Worry About Moral Hazard? Estimation of the Slutsky Equation Using Indemnity Health Insurance Contracts

4.1. Introduction and Motivation

The increase in the consumption of medical care when individuals are insured, known as moral hazard, has been identified as one of the main inefficiencies of health insurance (Pauly, 1968). The price distortion, generated by insurance coverage, alters the optimal medical care consumption by individuals as long as the demand is not perfectly inelastic, as pointed out in Pauly's seminal contribution, and could result in individuals purchasing more medical care units for which their value is below marginal cost. This increase in medical care consumption due to insurance was tested in the RAND Health Insurance Experiment (Manning et al., 1987) and found to be empirically relevant.

An assumption that is present in Pauly's original paper is the absence of "significant income effect", and therefore, all the increase in the quantity of medical care due to insurance is attributed to the distortion in prices introduced by insurance, and therefore, is considered inefficient. Under this framework, it is then optimal that insurance exhibit gaps in coverage, especially for those more price elastic services, and the use of deductibles and coinsurance to control medical care utilization would increase efficiency. This view is challenged by Nyman (1999), who emphasized the access motive of health insurance arguing that the income effects could be substantial.

In particular, Nyman (1999) argues that using the Marshallian demand to estimate the welfare loss from insurance results in an overstatement of that welfare loss, as it also includes the effects of income transfers from the healthy to the sick, which do not have a distortionary effect on prices. According to Nyman's calculations, the true welfare loss is about one third of the one estimated ignoring the income effects. This new insight puts into question the expenditure controls introduced by the insurance industry in their contract designs and the whole managed care model, calling for a reevaluation of the focus on cost-sharing and managed care by academics and policymakers.

Not surprisingly, the discussion at the beginning of this century was very active and controversial, not only in academia, but also in important popular press outlets (Gladwell, 2005). Despite the active discussion, no conclusion was reached as there was no credible measurement of the income and substitution effects of insurance. According to Nyman (1999) the ideal setting to answer this question would be one where indemnity and traditional insurance coexist. In the case of indemnity insurance, the insured receives a lump sum transfer of money in the bad state without affecting the relative prices, which identifies the income effect of insurance. Indemnity insurance is virtually extinct in most countries, which frustrated the efforts to provide a conclusion to this debate based on empirical analysis.

In this paper, we are able to measure the relative magnitude of the income and substitution effect of insurance by studying the market for supplemental health insurance in South Korea, where both indemnity and traditional insurance coexist. The National Health Insurance system in South Korea provides virtually universal coverage to all South Korean citizens, who are mandated to contribute to this system. This public insurance presents large gaps that implies high out-of-pocket payments by citizens, who then seek to alleviate the financial burden of these gaps by contracting with private companies through either indemnity or traditional health insurance policies.

In this setting, we are able to estimate the components of the Slutzky equation, which allows us to identify the elasticities of the Marshallian and Hicksian demands. Because we can estimate the response of consumers to price-reducing insurance plans and to income-transferring insurance plans, we can measure how much of the total moral hazard effect is welfare-reducing. Our data comes from the Korea Health Panel, which is a nationally representative annual survey tracking close to 8,000 households and more than 24,000 individuals over time. Important for this study, the survey provides data on medical care utilization, insurance status, and demographics among other important variables.

We find that 94% of moral hazard comes from the substitution effect, and therefore, the role of income effects is very limited. This result holds in general across several conditions and across categories of care. For patients with diabetes, heart diseases and stroke, chronic back conditions, and arthritis, we find similar magnitudes as the general results, and for patients with cancer, the income effect seems to be larger. For patients with cancer, for example, 59% of moral hazard is welfare-reducing. These results are also robust for different types of care, such as inpatient, outpatient and emergency care. In the case of inpatient care, 93% of moral hazard corresponds to the substitution effect, and similarly 95% of moral hazard is welfare-reducing. These results, contrast with the emergency care setting, where a lower 79% is due to inefficiencies.

Because our data comes from the decisions that people make in the market, the selection into the different kinds of plans is a concern. We exploit the institutions of the South Korean supplemental insurance market to assuage the concerns about selection, and in addition, we implement several tests that are commonly used in the literature, such as the correlation test of Chiappori and Salanié (2000), identifying the marginal consumer in a matching model, and explicitly adding consumers' private information to the estimation among others. Our results from these robustness analyses indicate that selection does not introduce major concerns for the interpretation of our results.

4.2. Theory and Empirical Strategy

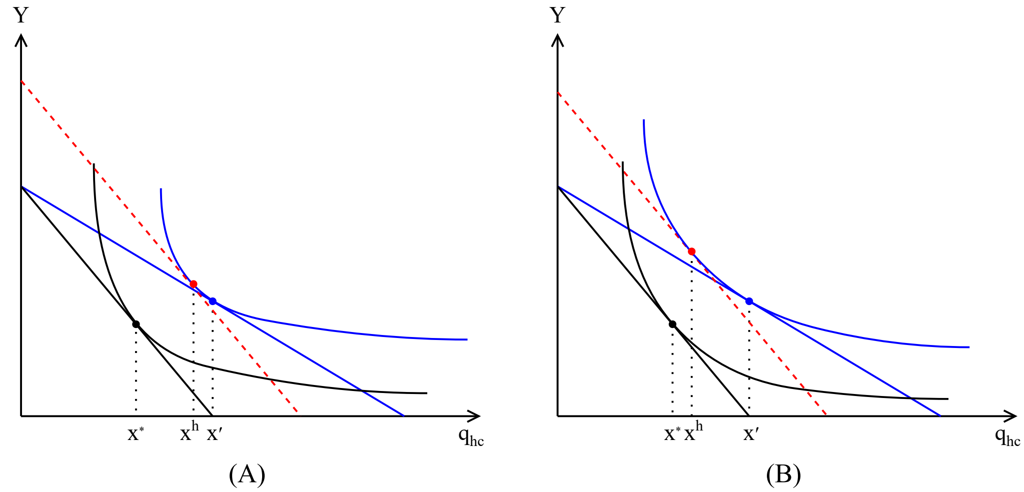
4.2.1. Income and substitution effect

To set up the frame in which our empirical approach is used, we first decompose the phenomenon in which the level of consumption changes in response to a change in the price of the same good into two parts: the income effect and the substitution effect. The two components of the total price effect are related to the two effects that consumers face when the price of a good changes. Consider the case of a decrease in purchasing price of health care due to the consumer having access to health insurance. First, because the price of health care decreased, now the consumer has access to additional purchasing power and is free to spend it however they want. Given that health care is a normal good, the consumption of health care will increase alongside with overall increase in the consumption of all other normal goods. The income effect refers to such increase in the level of consumption, represented by movement from x^* to x^h in Figure 4.1, that is the result of the increase in the overall purchasing power of the consumer. It is important to note that the portion of the total price effect that is attributable to the income effect does not take distortions in prices into account, rather, it is simply the response of the consumer facing increased purchasing power.

Second, because the price of health care decreased, now the consumer faces a choice set in which health care becomes relatively cheaper than all other goods, making it more attractive to the consumer. The resulting increase in the consumption of health care is referred to as the substitution effect. While there are different ways of identifying the substitution effect, we present the method using equivalent variation, which is a deliberate choice given the empirical setting that we exploit in this paper. Using the equivalent variation approach, the magnitude of the substitution effect is the difference between the level of consumption after the price change, represented by x' in Figure 4.1, and the level of consumption the consumer would have chosen under original prices but if they were given lump sum transfers that made them as satisfied as after the price change, represented by x^h. As seen in Figure 4.1, the sum of the income and the substitution effect equals the total price effect.

Note that both panel (A) and panel (B) in Figure 4.1 portray the same amount of increase in the level of consumption caused by the same amount of decrease in price. The key difference between the two cases is the degree to which the total price effect can be explained by the income effect. In panel (A), the income effect is the dominating force as most of the increase in the level of consumption would have occurred even if the consumer were simply given a lump sum transfer. On the other hand, in panel (B), the income effect explains very little of the total price effect, and the substitution effect is the dominating force. Most of the increase in the level of consumption only occurs due to changes in the relative prices of health care

Figure 4.1: Indifference curve decomposition of the price effect

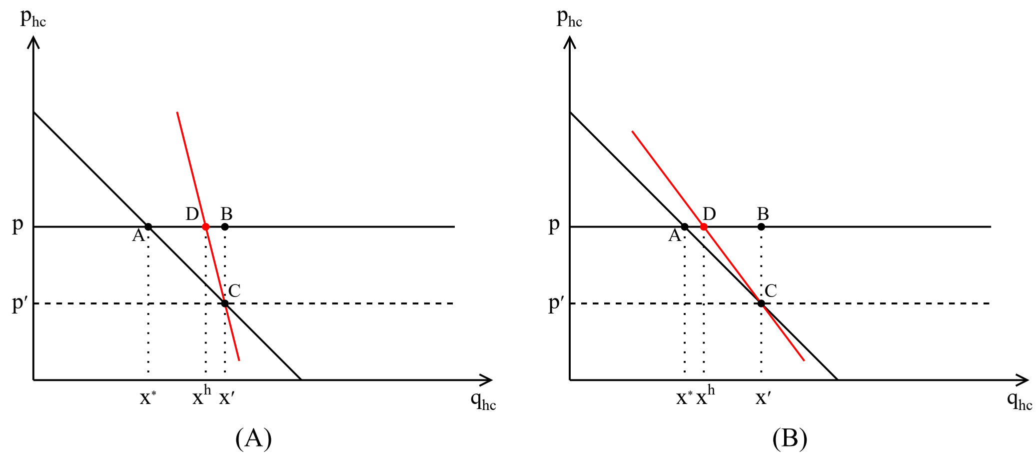


against all other goods. In this case, if the consumer were given a lump sum transfer, they would spend minimal amount on seeking additional health care.

4.2.2. Welfare implications

The welfare implications of the two scenarios depicted in the two panels can be seen easily by the resulting Marshallian and Hicksian demand curves following the decomposition approach using the equivalent variation. Figure 4.2 presents the two types of demand curves that are the respective results of the different shapes of the indifference curves depicted in panel (A) and (B) of Figure 4.1. Note that the slopes of the Marshallian demand curves are the same across the two panels because a Marshallian demand curve plots the overall changes in the level of demand following changes in the price. On the other hand, the slopes of the Hicksian demand curves are determined by the relative magnitude of the income effect as the Hicksian demand curve plots the level of demand holding utility fixed at the post-price change level. While a Hicksian demand curve will always be weakly steeper than a Marshallian demand curve, if the income effect is the dominating force as in panel (A) of Figure 4.2, the Hicksian demand curve will have a much steeper slope that departs away from the Marshallian demand curve. On the other hand, if the substitution effect is the dominating force as depicted in panel (B), the Hicksian demand curve will have a relatively

Figure 4.2: Hicksian demand curve decomposition of the price effect



similar slope as the Marshallian demand curve.

The critical welfare implication of the two alternate scenarios is the difference in magnitudes of deadweight loss. The deadweight loss must be calculated by using the Hicksian demand curve (Hausman, 1981) because the portion of the total price effect explained by the income effect is not caused by distortions in price, and therefore only the portion of the total price effect explained by the substitution effect has bearings on market efficiency. Therefore the degree of market inefficiency can be represented by the triangle BCD in Figure 4.2. Depending on which panel in Figure 4.2 we believe better represents the market for health care, magnitude of deadweight loss caused by moral hazard can be hugely different. If the income effect dominates as in panel (A), most of the changes in consumer behavior is, in fact, optimal. There is little deadweight loss from health insurance because it is the increases in consumption power that drives increases in health care consumption, rather than the distortion in relative price of health care. On the other hand, if the substitution effect dominates as in panel (B), most of the changes in consumer behavior results in deadweight loss. The suboptimal level of consumption is caused by the fact that health care simply becomes relatively cheaper, and health insurance induces overconsumption of health care that individuals would not have consumed if they were given a lump sum transfer instead.

4.2.3. Empirical strategy

Since the decomposition involves three components, total price effect, income effect, and substitution effect, the estimation of the decomposition of the price effect will require estimating at least two of the components. Directly estimating the substitution effect is a challenge because the estimation process would involve observing changes in consumer behavior holding the level of utility fixed. Given that utility is not easily measurable, a better empirical strategy would be to estimate the total price effect and the income effect to make inference on the magnitude of the substitution effect. One way to directly estimate the income effect would be to observe a market where consumers were randomly assigned to insurance plans that give out lump sum transfers. Observing the baseline expenditure, the amount of lump sum transfer, and the resulting changes in the level of health care expenditure would allow the econometrician to infer the magnitude of the income effect by measuring the proportion of each unit of additional income that is spent in health care. If the magnitude of the total price effect can be measured in the same market, for example by randomly assigning consumers to insurance plans that reduce the purchasing price of health care, then the measurement of the total price effect and the income effect can be used to infer the magnitude of the substitution effect, i.e., the portion of moral hazard that results in deadweight loss.

Suppose there is a market where individuals are randomly assigned to one of the following: (1) insurance that reduces the purchasing price of health care, (2) insurance that provides lump sum transfers in case the consumer requires health care, and (3) no insurance. Given that we observe the level of health care expenditure of each individual as well as the amount of insurance payments that individuals received from each type of insurance, we can estimate the following models.

$$y_{it} = \beta_0 + \beta_1 \times \textit{payments from insurance type 1}_{it} \\ + \beta_2 \times \textit{payments from insurance type 2}_{it} + \mathbf{X}_{it}\gamma_{it} + \epsilon_{it} \tag{4.1}$$

$$\log(y_{it}) = \delta_0 + \delta_1 \times \textit{enrolled in insurance type 1}_{it} \\ + \delta_2 \times \textit{enrolled in insurance type 2}_{it} + \mathbf{X}_{it}\gamma_{it} + \epsilon_{it} \tag{4.2}$$

where y_{it} measures the individual's level of health care expenditure, and $\mathbf{X}_{it}$ includes individual characteristics such as age, gender, overall level of health, and binary variables indicating insurance enrollment in each type. Note that y_{it} is not a measure of out-of-pocket spending, rather it measures the level of expenditure, i.e., spending on health care before insurance. *payments from insurance type 1* measures the amount of payments that individuals received from price-reducing insurance. In other words, if such insurance plan reimburses 90% of spending on health care and a given individual's spending was \$100, then the amount of payment from the insurance plan would be \$90. *payments from insurance type 2* measures the amount of payments that individuals received from plans with lump sum transfer contracts. Such plan would involve pre-specified amount of lump sum transfer that triggers when an individual ends up in a "bad" state.

With careful interpretation, the coefficients estimated from models (4.1) and (4.2) allow us to recover the decomposition of the total price effect. First, note that β_1 measures the amount of increases in health care expenditure per each dollar of payment from a price-reducing insurance plan. Defining $\bar{y}$ as the average level of health care spending by an individual with no insurance and r as the proportion of total spending to be covered by insurance, we can intuitively write β_1 as follows.

$$\beta_1 = \frac{y - \bar{y}}{ry} \tag{4.3}$$

Note that $r = 1 - p'$, where p' is the relative price of health care after insurance. We can then rewrite (4.3) as follows.

$$\beta_1 = \frac{y - \bar{y}}{y} \times \frac{1}{1 - p'} \tag{4.4}$$

Then, we use δ_1, which measures the percentage change in health care expenditure for those enrolled in price-reducing health insurance, as follows.

$$\begin{aligned}
\beta_1(\delta_1 + 1) &= \frac{y - \bar{y}}{y} \times \frac{1}{1 - p'} \times \Big(\frac{y - \bar{y}}{\bar{y}} + 1\Big) \\
&= \frac{y - \bar{y}}{\bar{y}} \times \frac{1}{1 - p'} \\
&= \frac{\text{\% change in health care expenditure}}{\text{\% change in price of health care}}
\end{aligned} \tag{4.5}$$

Therefore, $\beta_1(\delta_1 + 1)$ represents the Marshallian price elasticity of demand for health care.

On the other hand, β_2 measures the amount of increases in the level of health care expenditure per one dollar of lump sum transfer by an insurance plan. Note that such lump sum transfer does not affect the price of health care in any way. Defining T as the amount of lump sum transfer by insurance, we can write β_2 as follows.

$$\beta_2 = \frac{y - \bar{y}}{T} \tag{4.6}$$

By multiplying (4.6) by $\frac{w\bar{y}}{w\bar{y}}$, where w is the level of income, we can rewrite (4.6) as follows.

$$\begin{aligned}
\beta_2 &= \frac{\bar{y}}{w} \times \frac{y - \bar{y}}{\bar{y}} \times \frac{w}{T} \\
&= \alpha \times \frac{\text{\% change in health care expenditure}}{\text{\% change in income due to lump sum transfer}},
\end{aligned} \tag{4.7}$$

where α is the proportion of total income spent on health care. Therefore, β_2 represents the income elasticity of demand for health care multiplied by proportion of total income spent on health care.

Using the Slutsky equation, the Marshallian price elasticity is,

$$\eta_p = \eta_p^h + \alpha \eta_w, \tag{4.8}$$

where η_p^h is the Hicksian compensated price elasticity and η_w is the income elasticity. By substituting the results in (4.5) and (4.7), we can infer the Hicksian compensated price elasticity.

$$\eta_p^h = \beta_1(\delta_1 + 1) - \beta_2 \tag{4.9}$$

Therefore, given the ability to estimate the consumer response to price-reducing insurance plans and income-transferring insurance plans as modelled in (4.1) and (4.2), we can use the estimates to infer how much of moral hazard caused by price-reducing health insurance is in fact welfare-reducing.

4.3. Empirical Setting

The practical problem of estimating the decomposition is that insurance plans that involve lump sum transfers has been extinct at least in the U.S. for quite some time. Therefore, we exploit the private insurance market in South Korea where there exists insurance contracts with lump sum transfers that are designed to trigger when consumers are diagnosed with pre-specified conditions. In this section, we outline the basic institutional details of the health insurance market in South Korea, provide an overview of the private insurance market, and discuss insurance plans offered in the private insurance market and how they fit our setting.

4.3.1. The National Health Insurance in South Korea

The health care system in South Korea can be characterized by the universal coverage delivered through the National Health Insurance (NHI) system.[7] All Korean citizens, legal residents, and foreign nationals who reside in Korea for more than six months are legally required to enroll in the NHI, resulting in approximately 97% coverage of the entire population. Contributions to the NHI are proportional to wage (around 5%) for employees of

[7]We refer readers who are interested in further details on the National Health Insurance and the general health care system of South Korea to (Kwon et al., 2015)

private firms and organizations, government, and private and public schools. For the self-employed, contributions take both the level of income and the value of total assets into account. There is only a single uniform plan which provides all enrollees with access to the identical benefits package.

The benefits are explicitly defined in a positive list and covers essential health care services including diagnostic services, treatments, emergency care, pharmaceuticals, and dental care as well as medical check-ups. Quasi-government agencies decide the set of services and drugs to be included in the package mostly based on cost-effectiveness criteria and budgetary concerns. Figure 4.3 outlines the coinsurance rates for the services that are covered by the NHI. The coinsurance rate ranges from 20% for inpatient care to 35-60% for outpatient care depending on the type of provider. For example, the coinsurance rate for a consultation visit at a hospital outpatient is set higher than the same visit provided at a physician clinic in order to induce people to utilize low cost options first. There is an out-of-pocket ceiling with the amount depending on income. However, the ceiling only applies to services that are covered by the NHI, and therefore spending on care not covered by the NHI are not subject to any financial protection.

With limited amount of funding, covering the entire population was and still is the priority more so than the scope of services that are covered. Limited funding, together with the strict positive system of coverage, leads to many services and procedures to be left out of the NHI's coverage. Services and procedures that are not covered by the NHI generally fall into two categories: diagnostic/curative services and elective services. The former category includes medically necessary procedures that are not evaluated yet or deemed not as cost-effective. It also includes services that are conditionally covered, such as MRIs where only up to a certain number of tests are covered. Spending on medical care in this category is often a source of financial burden for consumers. The latter category includes services that are deemed to be non-medically necessary, e.g. elective cosmetic surgeries.

Together with non-covered services and relatively high coinsurance rate, the percentage of

Figure 4.3: Coinsurance Rates

User charges for health services

Health Service	Type of user charge in place	Exemptions and/or discounted rates	Cap on OOP spending	Protection mechanisms for children and the elderly	Other protection mechanisms
Public Health Center	Fixed rate: 30%	Flat amount (1100 KRW + prescription fee 500 KRW) when total expenditure is less than 120 000 KRW	Cap on OOP payment for 6 months - for lower income percentile 50%: 2 million KRW - for middle income percentile 30%: 3 million KRW - for higher income percentile 20%: 4 million KRW		**Reduced Co-payment Rate** - Severe disease patients (e.g. cancer) in hospital, outpatients and prescriptions are subject to 5% co-payment for 5 years from registration - for chronic renal failure: 10% - for cardio-cerebrovascular patients getting operations: up to 5% (maximum length of inpatient stay: 30 days) - for severe burns: up to 5% - for "unregistered" cancer patients: 20% - for "unregistered" rare and incurable disease: 30–60%
Physician Clinics (Primary care)	Fixed rate: 30%			For the elderly over 65 years, flat amount (1500 KRW) if total expenditure is less than 15000 KRW	
Outpatient care units of hospitals (Outpatient specialist visit)	Fixed rate: 40% (hospital), 50% (secondary hospital), 60% (tertiary hospital)	For rural areas, fixed rate is reduced by 5% (i.e. 35% for hospital, 45% for secondary hospital in rural areas)			
Outpatient prescription drugs	Fixed rate: 30%	Higher rate (40% for secondary hospital, 50% for tertiary hospital) for minor diseases that do not require upper-level hospital care		For those over 65 years, flat amount (1200 KRW) if total expenditure is less than 10 000 KRW	
Inpatient stay	Fixed rate: 20%			For children under 6, fixed rate: 10%	

Source: NHIS, 2014

total medical expenditure that the NHI accounts for has been fairly low. According to the *2017 Korea Healthcare Quality Report* by the Korea Institute for Health and Social Affairs (Kang et al., 2017), the government and the NHI spending only accounted for 56% of the total health care expenditure in 2016. It is estimated that 37% of the total health care expenditure came from out-of-pocket payments, which is fairly high when compared to the OECD average of 20% despite having universal coverage. Private health insurance accounted for the remaining 7% of total health care expenditure.

4.3.2. The private health insurance market

One of the solutions that consumers have turned to alleviate high burden of out-of-pocket costs is private health insurance that provides coverage in addition to the mandatory insurance. There is no clear data on the size of the private health insurance market due to the difficulty of collecting proprietary data across multiple firms, but in 2016 around 73% of total households were estimated to have private health insurance (Choi and Lee, 2017). The general consensus is that the private health insurance market is still growing.

The most critical aspect of the private health insurance market that we exploit in this paper is the two types of bases in which insurance payments are reimbursed to the consumers. Insurance plans sold in the private insurance market can be broadly categorized into plans that pay a pre-specified amount of money based on various triggers and plans that provide reimbursements based on the actual spending incurred by the consumer. In this paper, we define *fixed indemnity plans* as the plans that have a predefined set of triggers and associated set amount of money to be paid out to the consumers. Conversely, we define *supplementary plans* to be the plans that reimburse consumers based on how much they actually spent on eligible medical care.

First, supplementary insurance is the type of health insurance plans that consumers in the U.S. are familiar with. These plans provide reimbursements for out-of-pocket medical expenses including copays and coinsurance for services covered by the NHI as well as medically necessary services and procedures that are not covered by the NHI. In other words, supplementary insurance reimburses consumers for coinsurance paid on all medically necessary procedures and services regardless of whether they are covered by the NHI or not. Supplementary plans do not provide coverage for non-medically necessary elective procedures, such as plastic surgery performed for cosmetic reasons. Unlike the NHI and health insurance plans sold in the U.S., supplementary plans in South Korea usually have annual maximum reimbursements ranging from approximately \$30,000 to \$100,000 for inpatient care and \$100-\$500 for each outpatient visit. Supplementary plans themselves also incorporate measures of coinsurance. For inpatient visits, the patient is expected to pay 10-20% of total reimbursable expenses. For example, for an individual who purchased a plan with 10% coinsurance rate, if the total inpatient cost was \$5000 and the NHI covered \$4000, the patient first pays \$1000 to the health care provider, then the patient can submit a reimbursement claim for \$900, lowering the net out-of-pocket payments to \$100. Similarly, for outpatient visits and prescription drugs, different rates of copays apply, usually in the range of \$10-\$20 (Shin, 2015). In essence, the behavior of the consumers with supplementary insurance allows us to estimate the total price effect of health insurance in the sense that these plans distort the purchasing price of health care exactly through the same mechanism as all insurance plans sold in the U.S. market.

Fixed indemnity insurance plans have been the more popular form of insurance in South Korea. These are contracts that pay a pre-specified amount of money contingent on various triggers that range from diagnosis of certain condition to certain types of surgery. For example, a fixed indemnity plan with a cancer diagnosis trigger will pay the pre-specified amount of money when the policy holder is diagnosed with cancer. The condition that is the basis of the contract can vary from cancer, acute myocardial infarction, cerebral hemorrhage, to bone fracture. There are also plans with procedural triggers that pay when policy holders receive pre-defined types of surgery, which can be combined with a restriction contingent on the diagnosis of a pre-specified condition. Most of the plans that are offered in the market have core packages that cover more serious conditions, such as cancer, AMI, and stroke. Once consumers pick the base coverage, they can often choose add-on coverage on other conditions including, but not limited to, bone fractures, heart conditions, arthritis, dementia, and dentures/implants. Some contracts also pay a fixed amount of money per hospital stay, either covering all or a specific set of diagnoses.

Critical to our study, fixed indemnity payments based on changes in the state of the consumers, e.g. diagnosis of a medical condition, do not affect the price of health care that the consumers face in any way. Payments from these contracts are not contingent on utilization (except the visit and any necessary diagnostic services to confirm the diagnosis), and consumers are free to spend the money in any way they see fit. The proportion of indemnity payments that consumers decide to spend on health care provide us with the ability to directly estimate the magnitude of the income effect and in turn the ability to disentangle the total price effect of health insurance.

4.3.3. Adverse selection in the private health insurance market

While we perform extensive statistical tests in Section 4.6 to confirm that there is minimal bias from adverse selection, there are institutional details of the private health insurance market that allows to explain the results of our tests. First, consumers are legally required to disclose previous medical history and utilization records when applying for private insurance plans. Using the information, the private insurance companies are freely allowed to employ differential pricing, reject applications, and deny coverage for pre-existing conditions. For example, if an insurance company decides an applicant is high-risk but does not warrant a complete rejection, it can raise premiums, reduce benefits, and restrict coverage based on pre-existing conditions (Shin, 2015).

In practice, insurance firms usually dispatch investigators when a consumer files a claim within three years of enrollment. Consumers are contractually required to allow the investigators to access their medical history to a certain degree, otherwise the insurance company can postpone or reject any claims, and the investigator pays close attention to any violation in the disclosure requirements. If the investigator discovers evidence of non-disclosed pre-existing conditions, the insurance company can not only deny the claim but can also cancel the enrollment altogether. Ultimately, these tools not only allow both low-risk and high-risk individuals to stay in the market but also resolve adverse selection not by relying on self-sorting, e.g., by offering a menu of plans that offer varying degrees of financial protection, but through the firm essentially offering a personalized plan with premiums and coverage determined for each individual. Such lack of regulation, compared to other countries that adopted various systems of community rating, allows us to rule out issues rising from selection bias to the degree in which insurance firms have access to information.

Furthermore, the mandatory NHI alleviates at least some portion of the health risk that individuals face, and therefore the degree of adverse selection arising from private information on health risk should be lower in the private insurance market. In contrast, insurance markets where plans that are sold are expected to be the primary source of financial protection face potentially larger effect of adverse selection. For instance, the insurance market for individuals who do not have access to employer-based government-based plans in the U.S., especially before the reform in 2010, has been known to suffer from adverse selection as evidenced by higher average risk and higher premiums (Browne, 1992; Hackmann et al., 2015). In this market, the only way for high-risk individuals without access to group insurance plans to acquire financial protection is to enroll in individual health insurance plans, ultimately driving both the average risk and premiums higher. The idea is that if these individuals had access to basic insurance plans offered by the government, the financial risk that consumers face, on average, would be lower, and therefore the degree in which adverse selection occurs would be curtailed.

4.4. Data and Summary Statistics

4.4.1. Data

The data used in this paper were provided by the Korea Health Panel, an annual survey data jointly administered by the Korea Institute for Health and Social Affairs and the National Health Insurance Service. Being nationally representative, the Korea Health Panel data have been used in numerous studies across disciplines as well as policy designs and evaluations.

It began tracking 7,866 households and 24,616 individuals in 2008. By 2016, the latest data available, the retention rate was approximately 65%. To maintain adequate number of observations as well as representativeness, 2,520 households and 7,387 individuals were added in 2012 (the retention rate for the new group in 2016 was approximately 75%). We drop data before 2010 because important covariates including self-reported health began to be collected from 2010. Also, we exclude 2016 data since there can be time lags between when consumers receive health care and submit claims to private health insurance companies, and therefore a big portion of claims for care received in 2016 has not been included. From 2010 to 2015, there are total of 102,980 observations, and for our analysis we drop observations with missing data on individual's health and household characteristics, resulting in the final sample size of 72,925. For the purposes of this paper, we construct a dataset consisting of comprehensive variables related to health care consumption, private health insurance enrollment and reimbursements, and sociodemographic characteristics. All monetary variables are converted into 1,000 South Korean won (KRW). For simplicity, we refer to 1,000 KRW as 1 U.S. dollar (USD) when interpreting our results.[8]

As the main outcome variable, we use the total amount of each individual's spending on health care including inpatient, outpatient, emergency services, and prescription drug utilization. Note that the amount that we observe is the total amount of spending after NHI coverage, but not taking into account any reimbursements from private insurance, because the NHI preemptively adjusts the purchasing price of health care. In contrast, all private health insurance plans operate under a strict ex-post reimbursement system. Patients need to submit evidence, such as payment receipts and/or medical records, to the insurance company to receive reimbursements in an ex-post fashion after the actual utilization occurs. Therefore, the compulsory nature of the NHI implies that the total amount of health care spending that we observe essentially measures the total health care expenditure of each individual. As a robustness check, we also perform estimations using spending at the household level. Similar to the individual spending, the household spending is calculated by

[8] At the time of writing, the U.S. dollar to South Korean won exchange rate is at 1133.60.

aggregating health care spending of all members of the household. The additional benefit of confirming our analysis using the household-level variable is that it includes the amount of health care related expenses, such as transportation costs, midwives, long-term care, and over-the-counter drugs.

Private health insurance enrollment is tracked annually for all individuals that are part of the data. Although the specific plan that an individual is enrolled is tracked using a masked plan ID, it is not identifiable, i.e., we cannot look up plan details such as which insurance company offered the plan. However, plan type (indemnity versus supplementary) associated with each plan ID is available, giving us the ability to track whether an individual was enrolled in a plan offering indemnity insurance and/or a plan offering supplementary insurance. Data on the amount as well as the type of payments received from private health insurance are available at the plan level for each individual, enabling us to observe how much each individual received from private insurance by three categories: fixed indemnity payments based on changes in state of the individual (i.e., diagnosis, accidents, etc.), fixed indemnity payments triggered by utilization (i.e., fixed amount per hospital day), and supplementary payments on actual medical spending.

Detailed data on socioeconomic characteristics allow us to control for a rich set of covariates. We control for gender, age, marital status, employment status, years of education, logged household size, logged household income, and city/province of residence as sociodemographic characteristics. To control for variations in underlying health conditions and lifestyles that may affect both purchasing decisions of private health insurance and health care consumption, we control for whether the individual had a childbirth in the year of the survey, presence of chronic conditions, disability, physical limitation, days engaged in vigorous or moderate physical activities, days walked for more than ten minutes, and self-reported health status structured as a five-level Likert scale. Lastly, we exploit the panel nature of the data and include individual fixed effects to control for time-invariant unobserved individual characteristics.

Table 4.1: Summary statistics

	Mean	SD	Min	Max
Individual spending	601.8	1250.7	0	52484.3
Enrolled in supplementary insurance (%)	29.5	45.6	-	-
Enrolled in indemnity insurance (%)	67.9	46.7	-	-
Supplementary insurance payments	18.8	273.5	0	35000
Indemnity payments based on state	68.2	2012.2	0	400000
Indemnity payments based on utilization	30.0	569.3	0	110000
Female (%)	54.8	49.8	-	-
Age	51.0	16.9	17	102
Married (%)	71.1	45.3	-	-
Employed (%)	59.7	49.0	-	-
Years of education	11.2	4.28	0	19
Household income	42917.6	33074.3	20	1036000
Number of household members	3.24	1.31	1	11
Childbirth (%)	0.94	9.67	-	-
Chronic conditions (%)	61.8	48.6	-	-
Disabled (%)	6.21	24.1	-	-
Physical limitation (%)	5.56	22.9	-	-
Days in a week engaged in vigorous physical activity	0.76	1.67	0	7
Days in a week engaged in moderate physical activity	1.49	2.26	0	7
Days in a week walked for more than ten minutes	3.94	2.74	0	7
Self-reported health				
Very good (%)	5.85	23.5	-	-
Good (%)	37.7	48.5	-	-
Moderate (%)	41.2	49.2	-	-
Bad (%)	13.8	34.4	-	-
Very bad (%)	1.45	12.0	-	-

4.4.2. Summary statistics

Table 4.1 provides the summary statistics of the variables used in the analysis as well as the unconditional probabilities of enrollment in private health insurance plans by type. 67.9% and 29.5% of the total population respectively was enrolled in indemnity insurance plans and supplementary insurance plans. The summary statistics confirm previous research and show that the enrollment rate for the indemnity insurance plans are much higher than the supplementary insurance. Average payments from private insurance plans shown in Table 4.1 are calculated over the entire population as opposed to only those enrolled in each type of insurance.

Table 4.2 provides the summary statistics by enrollment in each type of insurance. Compar-

Table 4.2: Summary statistics by enrollment in each type of private insurance

	(1) Indemnity		(2) Supplementary	
	Mean	SD	Mean	SD
Individual spending	573.5	1225.5	554.4	1207.5
Female (%)	55.1	49.7	57.1	49.5
Age	47.2	14.1	43.3	12.8
Married (%)	76.2	42.6	75.7	42.9
Employed (%)	66.3	47.3	68.1	46.6
Years of education	12.1	3.62	12.9	2.97
Household income	49129.2	33945.9	53598.4	33724.7
Number of household members	3.43	1.21	3.59	1.13
Childbirth (%)	1.17	10.7	1.67	12.8
Chronic conditions (%)	56.8	49.5	50.2	50.0
Disabled (%)	3.84	19.2	1.98	13.9
Physical limitation (%)	2.91	16.8	1.65	12.7
Days in a week engaged in vigorous physical activity	0.86	1.74	0.88	1.74
Days in a week engaged in moderate physical activity	1.60	2.27	1.59	2.23
Days in a week walked for more than ten minutes	3.92	2.70	3.90	2.69
Self-reported health				
Very good (%)	6.18	24.1	6.71	25.0
Good (%)	40.7	49.1	42.9	49.5
Moderate (%)	42.1	49.4	42.2	49.4
Bad (%)	10.2	30.3	7.77	26.8
Very bad (%)	0.72	8.47	0.40	6.31

ing the average statistics of individuals enrolled in indemnity and supplementary insurance shows that the two groups are relatively similar in terms of most covariates including gender composition, marital status, education, household size, and self-reported health. Individuals who purchased supplementary insurance are slightly younger, have higher income, and are less likely to be disabled or have physical limitations. Compared to the population average, being younger, married, employed, having more education and higher household income, and being in a larger household are all correlated with private insurance enrollment. On the contrary, having a chronic disease, being disabled, and having limitations in daily activities are all negatively correlated with private health insurance enrollment. The negative correlation is the opposite of what economic theory of adverse selection would predict. This pattern is again evident in the self-reported health variable, as individuals with the worst health have the lowest rates of insurance enrollment. These results are likely due to the fact that the insurers are relatively successful in preventing adverse selection by being able to deny applicants without any restrictions, deny specific coverages based on pre-existing conditions, and transfer any calculated risk to the consumer by increasing the premium.

4.4.3. Actuarial values of private insurance plans

The estimated actuarial values of the two types of private insurance plans can be calculated by taking the sample mean of the amount of insurance payments that enrollees receive. First, we find the likelihood of receiving insurance payments from an indemnity plan to be 4.87% given enrollment. More specifically, 2.15% of indemnity plan enrollees received payments triggered by changes in state, and 3.09% of enrollees received payments triggered by health care utilization. As for supplementary plans, 9.37% of enrollees received payments. The higher likelihood of receiving payments from supplementary plans is expected because indemnity plans require more specific conditions to trigger payments.

Conditional on receiving any amount of payment from respective type of insurance, individuals received $4,673 from indemnity payments triggered by state, $1,430 from indemnity payments triggered by utilization, and $681 from supplementary payments. Again, the larger amount of average payments from indemnity plans is expected due to the nature of the differences in the type of insurance. Combining the likelihood of receiving any amount of payment and the conditional average amount of payments, the estimated average actuarial value of indemnity plans based on state is \$100.44, \$44.16 for indemnity plans based on utilization, and \$63.79 for supplementary plans.

4.5. Results

Table 4.3 presents the result of the estimation of model (4.1). We find that a one dollar of insurance payment from a supplementary plan is associated with \$0.54 increase in health care expenditure. On the other hand, a one dollar of payment from indemnity insurance triggered by changes in the state of an individual is associated with \$0.04 increase in health care expenditure. We present the result of the estimation of model (4.2) in Table 4.4. We find that enrollment in a supplementary insurance plan is associated with 19.84% increase in health care expenditure by exponentiating the point estimate. In other words, we estimate β_1, β_2, and δ_1 to be 0.536, 0.0404, and 0.181 respectively.

Using the Slutsky equation and the formulation in (4.9), we estimate the Hicksian compensated price elasticity of health care to be 0.593. In comparison to the total uncompensated price elasticity, which we estimate to be 0.633, we find that 94% of moral hazard caused by health insurance can be attributed as market inefficiency.

4.5.1. Nonlinear marginal effects

Our finding that 94% of moral hazard caused by health insurance leads to market inefficiency applies to the overall market for health care. However, at the individual level, it is plausible that both the magnitude of the total price effect and the decomposition into income and substitution effect can have large heterogeneity. For example, the demand for health care would be much more inelastic for individuals with a more serious condition compared to the average consumer. As a first test to detect such variation across the population, we perform nonlinear analysis by estimating polynomial specifications of model (4.1). Specifically, we fit a model with quadratic terms of payments from each type of private insurance plan. We

Table 4.3: Estimation of model (4.1)

	(1) Health care expenditure	(2) Health care expenditure
Supplementary insurance payments	0.572***	0.536**
	(0.170)	(0.173)
Indemnity payments based on state	0.0455	0.0404
	(0.0257)	(0.0235)
Control variables	Yes	Yes
Individual fixed effects	No	Yes
Observations	72925	72925

Note. Robust standard errors in parenthesis. Controls not shown include sex, age, marital status, employment status, years of education, logged household income, logged household size, childbirth, chronic conditions, disability, physical limitations, days engaged in moderate physical activities, days engaged in vigorous physical activities, days walked for more than ten minutes, self-reported health, city/province of residence, and year fixed effects. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 4.4: Estimation of model (4.2)

	(1) Logged health care expenditure	(2) Logged health care expenditure
Enrolled in supplementary insurance	0.304***	0.181***
	(0.0191)	(0.0328)
Enrolled in indemnity insurance	0.296***	0.0582
	(0.0186)	(0.0413)
Control variables	Yes	Yes
Individual fixed effects	No	Yes
Observations	72925	72925

Note. Robust standard errors in parenthesis. Controls not shown include sex, age, marital status, employment status, years of education, logged household income, logged household size, childbirth, chronic conditions, disability, physical limitations, days engaged in moderate physical activities, days engaged in vigorous physical activities, days walked for more than ten minutes, self-reported health, city/province of residence, and year fixed effects. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

present the results in Table 4.5.

As theory would suggest, the marginal effects of all types of insurance payments on health care consumption decrease as the amount of payments increase. One dollar of payment from supplementary insurance is associated with \$0.99 increase in health care expenditure when the amount of insurance payment is very small. The magnitude of the consumer response to supplementary insurance decreases until the amount of insurance payment reaches \$14,500 at which point the demand for health care becomes perfectly inelastic to price. On the other hand, one dollar of indemnity payment based on changes in the state of an individual is associated with \$0.11 increase in health care expenditure when the amount of insurance payment is very small. The consumer response to additional purchasing power on health care consumption decreases until the amount of indemnity payment reaches \$191,800 at which point the demand for health care becomes perfectly inelastic to income. Note that supplementary payment of \$14,500 and indemnity payment of \$191,800 are larger than the top 0.1% among all individuals who received any amount of payment from each insurance type.

This suggests that health care as a good does not become perfectly inelastic to either price or income except for individuals in extreme circumstances who are in the top 0.1% in terms of health care spending among all individuals with any level of health care consumption. Additionally, we compute the predicted amount of insurance payments from each type of insurance at different levels of health care spending by regressing the amount of insurance payments on health care spending among those who received insurance payments from each type of insurance. Then, using the marginal effects recovered in Table 4.5, we estimate the total uncompensated price elasticity and the income elasticity to infer the Hicksian compensated elasticity. This in turn allows us to measure the degree in which moral hazard from health insurance is welfare-reducing at different levels of health care spending. We find that even at the top 99 percentile in terms of health care spending among individuals with any level of health care consumption, 89% of the total price effect is caused by substitution

Table 4.5: Quadratic specification of model (4.1)

	(1) Health care expenditure	(2) Health care expenditure
Supplementary insurance payments	1.0264***	.9943***
	(.06479)	(.06714)
Supplementary insurance payments^2	-3.5e-05***	-3.4e-05***
	(3.4e-06)	(3.2e-06)
Indemnity payments based on state	.11416***	.10663***
	(.01472)	(.01421)
Indemnity payments based on state^2	-3.0e-07***	-2.8e-07***
	(3.8e-08)	(3.6e-08)
Control variables	Yes	Yes
Individual fixed effects	No	Yes
Observations	72925	72925

Note. Robust standard errors in parenthesis. Controls not shown include sex, age, marital status, employment status, years of education, logged household income, logged household size, childbirth, chronic conditions, disability, physical limitations, days engaged in moderate physical activities, days engaged in vigorous physical activities, days walked for more than ten minutes, self-reported health, city/province of residence, and year fixed effects. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

effect, and therefore results in market inefficiency.

4.5.2. Disease category

A more direct approach to measure variations in the magnitude of the total price effect and the decomposition of the price effect into income and substitution effect is to consider individuals with specific health conditions. Cancer, diabetes, heart diseases and stroke, hypertension, arthritis and osteoporosis, and chronic back conditions are selected based on topics of priority from CDC's initiative Healthy People 2020 and rate of prevalence in the South Korean population.

First, we identify individuals with positive spending related to each medical condition using diagnosis codes attached to health care utilization reports. Then, we estimate models (4.1) and (4.2) on subsets of data in which individuals had positive spending on each type of medical condition during the year. We summarize the results in Table 4.6. With the exception of individuals diagnosed with cancer in which we find that 59% of the total price effect of

Table 4.6: Subgroup analysis of individuals with positive spending in each condition

	N	(1) β_1	(2) β_2	(3) δ_1	(4) η_p^h/η_p
Cancer	2150	0.180	0.0822***	0.107	59%
		(0.104)	(0.0143)	(0.0792)	
Diabetes	4615	1.338***	0.122*	-0.0195	91%
		(0.285)	(0.0519)	(0.0472)	
Heart diseases and stroke	3490	0.733***	0.0720	0.0724	91%
		(0.142)	(0.0379)	(0.0574)	
Hypertension	14277	0.773***	0.169***	0.0869**	80%
		(0.185)	(0.0350)	(0.0279)	
Arthritis and osteoporosis	10211	0.995***	0.135**	0.0998**	88%
		(0.0916)	(0.0474)	(0.0372)	
Chronic back conditions	10361	1.085***	0.0922**	0.210***	93%
		(0.0969)	(0.0303)	(0.0336)	

Note. For individuals with diabetes, we assume $\delta_1 = 0$. Robust standard errors in parenthesis. Controls not shown include sex, age, marital status, employment status, years of education, logged household income, logged household size, childbirth, chronic conditions, disability, physical limitations, days engaged in moderate physical activities, days engaged in vigorous physical activities, days walked for more than ten minutes, self-reported health, city/province of residence, and year fixed effects. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

health insurance is welfare-reducing, the point estimates show that roughly 90% of the total price effect leads to market inefficiency across individuals with medical conditions with varying degrees of severity and acuteness. This result is consistent with our finding using quadratic models presented in Section 4.5.1, and it provides further evidence that a large portion of moral hazard caused by health insurance results in market inefficiency even among individuals with medical conditions.

4.6. Selection Bias

So far, we have interpreted the results in Section 4.5 as if individuals were randomly selected to be enrolled in each type of insurance. We are especially concerned about the estimate of δ_1. Because we would over-estimate δ_1 if adverse selection occurs for supplementary plans, we would also over-estimate the compensated price elasticity. On the other hand, if advantageous selection occurs in the market for supplementary plans, we would under-estimate δ_1 and therefore under-estimate the compensated price elasticity. While in an ideal world, we would have had the ability to randomly assign individuals to each type of insurance, we rely on the features of the private insurance market and utilize series of statistical tests to confirm that our estimation strategy does not suffer from significant bias from selection.

First, we exploit the fact that individuals are asked to rate their own health, which may contain private information about the health status of the individuals, and use this information to see if it is a strong predictor of private insurance enrollment. Secondly, we use residuals from the main models to detect any presence of unobserved factors, conditional on all of the control variables, that may affect both the decision on the level of health care consumption and enrollment in private insurance. Lastly, we identify a group of marginal consumers who are indifferent between the two different types of insurance, and we confirm our findings by performing the estimations using only the group of individuals whom we identify as marginal consumers. In all of our tests, we either do not find significant presence of selection bias or find that there is a slight advantageous selection in the supplementary insurance market, suggesting that our estimate of the compensated price elasticity and therefore the welfare-reducing proportion of the total price effect are most likely a lower bound.

4.6.1. Self-reported health as predictor of private insurance enrollment

Although private insurance firms in South Korea can ask for nearly all information regarding an applicant's health including previous utilization and medical history, there is one particular variable that the survey asks that the firms do not have access: self-reported health. Assuming that individuals partaking the survey were honest when they were asked to rate their overall health, self-reported health may contain information that is known to the individual but is not revealed to the insurer through previous utilization and/or diagnostic history. One way to test whether individuals are successful in withholding information from the firm and select into the "right" plan is to see if current and/or previous health status can predict private health insurance enrollment. Table 4.7 presents the result of this exercise and shows that self-reported health is not a particularly strong predictor of insurance enrollment. When considering the point estimates, individuals that indicated their health to be worse are in fact slightly less likely to enroll in both indemnity and supplementary

Table 4.7: Self-reported health as predictor of private insurance enrollment

	(1) Indemnity		(2) Supplementary	
Self-reported health lagged by one year				
Good	0.178***	(0.0485)	0.0142	(0.0442)
Moderate	0.146**	(0.0489)	0.000490	(0.0448)
Bad	-0.00449	(0.0558)	-0.156**	(0.0554)
Very bad	-0.124	(0.103)	-0.244	(0.136)
Self-reported health				
Good	0.100	(0.0511)	-0.0368	(0.0455)
Moderate	0.0829	(0.0515)	-0.0122	(0.0461)
Bad	-0.0522	(0.0584)	-0.143*	(0.0567)
Very bad	-0.170	(0.106)	-0.369*	(0.143)
Control variables	Yes		Yes	
Observations	54278		54278	

Note. Self-reported health: Excellent is the baseline. Robust standard errors in parenthesis. Controls not shown include sex, age, marital status, employment status, years of education, logged household income, logged household size, childbirth, chronic conditions, disability, physical limitations, days engaged in moderate physical activities, days engaged in vigorous physical activities, days walked for more than ten minutes, self-reported health, city/province of residence, year fixed effects, and individual fixed effects. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

insurance. This provides an evidence against any adverse selection, and helps us interpret our results on changes in health care expenditure as the response to insurance enrollment, rather than the effect of hidden information.

4.6.2. Correlation test

Another statistical test to detect selection bias resulting from unobserved factors can be performed using the residuals from model (4.1) presented in Table 4.3. The residuals of this estimation represent unobserved factors that are correlated with health care consumption after taking into account the main explanatory variables as well as all of the control variables. If we find that the residuals are a strong predictor of enrollment in private insurance, we would worry that there is an unobserved variable that enters both the spending decision as well as the decision to enroll in private health insurance, such as private information on risk type, resulting in selection bias. We present the result in Table 4.8 where we use the residuals from the model with individual fixed effects in Table 4.3 as a predictor of private insurance enrollment in a logit model. We find that the residuals from model (4.1) do not

Table 4.8: Correlation test using residuals from model (4.1)

	(1) Indemnity	(2) Supplementary
Residuals	-2.8e-06	1.2e-05
	(2.3e-05)	(1.9e-05)
Individual fixed effects	Yes	Yes
Model	Logit	Logit
Observations	9366	17734

Note. Robust standard errors in parenthesis. Controls not shown include sex, age, marital status, employment status, years of education, logged household income, logged household size, childbirth, chronic conditions, disability, physical limitations, days engaged in moderate physical activities, days engaged in vigorous physical activities, days walked for more than ten minutes, self-reported health, city/province of residence, and year fixed effects. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

have statistically significant association with enrollment in either type of private insurance, suggesting that selection bias is mitigated through features of the private insurance market alongside the available control variables.

4.6.3. Identifying the Marginal Consumer

Yet another possible way of detecting for any potential selection bias is to focus the analysis on the individuals whom we believe are the marginal consumers who are indifferent between purchasing indemnity and supplementary insurance. One method of identifying the marginal consumers is to compute the likelihood of each individual purchasing either type of insurance based on all of the covariates that are available, then selecting a subset of individuals with approximately equal likelihood of choosing either type.

An issue with implementing this approach is that the number of individuals with supplementary insurance is lower than the number of individuals with indemnity insurance. Therefore, a simple predictive model, such as logit or probit, on the full sample of individuals will be skewed towards indemnity insurance such that the likelihood of purchasing indemnity insurance will be higher than the likelihood of purchasing supplementary insurance. To mitigate this issue, we use propensity score matching with no replacement to first create a balanced subsample consisting of all individuals with only supplementary insurance and their matching counterpart with only indemnity insurance. The same control variables that are used

Table 4.9: Estimation of model (4.1) among marginal consumers

	(1) Health care expenditure	(2) Health care expenditure
Supplementary insurance payments	0.861***	0.799***
	(0.0927)	(0.0965)
Indemnity payments based on state	0.140***	0.142***
	(0.0269)	(0.00424)
Control variables	Yes	Yes
Individual fixed effects	No	Yes
Observations	2549	2549

Note. Robust standard errors in parenthesis. Controls not shown include sex, age, marital status, employment status, years of education, logged household income, logged household size, childbirth, chronic conditions, disability, physical limitations, days engaged in moderate physical activities, days engaged in vigorous physical activities, days walked for more than ten minutes, self-reported health, city/province of residence, and year fixed effects. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

in the main specification are used for matching. Using this method, a subsample of 2,453 observations in each group (total number of observations being 4,906 with additional 2,453 matched samples) is selected.

Next, we fit a logit model on this subsample to estimate the probability of each individual choosing either supplementary or indemnity insurance. To be consistent with the main specification, all of the control variables used in the main specification are used in the logit model. Then, individuals with close to equal probability of choosing the two types of insurance are selected to be the best representation of the marginal consumers. Table 4.9 presents the result on the final subsample of 2,549 individuals with probability of choosing either type of insurance ranging from 0.48 to 0.52. The results are consistent with the estimates using the full sample in both specifications including or excluding individual fixed effects.

4.7. Extensions and Robustness Checks

4.7.1. Including expenses related to health care consumption

The household spending variable captures total health care spending as well as relevant expenses, such as transportation costs, midwives, long-term care, and over-the-counter drugs. Therefore, performing estimations using the household spending as the dependent variable allows us to test the possibility that indemnity payments are used to cover health care related expenses. We present the results in Table 4.10. We find that the relationship between supplementary insurance payments and spending including relevant expenses is not statistically distinguishable from the same relationship with the spending variable that only focuses on direct health care expenses. This is expected as supplementary insurance payments are only based on direct health care spending, i.e., supplementary plans do not reimburse individuals for relevant expenses, such as transportation or over-the-counter drugs. We also find that indemnity payments are not associated with larger increases in health care spending even when accounting for related expenses. In other words, the results provide an evidence against the possibility that a significant portion of indemnity payments are used to cover health care related expenses.

Another extension that is possible using the household spending variable is to focus on households of size one. This allows us to capture individual spending including any relevant expenses without complicating the result with decisions of other household members. In our data, there are total of 1,527 unique single-member households. The results of the estimation performed in the subgroup is presented in the second row of Table 4.10. While the results should be taken with a grain of salt given the significant reduction in the number of samples, we do find that indemnity payments are marginally associated with higher increase in spending including related expenses compared to the increase in spending excluding related expenses. We find that 90% of the total price effect can be attributed to substitution effect, and therefore leads to market inefficiency.

4.7.2. Health status-dependent marginal effect

Another way to measure differential effect across individuals with different health status is to allow the effect of insurance payments and enrollment to differ across health status. For the purposes of this analysis, we group individuals who answered that their health is good or very good as "good" health and those who answered that their health is moderate, bad, or very bad as "bad" health. Then, we use this binary indicator of good health and interact it with

Table 4.10: Estimation using household spending including relevant expenses

	N	(1) β_1	(2) β_2	(3) δ_1	(4) η_p^h/η_p
Full sample	72848	0.628***	0.0496	0.0814***	93%
		(0.188)	(0.0281)	(0.0101)	
Single-member households	5581	0.611**	0.0692	0.122	90%
		(0.198)	(0.0423)	(0.0759)	

Note. Robust standard errors in parenthesis. Controls not shown include sex, age, marital status, employment status, years of education, logged household income, logged household size, childbirth, chronic conditions, disability, physical limitations, days engaged in moderate physical activities, days engaged in vigorous physical activities, days walked for more than ten minutes, self-reported health, city/province of residence, and year fixed effects. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 4.11: Interaction using self-reported health

	(1) β_1	(2) β_2	(3) δ_1	(4) η_p^h/η_p
Baseline	0.391*	0.0319	0.175***	93%
	(0.196)	(0.0218)	(0.0361)	
Baseline × good health	0.451	0.0518	0.0121	92%
	(0.236)	(0.0304)	(0.0381)	

Note. $N = 72,925$. Robust standard errors in parenthesis. Controls not shown include sex, age, marital status, employment status, years of education, logged household income, logged household size, childbirth, chronic conditions, disability, physical limitations, days engaged in moderate physical activities, days engaged in vigorous physical activities, days walked for more than ten minutes, self-reported health, city/province of residence, year fixed effects, and individual fixed effects. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

insurance payments and enrollment. The first row of Table 4.11 provides the result of the estimation on the baseline group, which include individuals in bad health. The second row provides the coefficients of the interaction terms. We compute the proportion of total price effect that is welfare-reducing for individuals with good health by adding the point estimate of the interaction term and the baseline term. First, the result shows that individuals with good health show much more elastic behavior in terms of both price and income, which is expected and confirms our findings in previous analyses. Second, even though the magnitude of the price effect is different between the two groups, the proportion that is welfare-reducing is surprisingly consistent. We find that 93% and 92% of the total price effect is associated with market inefficiency among individuals with bad and good health respectively.

Table 4.12: Individuals with income below poverty line

	N	(1) β_1	(2) β_2	(3) δ_1	(4) η_p^h/η_p
Income below poverty line	4072	0.813***	0.422***	0.428*	64%
		(0.128)	(0.107)	(0.202)	
Income above poverty line	68853	0.531**	0.0400	0.175***	94%
		(0.173)	(0.0233)	(0.0332)	

Note. Robust standard errors in parenthesis. Controls not shown include sex, age, marital status, employment status, years of education, logged household income, logged household size, childbirth, chronic conditions, disability, physical limitations, days engaged in moderate physical activities, days engaged in vigorous physical activities, days walked for more than ten minutes, self-reported health, city/province of residence, year fixed effects, and individual fixed effects. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

4.7.3. Individuals with income below poverty line

Next, we examine whether individuals with financial constraints have different price and income elasticities and consequently the welfare implication of moral hazard. We stratify the sample into two groups: (1) individuals whose household income was below the poverty line for at least one year during the data collection period and (2) individuals whose household income was never below the poverty line. The poverty line is set by the Ministry of Health and Welfare and is around $1,000 USD for single-member households and increases as the number of individuals in the household increases. Approximately 6% of individuals in the sample were ever considered to have income below the poverty line. Results are presented in Table 4.12, and we find that individuals with income below the poverty line have both higher price and income elasticities. Furthermore, we find that the proportion of the total price effect that can be attributed to the substitution effect is 64% among the individuals with income below the poverty line.

4.7.4. Inpatient, outpatient, vs. emergency services

We measure differential effect based on the type of health care service by dividing individuals' health care spending into expenses incurred for inpatient, outpatient, and emergency services. Table 4.13 shows that 93% and 95% of the total price effect on inpatient and outpatient care respectively can be attributed as substitution effect. We find that the price and income

Table 4.13: Estimation using spending in each category of care

	(1) β_1	(2) β_2	(3) δ_1	(4) η_p^h/η_p
Inpatient	0.413**	0.0358	0.216***	93%
	(0.135)	(0.0201)	(0.0179)	
Outpatient	0.152***	0.00929	0.252***	95%
	(0.0459)	(0.00541)	(0.0172)	
Emergency care	0.00289*	0.000648	0.0506***	79%
	(0.00125)	(0.000450)	(0.00862)	

Note. $N = 72,925$. Robust standard errors in parenthesis. Controls not shown include sex, age, marital status, employment status, years of education, logged household income, logged household size, childbirth, chronic conditions, disability, physical limitations, days engaged in moderate physical activities, days engaged in vigorous physical activities, days walked for more than ten minutes, self-reported health, city/province of residence, and year fixed effects. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

elasticities for emergency care are significantly smaller than elasticities for inpatient and outpatient care. Furthermore, we find that 79% of the total price effect is welfare-reducing, which is relatively smaller than what we find for inpatient and outpatient care.

4.8. Discussion and Conclusion

In this paper, we use the unique private health insurance market in South Korea to provide an empirical evidence on the degree in which moral hazard in health care utilization is welfare-reducing. Our estimate of the price elasticity of health care utilization in the South Korean market is 0.633, of which 94% can be attributed to the substitution effect. We conduct multiple analyses that estimate the decomposition of moral hazard across different consumer types and care settings. Our estimates are consistent with economic theory in that the welfare-reducing proportion of the total price effect is smaller for patients with severe and acute medical conditions and individuals with lower income. We also find that the deadweight loss from moral hazard is less of a concern in medical care provided in emergency care settings.

Our work has policy implications that are in line with the theoretical development around issues of *ex-post* moral hazard in health care utilization. Various measures that can curtail moral hazard, such as coinsurance and deductibles, have been the focal point of countless policy debates in the U.S. Our estimates of the decomposition of moral hazard support such concerns in that insurance-induced overconsumption of health care carries significant deadweight welfare loss in the general population. At the same time, our work also supports proponents of increased financial protection for individuals with lower income as well as individuals with severe and acute diseases as the welfare-reducing proportion of moral hazard is lower among these individuals.

CHAPTER 5

Conclusion and Future Directions

Each essay in this dissertation takes a unique perspective that together comprises a complex healthcare system. In Chapter 2, the focus was on the decisions of the regulator in managing the incentives of healthcare organizations around quality improvements. In evaluating the VBP program, the essay takes the current measures of quality as given and focuses more on designing the financial aspects of the incentive program. However, questions still remain as to the validity of the quality measures that are currently used. Quality, especially in healthcare, is multidimensional by nature. Therefore, future research needs to evaluate the effectiveness of different quality measures in distinguishing true underlying quality as well as incentivizing changes in organizational behaviors. For instance, mortality rates, while included in the VBP program, can be difficult to accurately measure, let alone improve. Further analysis should consider a more focused incentive program that targets a narrower, strategically selected sets of quality measures.

Furthermore, any unintended negative consequence of such incentive programs should be analyzed. The operations management literature has well documented the "speed-quality" tradeoff similar to the idea of the iron triangle in health economics. However, the possibility of a "quality-quality" tradeoff where improving one aspect of quality can lead to decreases in, or be at the expense of, another aspect of quality. Ultimately, the response of healthcare organizations on incentive programs is a resource allocation problem. Therefore, careful analysis is needed to make sure that placing financial or non-financial incentives on very specific quality measures does not harm quality in other aspects, possibly manifesting through different patient types or procedures that are not directly incentivized. On the other hand, such analysis could in fact reveal that there is a positive spillover effect, where a healthcare organization's effort to improve quality in one aspect can indirectly lead to quality improvements in other aspects. Understanding such spillover effects can be an important factor in designing incentive programs in the future.

Chapter 3 studies the impact of operational decisions of a healthcare organization. In conjunction with the incentives set forth by the regulator, whether directly through incentive programs like the VBP program or indirectly through payment schemes, a natural followup question is the impact of operational decisions on the ability of healthcare organizations to respond to the designed incentives. For instance, the relationship between physicians and hospitals have been changing towards a more integrated model where hospitals tend to directly hire physicians rather than giving them admitting privileges. Analyzing the impact of such operational and organizational decisions is an important aspect in understanding how the healthcare system operates and how to best design incentives that can promote efficiency and quality of care. This is especially important and timely as innovative delivery models, such as telemedicine and personalized care, emerge. It is important to develop a better understanding of the impact of these new delivery models in conjunction with other operational decisions as well as the incentives faced by healthcare organization.

Chapter 4 illustrates the importance of patient-level decisions and the impact they have on healthcare systems. Understanding the incentives that the patients have and how their behaviors change accordingly have significant implications on organization-level an regulator-level decisions. For instance, various national incentive programs are grounded in the idea that policy interventions are necessary due to the lack of quality competition. While a positive relationship between demand and quality seems intuitive and reasonable, it is unclear what the effect would be when throughput is considered. If certain healthcare organizations are already at full capacity, improving quality could in fact lead to lower throughput, harming patient access. Therefore, identifying the market-level characteristics that determine the degree of consumer response to quality as well as the mechanism in which quality improvements lead to changes in patient demand and throughput has an important implication on improving the performance of healthcare systems.

Moving forward, I hope to continue to examine the interconnected nature of different entities within the healthcare system and study how operational challenges experienced by healthcare organizations relate to changes in the regulator's decisions as well as patient behavior. My vision is to produce research that helps managers at healthcare organizations, insurance firms, and policymakers to improve healthcare systems to be as productive as possible in delivering high quality and efficient care.

APPENDIX A

Appendix for "Searching for the Best Yardstick: Cost of Quality Improvements in the U.S. Hospital Industry"

A.1. Additional figures and tables

Table A.1: List of hospital quality measures and domains

Measure ID	Measure Description	Domain
AMI-7a	Fibrinolytic therapy received within 30 minutes of hospital arrival	Clinical Care - Process
AMI-8a	Primary PCI received within 90 minutes of hospital arrival	Clinical Care - Process
HF-1	Discharge instructions	Clinical Care - Process
IMM-2	Influenza immunization	Clinical Care - Process
PN-3b	Blood cultures performed in the ED prior to initial antibiotic received in hospital	Clinical Care - Process
PN-6	Initial antibiotic selection for CAP in immunocompetent patient	Clinical Care - Process
SCIP-Card-2	Surgery patients on a beta blocker prior to arrival that received a beta blocker during the perioperative period	Clinical Care - Process
SCIP-Inf-1	Prophylactic antibiotic received within one hour prior to surgical incision	Clinical Care - Process
SCIP-Inf-2	Prophylactic antibiotic selection for surgical patients	Clinical Care - Process
SCIP-Inf-3	Prophylactic antibiotics discontinued within 24 hours after surgery end time	Clinical Care - Process
SCIP-Inf-4	Cardiac surgery patients with controlled 6AM postoperative serum glucose	Clinical Care - Process
SCIP-Inf-9	Postoperative urinary catheter removal on post operative day 1 or 2	Clinical Care - Process
SCIP-VTE-1	Surgery patients with recommended venous thromboembolism prophylaxis ordered	Clinical Care - Process
SCIP-VTE-2	Surgery patients who received appropriate venous thromboembolism prophylaxis within 24 hours prior to surgery to 24 hours after surgery	Clinical Care - Process
PC-01	Elective delivery prior to 39 completed weeks gestation	Clinical Care - Process (Safety from FY 2018)
MORT-30-AMI	Acute myocardial infarction (AMI) 30-day mortality rate	Clinical Care - Outcomes
MORT-30-HF	Heart failure (HF) 30-day mortality rate	Clinical Care - Outcomes
MORT-30-PN	Pneumonia (PN) 30-day mortality rate	Clinical Care - Outcomes
COMP-HIP-KNEE	Total hip arthroplasty (THA)/total knee arthroplasty (TKA) complication rate	Clinical Care - Outcomes
PSI-90	Complication/patient safety for selected indicators	Clinical Care - Outcomes (Safety from FY 2017)
HAI-1	Central line-associated blood stream infection	Clinical Care - Outcomes (Safety from FY 2017)
HAI-2	Catheter-associated urinary tract infection	Clinical Care - Outcomes (Safety from FY 2017)
HAI-3	Surgical site infection (SSI) - colon surgery	Clinical Care - Outcomes (Safety from FY 2017)
HAI-4	Surgical site infection (SSI) - abdominal hysterectomy	Clinical Care - Outcomes (Safety from FY 2017)
HAI-6	Clostridium difficile infection	Safety
HAI-5	Methicillin-resistant staphylococcus aureus	Safety
HCAHPS-CN	Communication with nurses	Patient Experience
HCAHPS-CD	Communication with doctors	Patient Experience
HCAHPS-RS	Responsiveness of hospital staff	Patient Experience
HCAHPS-PM	Pain management	Patient Experience
HCAHPS-CM	Communication about medicines	Patient Experience
HCAHPS-CQ	Cleanliness and quietness of hospital environment	Patient Experience
HCAHPS-DI	Discharge information	Patient Experience
HCAHPS-OVR	Overall rating of hospital	Patient Experience
HCAHPS-CT	Care transition	Patient Experience
MSPB-1	Medicare spending per beneficiary	Efficiency

Table A.2: Measures included in the Hospital VBP Program from FY 2013 to FY 2020

Measure ID	**FY 2013**	**FY 2014**	**FY 2015**	**FY 2016**	**FY 2017**	**FY 2018**	**FY 2019**	**FY 2020**
AMI-7a	✓	✓	✓	✓	✓			
AMI-8a	✓	✓	✓					
HF-1	✓	✓	✓					
IMM-2				✓	✓			
PN-3b	✓	✓	✓					
PN-6	✓	✓	✓	✓				
SCIP-Card-2	✓	✓	✓	✓				
SCIP-Inf-1	✓	✓	✓					
SCIP-Inf-2	✓	✓	✓	✓				
SCIP-Inf-3	✓	✓	✓	✓				
SCIP-Inf-4	✓	✓	✓					
SCIP-Inf-9		✓	✓	✓				
SCIP-VTE-1	✓	✓						
SCIP-VTE-2	✓	✓	✓	✓				
PC-01					✓	✓	✓	✓
MORT-30-AMI		✓	✓	✓	✓	✓	✓	✓
MORT-30-HF		✓	✓	✓	✓	✓	✓	✓
MORT-30-PN		✓	✓	✓	✓	✓	✓	✓
COMP-HIP-KNEE							✓	✓
PSI-90			✓	✓	✓	✓		
HAI-1			✓	✓	✓	✓	✓	✓
HAI-2				✓	✓	✓	✓	✓
HAI-3				✓	✓	✓	✓	✓
HAI-4				✓	✓	✓	✓	✓
HAI-6					✓	✓	✓	✓
HAI-5					✓	✓	✓	✓
HCAHPS-CN	✓	✓	✓	✓	✓	✓	✓	✓
HCAHPS-CD	✓	✓	✓	✓	✓	✓	✓	✓
HCAHPS-RS	✓	✓	✓	✓	✓	✓	✓	✓
HCAHPS-PM	✓	✓	✓	✓	✓			
HCAHPS-CM	✓	✓	✓	✓	✓	✓	✓	✓
HCAHPS-CQ	✓	✓	✓	✓	✓	✓	✓	✓
HCAHPS-DI	✓	✓	✓	✓	✓	✓	✓	✓
HCAHPS-OVR	✓	✓	✓	✓	✓	✓	✓	✓
HCAHPS-CT						✓	✓	✓
MSPB-1			✓	✓	✓	✓	✓	✓

Table A.3: List of domains and their weights used in the Hospital VBP Program from FY 2013 to FY 2020

Domains	**FY 2013**	**FY 2014**	**FY 2015**	**FY 2016**	**FY 2017**	**FY 2018**	**FY 2019**	**FY 2020**
Clinical Care - Process	70%	45%	20%	10%	5%			
Clinical Care - Outcomes		25%	30%	40%	25%	25%	25%	25%
Safety					20%	25%	25%	25%
Patient Experience	30%	30%	30%	25%	25%	25%	25%	25%
Efficiency			20%	25%	25%	25%	25%	25%

Table A.4: Medicare payment withholding percentages used in counterfactual analyses

Target Percentage	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8
1.00%	1.00%	1.00%	1.00%	1.00%	1.00%	1.00%	1.00%	1.00%
1.25%	1.00%	1.06%	1.13%	1.19%	1.25%	1.25%	1.25%	1.25%
1.50%	1.00%	1.13%	1.25%	1.38%	1.50%	1.50%	1.50%	1.50%
1.75%	1.00%	1.19%	1.38%	1.56%	1.75%	1.75%	1.75%	1.75%
2.00%	1.00%	1.25%	1.50%	1.75%	2.00%	2.00%	2.00%	2.00%
2.25%	1.00%	1.31%	1.63%	1.94%	2.25%	2.25%	2.25%	2.25%
2.50%	1.00%	1.38%	1.75%	2.13%	2.50%	2.50%	2.50%	2.50%
2.75%	1.00%	1.44%	1.88%	2.31%	2.75%	2.75%	2.75%	2.75%
3.00%	1.00%	1.50%	2.00%	2.50%	3.00%	3.00%	3.00%	3.00%
3.25%	1.00%	1.56%	2.13%	2.69%	3.25%	3.25%	3.25%	3.25%
3.50%	1.00%	1.63%	2.25%	2.88%	3.50%	3.50%	3.50%	3.50%
3.75%	1.00%	1.69%	2.38%	3.06%	3.75%	3.75%	3.75%	3.75%
4.00%	1.00%	1.75%	2.50%	3.25%	4.00%	4.00%	4.00%	4.00%

Note. The first column represents target percentage of Medicare payments that withheld. Withholding percentages all begin at 1% and are gradually increased over a five-year period until the target percentage is reached in Year 5.

A.2. Throughput elasticity with respect to performances in quality measures

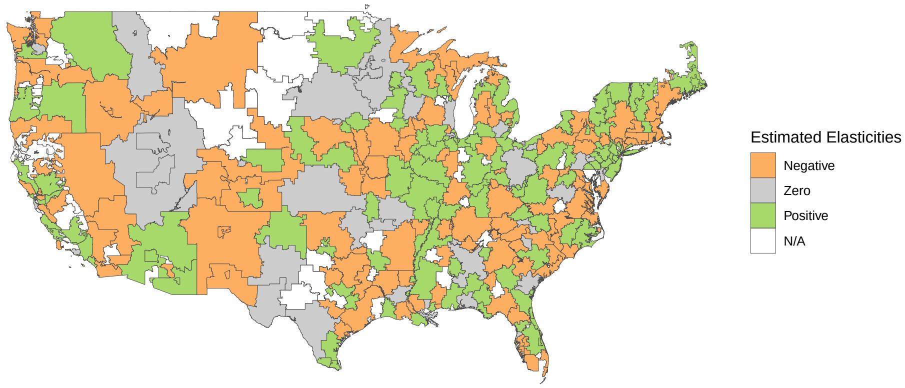
To empirically test the assumption that the trajectory of hospital throughput is largely independent of the performances in quality measures used in the VBP Program, we perform a regression analysis that examine the relationship between composites scores and hospital throughput. Specifically, we perform a fixed-effects Poisson regression using the following specification:

$$\log(E(\lambda_{ht}^M | S_{ht}, \bar{S}_{ht}, t)) = \alpha + \beta_{HRR} \log(S_{ht}) + \gamma \log(\bar{S}_{ht}) + \delta_t + \eta_h, \tag{A.1}$$

where $\bar{S}_{ht}$ is equal to the average composite scores of peer hospitals in the same Hospital Referral Region (HRR) excluding own score. The set of β_{HRR} is the estimated HRR-specific throughput elasticity with respect to quality.

We present the results in Figure A.1. First, we qualitatively verify that there isn't any discernible pattern in HRRs with positive versus negative elasticities. We further examine any correlation with market-level characteristics, including level of income, education, and

Figure A.1: Medicare patient throughput elasticity with respect to quality by hospital referral regions



hospital density, and find that none of the market-level characteristics that we examine are significantly correlated with throughput elasticity.

A.3. Structural estimation

The specific set of investment decisions prescribed by the optimal investment policy is a function of the parameters governing the distributions of performance scores conditional on quality levels $(\mu_\omega^j, \sigma_\omega^j)$, the initial distribution of quality levels F_0, the transition matrices conditional on investment decision F_x, and the parameters governing the common distributions of the operating and investment costs $(\mu_{\boldsymbol{c}}, \sigma_{\boldsymbol{c}}, \rho_c)$. The goal of the estimation process is to recover the true values of the parameters under the assumption that the data were generated by Markov perfect equilibrium behavior. As a byproduct of the estimation process, we also recover the most likely quality level of each hospital and the most likely investment decision it made at each period we have data.

The key challenge in the estimation process is that we do not observe which level within the quality ladder hospitals belonged, whether or not hospitals decided to invest in quality, and the breakdown of total costs into the baseline operating costs that hospitals incurred and the cost of investment that hospitals incurred or that may had prevented hospitals from investing in quality. Therefore, we treat positions within the quality ladder, investment decisions, operating costs, and investment costs as latent variables and utilize expectation-maximization (EM) algorithm to find the maximum likelihood estimates of the parameters.

The estimation process can be summarized as an iterative process that cycles between two main steps. In the first step, or the expectation step, we obtain posterior samples of the latent variables given candidate estimates of the parameters. We utilize Metropolis-Hastings (MH) algorithm, which can be used to draw random samples when the probability distribution itself is not known but a function proportional to the density function is known and can be computed. To draw posterior samples of the latent variables, we first draw candidate samples of cost values for each hospital. Then, we compute optimal investment policy using dynamic programming given the candidate cost values, then find the most likely trajectory of quality levels and associated investment decisions using a hidden Markov chain framework. Next, we evaluate the log-likelihood of each candidate sample of the latent variables, after which the log-likelihood is used to either accept or reject into the set of posterior samples. The second step consists of updating the estimates of the parameters. We take a sufficient statistics approach and identify the set of parameters that maximizes the expected log-likelihood given the posterior samples. In the next iteration, the updated parameters are used to again draw posterior samples, which are then used to further update the estimates of the parameters. The iterative process continues until the estimates of the parameters have converged to stationary points. In the rest of this section, we describe in detail each component of the estimation process.

A.3.1. Expectation-maximization algorithm

The three components that are used in the EM algorithm are observed data, latent data, and parameters. The observed data used in the estimation are the vector of set of performances in quality measures of hospital h over the study period, denoted $\hat{\boldsymbol{q}}_h = (\hat{\boldsymbol{q}}_{h1}, \ldots, \hat{\boldsymbol{q}}_{hT})$ where each $\hat{\boldsymbol{q}}_{ht}$ is a vector consisting of performances in all individual measures in period t, i.e., $\hat{\boldsymbol{q}}_{ht} = (\hat{q}^j_{ht}\ \forall j \in \mathcal{S}_t)$, and the observed total costs, denoted $\boldsymbol{\pi}^c_h = (\pi^c_{h1}, \ldots, \pi^c_{hT})$. T is the last period in which we have data. The benefit of using observed total costs, as opposed to total profits, is that we can separate out the incentive payments when estimating the cost parameters. In essence, performances in quality measures provide information on quality ladder, including transitions and emission distributions within each level, and the observed total costs provide information on cost parameters and how well the cost parameters fit the observed total costs. We denote the set of observed data for hospital h by $\boldsymbol{X}_h = (\hat{\boldsymbol{q}}_h, \boldsymbol{\pi}^c_h)$. The latent set of variables includes operating and investment costs per patient $\boldsymbol{c}_h$, sequence of quality levels, denoted $\boldsymbol{\omega}_h = (\omega_{h1}, \ldots, \omega_{hT})$, and sequence of investment decisions, denoted $\boldsymbol{x}_h = (x_{h1}, \ldots, x_{hT})$. We denote the latent data for hospital h by $\boldsymbol{Z}_h = (\boldsymbol{c}_h, \boldsymbol{\omega}_h, \boldsymbol{x}_h)$. For brevity, we define $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_H)$ and $\boldsymbol{Z} = (\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_H)$. Finally, the parameters to be estimated are $\boldsymbol{\theta} = (\mu_{\boldsymbol{c}}, \sigma_{\boldsymbol{c}}, \rho_c, F_0, F_x, G_\omega, \sigma_u)$. The EM algorithm iteratively finds the values of the parameters $\boldsymbol{\theta}$ that maximize the likelihood function

$$\begin{aligned} L(\boldsymbol{\theta}; \boldsymbol{X}) &= P(\boldsymbol{X}|\boldsymbol{\theta}) \\ &= \int P(\boldsymbol{X}, \boldsymbol{Z}|\boldsymbol{\theta}) d\boldsymbol{Z}. \end{aligned} \tag{A.2}$$

The E-step can be formally written as

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = E_{\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{\theta}^{(t)}} \log L(\boldsymbol{\theta}; \boldsymbol{X}, \boldsymbol{Z}). \tag{A.3}$$

In practice, computing the conditional expectation requires generating posterior samples that follow the conditional distribution of $\boldsymbol{Z}$ given $\boldsymbol{X}$ and the current estimates of the parameters $\boldsymbol{\theta}^{(t)}$. We generate the posterior samples by using MH algorithm, which can draw samples from any probability distribution as long as there is a function that is proportional to the density of the distribution, by exploiting the fact that $\log L(\boldsymbol{\theta}; \boldsymbol{X}, \boldsymbol{Z})$ is a function that is proportional to the conditional distribution of $\boldsymbol{Z}$.

The MH algorithm involves choosing a first sample and drawing random perturbations to draw candidate samples. The algorithm either accepts or rejects the candidate sample by evaluating the function that is proportional to the density of the distribution to be approximated. To facilitate the sample generation process, we use the current estimates of

$\mu_{\boldsymbol{c}}$ as the first sample of $\log \boldsymbol{c}_h$. Rather than drawing random samples of the path of quality ladder and investment decisions, we first recursively find the optimal policy by solving for the following Bellman equation given the sample draw of $\boldsymbol{c}_h$ and the current estimates of the parameters $\boldsymbol{\theta}^{(t)}$

$$V_{ht}(\omega_{ht}) = \max_{x_{ht}} E[\pi(\omega_{ht}, x_{ht})] + \beta \sum_{\Omega} P(\omega_{h,t+1}|\omega_{ht}, x_{ht}) V_{h,t+1}(\omega_{h,t+1}). \tag{A.4}$$

Given the optimal policy, we model transition of quality levels within the ladder as a hidden Markov process with choice-dependent transition matrices and find the most likely path of quality level, denoted $\boldsymbol{\omega}_h^*$, and the corresponding investment decisions, denoted $\boldsymbol{x}_h^*$. In turn, we evaluate $\log L_h(\boldsymbol{\theta}; \boldsymbol{X}_h, \boldsymbol{Z}_h)$ given the sample draw of $\boldsymbol{c}_h$, $\boldsymbol{\omega}_h^*$, and $\boldsymbol{x}_h^*$. We describe the hidden Markov model and the computation of $\log L_h(\boldsymbol{\theta}; \boldsymbol{X}_h, \boldsymbol{Z}_h)$ in detail in Section A.3.2. For the next iteration of the MH algorithm, we draw random perturbations of cost values by first using a zero-mean bivariate normal distribution with $(\sigma_{c^0}, \sigma_{c^1}, \rho_c)$ as the standard deviations and the correlation factor for candidate $\log c_h^0$ and $\log c_h^1$. Similarly, we draw random perturbation to draw candidate $\log c_h^2$ using a zero-mean normal distribution with standard deviation of σ_{c^2}. Then, using the candidate cost values, we again find the optimal policy, find the most likely path of quality levels and investment decisions, and evaluate $\log L_h(\boldsymbol{\theta}; \boldsymbol{X}_h, \boldsymbol{Z}_h)$ under the new candidate sample. We accept or reject the candidate sample by comparing the values of $\log L_h(\boldsymbol{\theta}; \boldsymbol{X}_h, \boldsymbol{Z}_h)$ under the previous sample and the new candidate sample. We allow a burn-in period of 4,000, i.e., we throw away the first 4,000 samples, then we take every 50th sample until we have 20 samples for each hospital. We parallelize this process at the hospital level, and end up with $20NT$ posterior samples at each iteration of the E-step.

The M-step involves finding the set of values of the parameters that maximize $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ and setting the argmax as $\boldsymbol{\theta}^{(t+1)}$ to be used in the next iteration of the algorithm. Rather than performing a numerical maximization to find the argmax, we exploit the posterior samples by using sufficient statistics approach. First, the parameters of the common distributions of the operating and investment costs $(\mu_{\boldsymbol{c}}, \sigma_{\boldsymbol{c}}, \rho_c)$ are estimated by taking the sample mean and the sample variance of the posterior samples. The initial distribution F_0 and the transition matrices F_x are estimated directly from the observed state transitions in the posterior samples. The emission distributions of performances in quality measures G_ω are estimated by maximum likelihood estimates on the performances of hospitals in the posterior samples that falls in each quality level. Finally, we construct a structural error term, u_{ht} by taking the difference between the observed profits and the estimated profits. We assume that it follows the distribution $u_{ht} \sim N(0, \sigma_u^2)$, and estimate the standard deviation of the structural error term σ_u by the sample standard deviation of u_{ht} in the posterior samples.

A.3.2. Hidden Markov model

The estimation process requires the computation of $\log L_h(\boldsymbol{\theta}; \boldsymbol{X}_h, \boldsymbol{Z}_h)$ given candidate $\boldsymbol{c}_h$. To do so, we model the path of quality levels as a hidden Markov process and find the most likely path of quality levels and investment decisions and, in turn, the log-likelihood given $\boldsymbol{c}_h$, $\boldsymbol{\omega}_h^*$, and $\boldsymbol{x}_h^*$. We can write and expand the likelihood

$$\begin{aligned} L_h(\boldsymbol{\theta}; \boldsymbol{X}_h, \boldsymbol{Z}_h) &= P(\boldsymbol{X}_h | \boldsymbol{Z}_h, \boldsymbol{\theta}) P(\boldsymbol{Z}_h | \boldsymbol{\theta}) \\ &= P(\hat{\boldsymbol{q}}_h | \boldsymbol{\omega}_h^*, \boldsymbol{\theta}) P(\boldsymbol{\pi}_h^c | \boldsymbol{c}_h, \boldsymbol{\omega}_h^*, \boldsymbol{x}_h^*, \boldsymbol{\theta}) P(\boldsymbol{\omega}_h^*, \boldsymbol{x}_h^* | \boldsymbol{c}_h, \boldsymbol{\theta}) P(\boldsymbol{c}_h | \boldsymbol{\theta}). \end{aligned} \tag{A.5}$$

To find the most likely path of quality ladder and investment decisions, we first find the optimal policy given $\boldsymbol{c}_h$ by solving the Bellman equation

$$V_{ht}(\omega_{ht}) = \max_{x_{ht}} E[\pi(\omega_{ht}, x_{ht})] + \beta \sum_{\Omega} P(\omega_{h,t+1} | \omega_{ht}, x_{ht}) V_{h,t+1}(\omega_{h,t+1}). \tag{A.6}$$

Given the optimal investment policy, finding the most likely path of quality levels $\boldsymbol{\omega}_h^*$ is sufficient to find the most likely path of investment decisions $\boldsymbol{x}_h^*$ since they are simply the investment decisions prescribed by the optimal investment policy at each point in the most likely path of quality levels. To find $\boldsymbol{\omega}_h^*$ conditional on a draw of $\boldsymbol{c}_h$ and current estimates of $\boldsymbol{\theta}$, we define for each possible quality level, denoted $k = 1, \ldots, 5$,

$$\alpha_1(k) = P(\hat{\boldsymbol{q}}_{h1}|\omega_{h1} = k)P(\pi^c_{h1}|\omega_{h1} = k)P(\omega_{h1} = k). \tag{A.7}$$

For $t = 2, \ldots, T$, we compute

$$\alpha_t(k) = \max_{m\in\{1,\ldots,5\}} P(\hat{\boldsymbol{q}}_{ht}|\omega_{ht} = k)P(\pi^c_{ht}|\omega_{ht} = k)P(\omega_{ht} = k|\omega_{h,t-1} = m)\alpha_{t-1}(m). \tag{A.8}$$

In (A.7) and (A.8), $P(\hat{\boldsymbol{q}}_{ht}|\omega_{ht} = k)$ is computed by following the rectified normal distribution with upper bounds d^j

$$P(\hat{\boldsymbol{q}}_{ht}|\omega_{ht} = k) = \prod_{j\in\mathcal{S}_t} \left(1 - \Phi\left(\frac{d^j - \mu^j_k}{\sigma^j_k}\right)\right)^{\mathbf{1}(\hat{q}^j_{ht}=d^j)} \left(\frac{1}{\sigma^j_k}\phi\left(\frac{\hat{q}^j_{ht} - \mu^j_k}{\sigma^j_k}\right)\right)^{\mathbf{1}(\hat{q}^j_{ht}<d^j)}. \tag{A.9}$$

$P(\pi^c_{ht}|\omega_{ht} = k)$ is computed by evaluating the difference between observed total costs and estimated costs derived from the candidate cost values, current quality level, and the corresponding investment decision specified by the optimal policy. $P(\omega_{h1} = k)$ is taken directly from the initial distribution within the quality ladder F_0, and lastly the transition matrices F_x are used to identify $P(\omega_{ht} = k|\omega_{h,t-1} = m)$.

We then find the most likely path of quality levels by finding the last quality level that maximizes $\alpha_T(k)$ and proceeding backwards by keeping track of each preceding state. Note that the maximum value conditional on $\boldsymbol{c}_h$ and $\boldsymbol{\theta}$

$$\alpha_T(k^*) = \max_{k\in 1,\ldots,5} \alpha_T(k) \tag{A.10}$$

$$= P(\hat{\boldsymbol{q}}_h|\boldsymbol{\omega}^*_h, \boldsymbol{\theta})P(\boldsymbol{\pi}^c_h|\boldsymbol{c}_h, \boldsymbol{\omega}^*_h, \boldsymbol{x}^*_h, \boldsymbol{\theta})P(\boldsymbol{\omega}^*_h, \boldsymbol{x}^*_h|\boldsymbol{c}_h, \boldsymbol{\theta}) \tag{A.11}$$

is equal to the first three terms in (A.5). Therefore, we can now compute

$$L_h(\boldsymbol{\theta}; \boldsymbol{X}_h, \boldsymbol{Z}_h) = \alpha_T(k^*)P(\boldsymbol{c}_h|\boldsymbol{\theta}), \tag{A.12}$$

where $P(\boldsymbol{c}_h|\boldsymbol{\theta})$ is computed using the density functions of the bivariate normal distribution with parameters $(\mu_{c^0}, \sigma^2_{c^0}, \mu_{c^1}, \sigma^2_{c^1}, \rho_c)$ and the normal distribution with parameters $(\mu_{c^2}, \sigma^2_{c^2})$. In practice, we compute the log-likelihood to avoid underflow. Using the log-likelihood function, we iterate over the EM algorithm until the estimated parameters converge. We discuss how we determine convergence and provide results in Appendix A.4.

A.4. Convergence criteria for estimates and counterfactual equilibria

To determine the convergence of the iterative estimation process, we first define $\bar{\theta}_i^{(k)}$ as the average of the last ten estimated values, i.e.,

$$\bar{\theta}_i^{(k)} = \frac{\sum_{j=i-9}^{i} \theta_j^{(k)}}{10}. \tag{A.13}$$

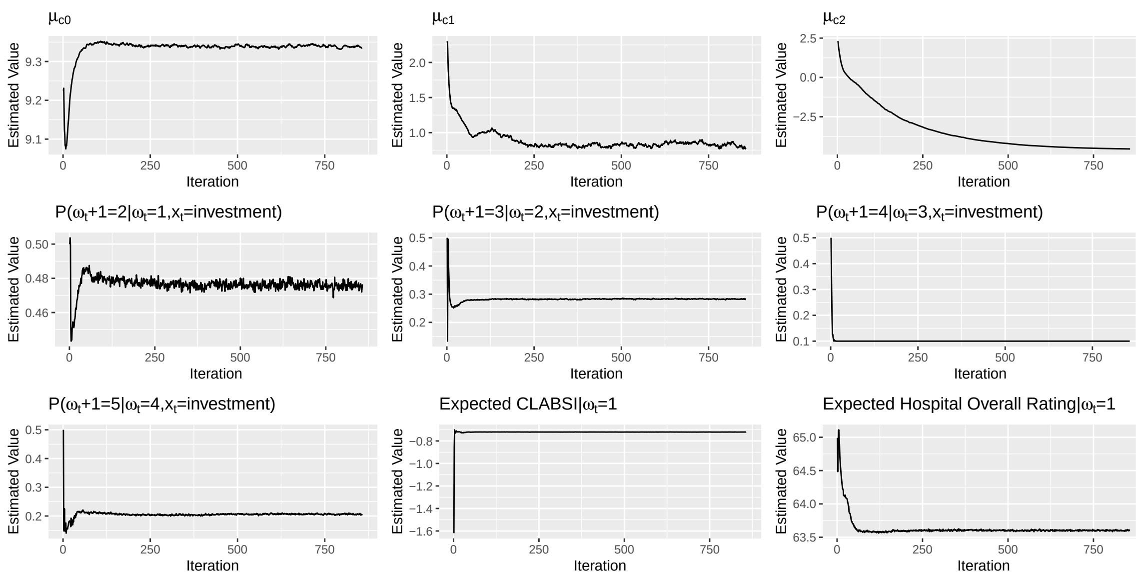
Then, we use the following two criteria to define convergence of the i^{th} estimate of parameter $\theta_i^{(k)}$:

$$\theta_i^{(k)} \text{ has converged if } \begin{cases} |(\bar{\theta}_i^{(k)} - \theta_i^{(k)})/\bar{\theta}_i^{(k)}| < \epsilon^{(k)} \quad \text{or} \\ |\bar{\theta}_i^{(k)} - \theta_i^{(k)}| < \delta^{(k)}. \end{cases} \tag{A.14}$$

In practice, we use $\epsilon^{(k)} = 0.025$ and $\delta^{(k)} = 0.001$. We also qualitatively verify that convergence has been reached using Figure A.2.

Similarly, we use the same criteria to determine the convergence in computing counterfactual equilibrium presented in Section 2.6. To estimate the 95% confidence interval of the key metrics we present, we independently repeat the computation of counterfactual equilibrium 100 times and choose the 2.5 and 97.5 percentile values as the lower bound and the upper bound of the confidence interval.

Figure A.2: Estimated values of selected parameters

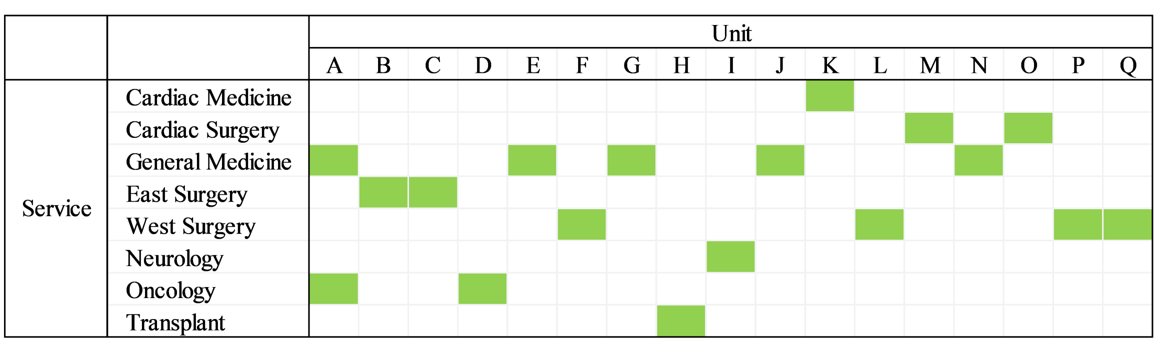


APPENDIX B

Appendix for "The Spillover Effects of Capacity Pooling in Hospitals"

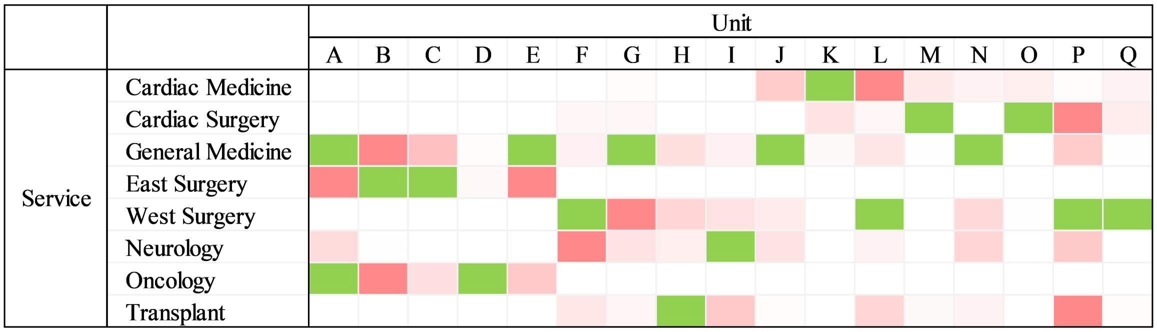
B.1. Additional figures and tables

Figure B.1: Service-to-unit mapping



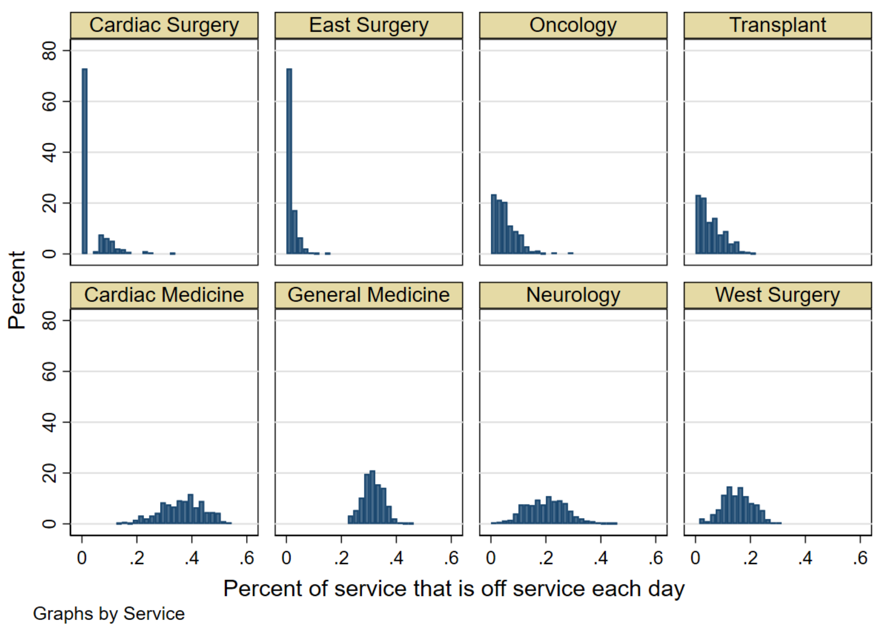
Note. This figure shows the *on*-service designation for each unit.

Figure B.2: Routing frequency to off-service units by service



Note. This figure shows which off-service units were used most frequently by each service. Darker shades of red indicate a higher frequency of off-service placement to that unit by a given service.

Figure B.3: Daily proportion off service for protected versus unprotected services



Note. This figure plots daily proportion of patients who are placed off service by each service. The first row shows the four services that are protected in the "on until a few beds + protected services" policy described in section 3.7.2.

Table B.1: Computed probabilities for the "*on until a few beds*" policy

# Open beds on service	Cardiac Medicine	Cardiac Surgery	East Surgery	General Medicine	Neurology	Oncology Medicine	Transplant	West Surgery
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	0.52	0.03	0.04	0.37	0.41	0.14	0.14	0.37
2	0.43	0.01	0.02	0.37	0.27	0.14	0.07	0.31
3	0.37	0.01	0.02	0.37	0.20	0.09	0.04	0.29
4	0.31	0.01	0.01	0.35	0.18	0.08	0.04	0.22
5	0.23	0.01	0.01	0.35	0.13	0.08	0.01	0.22

Note. This table shows the probabilities that are used to route patients in the "on until a few beds" policy described in section 3.7.2. Each value is computed by calculating the proportion of patients who were placed off service given the number of open beds on service at the time of admission.

B.2. Full results tables with control variables

Table B.2: Spillover effect of off-service placement, operationalized using the service's proportion of patients placed off service

	(1)	(2)	(3)	(4)	(5)
	Mean of proportion off service	Logged length of stay	Hospital readmission	Trigger activation	In-hospital mortality
Preadmission service-to-hospital utilization ratio	0.0699***				
	(0.0105)				
Preadmission hospital utilization excluding focal service	-0.125***				
	(0.0157)				
Mean of proportion off service		2.484***	-0.348	-1.263	-0.870
		(0.567)	(1.365)	(1.736)	(5.282)
SD of proportion off service	1.358***	7.753***	2.694	10.47***	-0.812
	(0.0559)	(0.867)	(2.066)	(2.629)	(7.731)
Age (years)	0.000178***	0.000514	-0.00204**	0.00675***	0.0192***
	(0.0000346)	(0.000275)	(0.000747)	(0.00113)	(0.00278)
Female indicator	0.00172	-0.00109	-0.0932***	0.102**	-0.0644
	(0.00122)	(0.00865)	(0.0235)	(0.0345)	(0.0747)
DRG cost weight	-0.000987*	0.0712***	-0.00560	0.0573***	0.0654*
	(0.000495)	(0.00432)	(0.0111)	(0.00917)	(0.0262)
Complications or comorbidities	-0.00460**	0.172***	0.106***	0.355***	0.666***
	(0.00156)	(0.0121)	(0.0252)	(0.0355)	(0.103)
Number of transfers	-0.00394***	0.0624***	-0.0563**	0.104***	0.141***
	(0.000810)	(0.00697)	(0.0200)	(0.0185)	(0.0418)
Unit-level utilization	-1.988***	11.18***	6.953	6.165	14.98
	(0.123)	(1.366)	(3.693)	(7.461)	(15.64)
(Unit-level utilization)^2	1.426***	-7.153***	-4.749	-3.101	-9.773
	(0.0734)	(0.926)	(2.453)	(4.478)	(10.23)
Service-level utilization	-1.960***	20.52***	4.135	12.89	24.12
	(0.234)	(2.011)	(4.775)	(8.870)	(28.59)
(Service-level utilization)^2	1.137***	-12.62***	-2.001	-8.337	-13.31
	(0.136)	(1.167)	(2.804)	(5.137)	(16.45)
Visited ICU during encounter	0.0378***	-0.393***	-0.123	-0.235**	0.609**
	(0.00197)	(0.0268)	(0.0800)	(0.0787)	(0.228)
Admitted during PM shift	0.00328	0.0799***	-0.112*	0.0381	0.564**
	(0.00235)	(0.0161)	(0.0460)	(0.0492)	(0.205)
Admitted during overnight shift	0.00472*	-0.0000911	-0.227***	0.0476	0.393
	(0.00211)	(0.0149)	(0.0423)	(0.0596)	(0.213)
Admitted on weekday	-0.0141***	0.0387*	0.0221	0.00353	0.0269
	(0.00182)	(0.0166)	(0.0344)	(0.0645)	(0.123)
Surgical service	-0.138***	0.200*	-0.539**	-0.332	-1.044
	(0.00167)	(0.0807)	(0.187)	(0.248)	(0.821)
Mixed service	-0.131***	0.182*	-0.137	-0.583*	-0.584
	(0.00204)	(0.0775)	(0.183)	(0.248)	(0.741)
Controls	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes
Model	2SLS	2SLS	2SRI	2SRI	2SRI
	1st stage	2nd stage	2nd stage	2nd stage	2nd stage
Observations	14793	14793	14793	14793	14793

Note. Standard errors (in parentheses) are heteroskedasticity robust for continuous outcome variables and bootstrapped for binary outcome variables. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.3: Spillover effect of off-service placement, operationalized using the number of units across which the service has patients placed off service

	(1) Mean of units with off-service patients	(2) Logged length of stay	(3) Hospital readmission	(4) Trigger activation	(5) In-hospital mortality
Preadmission service-to-hospital utilization ratio	4.558***				
	(0.323)				
Preadmission hospital utilization excluding focal service	-4.051***				
	(0.537)				
Mean of units with off-service patients		0.0441***	0.00658	-0.0188	-0.0134
		(0.0108)	(0.0305)	(0.0381)	(0.103)
SD of units with off-service patients	0.923***	0.319***	0.107**	0.276***	0.0557
	(0.0535)	(0.0178)	(0.0380)	(0.0434)	(0.115)
Age (years)	-0.000231	0.00125***	-0.00212**	0.00674***	0.0173***
	(0.00121)	(0.000240)	(0.000653)	(0.00103)	(0.00267)
Female indicator	0.128**	-0.00818	-0.0944***	0.0959**	-0.0789
	(0.0409)	(0.00809)	(0.0272)	(0.0295)	(0.0775)
DRG cost weight	-0.139***	0.0803***	-0.00513	0.0606***	0.0640*
	(0.0175)	(0.00444)	(0.00936)	(0.0129)	(0.0251)
Complications or comorbidities	0.165**	0.144***	0.101***	0.358***	0.637***
	(0.0532)	(0.0113)	(0.0271)	(0.0391)	(0.0976)
Number of transfers	-0.0771**	0.0570***	-0.0546**	0.110***	0.133**
	(0.0275)	(0.00621)	(0.0179)	(0.0210)	(0.0433)
Unit-level utilization	-78.86***	10.30***	8.233*	7.306	17.94
	(5.412)	(1.177)	(3.304)	(9.856)	(14.03)
(Unit-level utilization)^2	55.90***	-6.444***	-5.697**	-3.945	-11.76
	(3.195)	(0.770)	(2.175)	(5.878)	(8.925)
Service-level utilization	17.48*	13.49***	4.799	15.32	32.83
	(7.022)	(1.695)	(4.753)	(8.125)	(24.18)
(Service-level utilization)^2	-14.19***	-8.378***	-2.369	-9.778*	-18.68
	(4.112)	(0.987)	(2.769)	(4.710)	(13.64)
Visited ICU during encounter	1.324***	-0.354***	-0.161**	-0.253***	0.513**
	(0.0678)	(0.0207)	(0.0532)	(0.0715)	(0.158)
Admitted during PM shift	0.0779	0.0804***	-0.118**	0.0346	0.578**
	(0.0783)	(0.0147)	(0.0458)	(0.0668)	(0.219)
Admitted during overnight shift	0.189**	-0.00186	-0.234***	0.0383	0.415*
	(0.0711)	(0.0135)	(0.0416)	(0.0586)	(0.210)
Admitted on weekday	-0.251***	0.00370	0.0330	0.0145	0.0158
	(0.0609)	(0.0126)	(0.0374)	(0.0418)	(0.124)
Surgical service	-3.735***	-0.0449	-0.478***	-0.281	-0.971
	(0.0543)	(0.0432)	(0.119)	(0.147)	(0.497)
Mixed service	-4.330***	0.0748	-0.0566	-0.465**	-0.583
	(0.0560)	(0.0503)	(0.146)	(0.167)	(0.469)
Controls	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes
Model	2SLS	2SLS	2SRI	2SRI	2SRI
	1st stage	2nd stage	2nd stage	2nd stage	2nd stage
Observations	14793	14793	14793	14793	14793

Note. Standard errors (in parentheses) are heteroskedasticity robust for continuous outcome variables and bootstrapped for binary outcome variables. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.4: Spillover effect of off-service placement, operationalized using the number of movements into and out of off-service units

	(1) Logged length of stay	(2) Hospital readmission	(3) Trigger activation	(4) In-hospital mortality
Mean of proportion off service	2.593***	-0.216	-1.100	-1.086
	(0.470)	(1.638)	(1.837)	(4.223)
Demeaned count of movements into and out of off-service units	0.00521***	0.000862***	0.00433***	0.000729
	(0.000132)	(0.000224)	(0.000276)	(0.000522)
Age (years)	0.000734**	-0.00196*	0.00733***	0.0194***
	(0.000236)	(0.000820)	(0.00109)	(0.00316)
Female indicator	-0.00764	-0.0953***	0.0947***	-0.0635
	(0.00757)	(0.0246)	(0.0264)	(0.0715)
DRG cost weight	0.0519***	-0.00732	0.0434***	0.0554*
	(0.00499)	(0.0112)	(0.0130)	(0.0247)
Complications or comorbidities	0.0872***	0.0940**	0.294***	0.670***
	(0.0120)	(0.0336)	(0.0368)	(0.109)
Number of transfers	0.0255***	-0.0615***	0.0708***	0.134**
	(0.00640)	(0.0163)	(0.0192)	(0.0412)
Unit-level utilization	11.60***	7.407	8.291	15.04
	(1.226)	(4.156)	(10.61)	(15.14)
(Unit-level utilization)^2	-7.519***	-5.075	-4.667	-9.844
	(0.816)	(2.856)	(6.335)	(9.744)
Service-level utilization	17.68***	3.704	8.535	19.57
	(1.744)	(5.141)	(10.31)	(27.75)
(Service-level utilization)^2	-10.85***	-1.713	-5.535	-10.73
	(1.013)	(2.928)	(5.884)	(15.84)
Visited ICU during encounter	-0.416***	-0.121	-0.266**	0.598*
	(0.0268)	(0.0915)	(0.0916)	(0.244)
Admitted during PM shift	0.0751***	-0.113*	0.0390	0.566**
	(0.0144)	(0.0491)	(0.0536)	(0.203)
Admitted during overnight shift	0.0141	-0.227***	0.0568	0.402
	(0.0130)	(0.0424)	(0.0544)	(0.213)
Admitted on weekday	0.00776	0.0181	-0.0250	0.0309
	(0.0146)	(0.0559)	(0.0678)	(0.165)
Surgical service	0.167*	-0.538*	-0.320	-1.036
	(0.0720)	(0.246)	(0.288)	(0.737)
Mixed service	0.266***	-0.0974	-0.384	-0.576
	(0.0596)	(0.196)	(0.231)	(0.531)
Controls	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Model	2SLS 2nd stage	2SRI 2nd stage	2SRI 2nd stage	2SRI 2nd stage
Observations	14793	14793	14793	14793

Note. Standard errors (in parentheses) are heteroskedasticity robust for continuous outcome variables and bootstrapped for binary outcome variables. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.5: Seemingly unrelated regression on spillover effects of movements into and out of on-service units versus off-service units

	(1) Logged length of stay	Logged length of stay
Mean of proportion off service	2.777***	2.884***
	(0.410)	(0.375)
Demeaned count of movements into and out of off-service units	0.00267***	
	(0.000120)	
Demeaned count of movements into and out of on-service units		0.00153***
		(0.0000717)
Age (years)	0.000677**	0.000639**
	(0.000209)	(0.000203)
Female indicator	-0.00560	-0.00513
	(0.00687)	(0.00646)
DRG cost weight	0.0667***	0.0508***
	(0.00319)	(0.00348)
Complications or comorbidities	0.135***	0.128***
	(0.0102)	(0.0102)
Number of transfers	0.0466***	0.0376***
	(0.00588)	(0.00520)
Unit-level utilization	11.97***	11.62***
	(1.130)	(1.030)
(Unit-level utilization)^2	-7.729***	-7.533***
	(0.761)	(0.692)
Service-level utilization	19.33***	16.55***
	(1.698)	(1.562)
(Service-level utilization)^2	-11.84***	-10.10***
	(0.976)	(0.905)
Visited ICU during encounter	-0.386***	-0.396***
	(0.0253)	(0.0232)
Admitted during PM shift	0.0808***	0.0648***
	(0.0118)	(0.0109)
Admitted during overnight shift	0.00230	0.00171
	(0.0114)	(0.0102)
Admitted on weekday	0.0207	0.0248*
	(0.0133)	(0.0122)
Surgical service	0.189**	0.219***
	(0.0632)	(0.0583)
Mixed service	0.284***	0.299***
	(0.0517)	(0.0478)
Controls	Yes	Yes
Month FE	Yes	Yes
Model	2SRI 2nd stage	2SRI 2nd stage
Observations	14793	14793

Note. Standard errors (in parentheses) are heteroskedasticity robust for continuous outcome variables and bootstrapped for binary outcome variables. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.6: Spillover effect of off-service placement, operationalized using the maximum positive and negative deviations from the proportion off service at time of admission

	(1) Logged length of stay	(2) Hospital readmission	(3) Trigger activation	(4) In-hospital mortality
Mean of proportion off service	2.128***	-0.484	-1.852	-1.529
	(0.570)	(1.543)	(2.318)	(5.300)
Maximum positive deviation in proportion off service	3.530***	0.615	5.034***	1.996
	(0.354)	(0.931)	(1.408)	(3.089)
Maximum negative deviation in proportion off service	2.502***	0.941	4.001**	-2.020
	(0.343)	(0.799)	(1.276)	(3.016)
Age (years)	0.000413	-0.00204*	0.00663***	0.0199***
	(0.000258)	(0.000816)	(0.00115)	(0.00329)
Female indicator	-0.00214	-0.0932***	0.104***	-0.0770
	(0.00830)	(0.0264)	(0.0293)	(0.0790)
DRG cost weight	0.0625***	-0.00639	0.0426**	0.0575*
	(0.00448)	(0.0113)	(0.0143)	(0.0238)
Complications or comorbidities	0.164***	0.105***	0.346***	0.677***
	(0.0119)	(0.0302)	(0.0474)	(0.0873)
Number of transfers	0.0570***	-0.0582**	0.0921***	0.134**
	(0.00702)	(0.0197)	(0.0251)	(0.0436)
Unit-level utilization	10.64***	6.648	4.641	11.47
	(1.333)	(3.501)	(8.840)	(14.14)
(Unit-level utilization)^2	-6.817***	-4.543	-2.153	-7.622
	(0.903)	(2.352)	(5.433)	(9.327)
Service-level utilization	20.30***	3.806	13.05	28.37
	(1.964)	(4.487)	(10.32)	(23.62)
(Service-level utilization)^2	-12.51***	-1.803	-8.519	-15.93
	(1.139)	(2.605)	(5.930)	(13.57)
Visited ICU during encounter	-0.395***	-0.118	-0.248*	0.657***
	(0.0243)	(0.0707)	(0.0993)	(0.184)
Admitted during PM shift	0.0690***	-0.112**	0.0187	0.532*
	(0.0154)	(0.0365)	(0.0495)	(0.225)
Admitted during overnight shift	-0.000707	-0.225***	0.0455	0.362
	(0.0144)	(0.0372)	(0.0465)	(0.224)
Admitted on weekday	0.0339*	0.0189	-0.00120	0.00726
	(0.0159)	(0.0465)	(0.0619)	(0.145)
Surgical service	0.189*	-0.554**	-0.338	-1.113
	(0.0763)	(0.207)	(0.310)	(0.760)
Mixed service	0.132	-0.147	-0.677*	-0.654
	(0.0786)	(0.213)	(0.324)	(0.745)
Controls	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Model	2SLS 2nd stage	2SRI 2nd stage	2SRI 2nd stage	2SRI 2nd stage
Observations	14793	14793	14793	14793

Note. Standard errors (in parentheses) are heteroskedasticity robust for continuous outcome variables and bootstrapped for binary outcome variables. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

B.3. Heterogeneous treatment effects

Hospitals have multiple specialties that provide specific sets of services that can be broadly categorized as medical or surgical. Since these two types of specialties provide two distinct types of care, we examine spillover effects that are specific to each service type by stratifying the sample into two subsamples: medical specialties versus surgical specialties. To separately identify the spillover effects for each service type, we modify the analysis by including service-level and unit-level fixed effects for each patient. To address service-type differences in the level of off-service placement as well as differences in overall utilization, we vectorize the IV by creating eight variables, with each corresponding to a particular service that equals the service-to-hospital utilization ratio one hour prior to admission if the patient is in the service, and zero otherwise.

We find that while patients in both medical and surgical services experience similar level of spillover effects from the volatility of off-service placement when it comes to their lengths of stay, only patients in the medical services are affected by the spillovers from experiencing a high level of off-service placement. Furthermore, the volatility of off-service placement leads to increases in readmission likelihood for surgical patients, whereas it leads to increases in the likelihood of clinical trigger activation for both medical and surgical patients alike.

We apply service-type-specific estimates for our counterfactual analyses with a few necessary modifications. First, we use maximum likelihood estimators for the binary outcome variables instead of 2SRI estimators, so that we are able to calculate predicted probabilities. Second, we drop unit-level fixed effects from the maximum likelihood estimators to ensure convergence. We confirm that the coefficients derived with the modifications are not meaningfully different from the estimates presented in Table B.7 of Appendix B.3.

Table B.7: Spillover effect of off-service placement in medical and surgical services, operationalized using the service's proportion of patients placed off service

	Medical services			Surgical services		
	(1) Logged length of stay	(2) Hospital readmission	(3) Trigger activation	(4) Logged length of stay	(5) Hospital readmission	(6) Trigger activation
Mean of proportion off service	3.163*	-1.199	-1.035	-0.503	-5.550	-6.838
	(1.459)	(4.294)	(5.048)	(1.319)	(5.341)	(6.647)
SD of proportion off service	12.28***	2.975	10.30***	12.02***	10.37**	13.03**
	(0.770)	(1.567)	(1.855)	(1.236)	(3.572)	(4.061)
Age (years)	0.00118***	-0.00349***	0.00572***	0.00259***	0.00390**	0.0103***
	(0.000305)	(0.000911)	(0.00120)	(0.000344)	(0.00147)	(0.00176)
Female indicator	0.00655	-0.0499	0.0169	0.00274	-0.0697	0.189***
	(0.0103)	(0.0302)	(0.0388)	(0.0110)	(0.0460)	(0.0551)
DRG cost weight	0.0583***	-0.0211	0.0702***	0.0845***	-0.0214	0.0918***
	(0.00540)	(0.0149)	(0.0154)	(0.00652)	(0.0195)	(0.0195)
Complications or comorbidities	0.182***	0.112**	0.358***	0.0103	0.118	0.240**
	(0.0127)	(0.0359)	(0.0427)	(0.0207)	(0.0715)	(0.0782)
Number of transfers	0.0738***	-0.0209	0.0964***	0.0356***	-0.0400	0.114**
	(0.00760)	(0.0203)	(0.0236)	(0.00947)	(0.0340)	(0.0385)
Unit-level utilization	6.516***	-4.307	4.907	17.02***	11.38	36.84***
	(1.400)	(3.997)	(5.716)	(2.037)	(7.545)	(11.13)
(Unit-level utilization)^2	-3.734***	2.705	-2.729	-9.677***	-7.072	-20.92**
	(0.836)	(2.392)	(3.395)	(1.182)	(4.434)	(6.491)
Service-level utilization	12.81***	3.741	19.49*	0.190	-12.58	-22.24
	(2.127)	(6.480)	(9.006)	(2.840)	(10.78)	(13.90)
(Service-level utilization)^2	-8.176***	-2.192	-11.88*	-1.311	7.498	12.21
	(1.195)	(3.658)	(5.115)	(1.679)	(6.490)	(8.369)
Visited ICU during encounter	-0.345***	-0.0291	-0.109	-0.158***	-0.0816	-0.336*
	(0.0196)	(0.0536)	(0.0636)	(0.0333)	(0.112)	(0.147)
Admitted during PM shift	0.0713***	-0.212***	0.0455	0.127***	0.123	0.114
	(0.0185)	(0.0516)	(0.0695)	(0.0236)	(0.0870)	(0.107)
Admitted during overnight shift	0.0400*	-0.253***	0.0750	-0.0166	-0.0685	-0.0807
	(0.0181)	(0.0501)	(0.0677)	(0.0195)	(0.0763)	(0.0950)
Admitted on weekday	0.0275	0.0529	0.0107	-0.0961***	-0.220**	0.0189
	(0.0162)	(0.0468)	(0.0581)	(0.0222)	(0.0795)	(0.109)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Service FE	Yes	Yes	Yes	Yes	Yes	Yes
Unit FE	Yes	Yes	Yes	Yes	Yes	Yes
Model	2SLS 2nd stage	2SRI 2nd stage	2SRI 2nd stage	2SLS 2nd stage	2SRI 2nd stage	2SRI 2nd stage
Observations	8982	8981	8981	5811	5808	5425

Note. Standard errors (in parentheses) are heteroskedasticity robust for continuous outcome variables and bootstrapped for binary outcome variables. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

APPENDIX C

Appendix for "Should We Worry About Moral Hazard? Estimation of the Slutsky Equation Using Indemnity Health Insurance Contracts"

C.1. Full results tables with control variables

Table C.1: Estimation of model 4.1

	(1) Health care expenditure		(2) Health care expenditure	
Supplementary insurance payments	0.572***	(0.170)	0.536**	(0.173)
Indemnity payments based on state	0.0455	(0.0257)	0.0404	(0.0235)
Indemnity payments based on utilization	0.306*	(0.122)	0.279**	(0.106)
Enrolled in supplementary insurance	14.24	(13.97)	15.47	(25.90)
Enrolled in indemnity insurance	48.49***	(11.79)	11.72	(25.12)
Female	49.06***	(9.389)		
Age	5.742***	(0.472)	33.90***	(3.828)
Married	87.00***	(11.79)	101.7*	(42.56)
Employed	-123.4***	(9.850)	-49.45**	(17.76)
Years of education	-2.622	(1.646)	7.584	(12.91)
Logged household income	161.0***	(9.076)	41.91**	(15.85)
Logged number of household members	-260.1***	(14.67)	-163.1***	(43.85)
Childbirth	30.44	(30.14)	-280.3***	(39.03)
Chronic conditions	269.3***	(9.530)	84.68***	(20.67)
Disabled	-38.92	(25.00)	-158.8	(132.7)
Physical limitation	315.1***	(37.08)	299.4***	(43.71)
Days per week engaged in vigorous physical activity	-1.755	(2.647)	0.316	(3.093)
Days per week engaged in moderate physical activity	-3.659	(1.932)	-8.141***	(2.327)
Days per week walked for more than ten minutes	-1.428	(1.672)	-3.643	(2.080)
Self-reported health (low is good)=2	12.94	(13.47)	13.55	(15.49)
Self-reported health (low is good)=3	97.53***	(13.96)	58.57***	(16.72)
Self-reported health (low is good)=4	433.2***	(22.37)	203.1***	(26.39)
Self-reported health (low is good)=5	778.5***	(92.85)	419.0***	(95.01)
Constant	-1444.8***	(89.45)	-1554.6***	(256.3)
Individual fixed effects	No		Yes	
City/province fixed effects	Yes		Yes	
Year fixed effects	Yes		Yes	
Observations	72925		72925	

Note. Robust standard errors in parenthesis. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table C.2: Estimation of model 4.2

	(1) Logged health care expenditure		(2) Logged health care expenditure	
Enrolled in supplementary insurance	0.304***	(0.0191)	0.181***	(0.0328)
Enrolled in indemnity insurance	0.296***	(0.0186)	0.0582	(0.0413)
Female	0.629***	(0.0172)		
Age	0.0225***	(0.000800)	0.0723***	(0.00523)
Married	0.612***	(0.0206)	0.585***	(0.0798)
Employed	-0.267***	(0.0169)	-0.0541*	(0.0272)
Years of education	-0.00771**	(0.00253)	-0.00821	(0.0206)
Logged household income	0.351***	(0.0141)	0.0729***	(0.0215)
Logged number of household members	-0.586***	(0.0240)	-0.546***	(0.0622)
Childbirth	0.644***	(0.0779)	-0.599***	(0.0828)
Chronic conditions	1.665***	(0.0209)	0.507***	(0.0426)
Disabled	-0.0565	(0.0312)	-0.158	(0.101)
Physical limitation	0.273***	(0.0324)	0.210***	(0.0318)
Days per week engaged in vigorous physical activity	-0.000255	(0.00512)	-0.00805	(0.00517)
Days per week engaged in moderate physical activity	0.00970**	(0.00367)	-0.000866	(0.00365)
Days per week walked for more than ten minutes	0.00291	(0.00286)	-0.00454	(0.00287)
Self-reported health (low is good)=2	0.115**	(0.0382)	0.0524	(0.0381)
Self-reported health (low is good)=3	0.412***	(0.0382)	0.182***	(0.0390)
Self-reported health (low is good)=4	0.928***	(0.0417)	0.342***	(0.0427)
Self-reported health (low is good)=5	1.091***	(0.0692)	0.496***	(0.0651)
Constant	-1.820***	(0.149)	-0.00509	(0.417)
Individual fixed effects	No		Yes	
City/province fixed effects	Yes		Yes	
Year fixed effects	Yes		Yes	
Observations	72925		72925	

Note. Robust standard errors in parenthesis. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table C.3: Quadratic specification of model 4.1

	(1) Health care expenditure		(2) Health care expenditure	
Supplementary insurance payments	1.0264***	(.06479)	.9943***	(.06714)
Supplementary insurance payments^2	-3.5e-05***	(3.4e-06)	-3.4e-05***	(3.2e-06)
Indemnity payments based on state	.11416***	(.01472)	.10663***	(.01421)
Indemnity payments based on state^2	-3.0e-07***	(3.8e-08)	-2.8e-07***	(3.6e-08)
Indemnity payments based on utilization	.64394***	(.05707)	.58992***	(.05908)
Indemnity payments based on utilization^2	-5.2e-06***	(5.0e-07)	-4.7e-06***	(5.2e-07)
Enrolled in supplementary insurance	-13.999	(10.286)	-9.9573	(23.595)
Enrolled in indemnity insurance	28.668**	(10.623)	-3.969	(24.392)
Female	44.676***	(9.2443)		
Age	5.8449***	(.467)	33.58***	(3.8006)
Married	82.532***	(11.658)	87.479*	(42.141)
Employed	-123.47***	(9.7247)	-52.211**	(17.315)
Years of education	-2.4147	(1.6292)	8.6913	(12.815)
Logged household income	156.72***	(8.9689)	48.56**	(15.662)
Logged number of household members	-250.67***	(14.427)	-164.78***	(43.166)
Childbirth	47.283	(29.596)	-277.61***	(38.613)
Chronic conditions	259.29***	(9.0308)	88.692***	(20.089)
Disabled	-36.213	(24.991)	-147.36	(131.17)
Physical limitation	298.98***	(36.242)	281.07***	(43.156)
Days per week engaged in vigorous physical activity	-1.9615	(2.6107)	-.34209	(3.0129)
Days per week engaged in moderate physical activity	-3.1425	(1.8914)	-7.6004***	(2.2681)
Days per week walked for more than ten minutes	-1.2317	(1.6511)	-3.4973	(2.0482)
Self-reported health (low is good)=2	3.4535	(13.027)	5.4025	(15.002)
Self-reported health (low is good)=3	85.196***	(13.348)	49.84**	(15.944)
Self-reported health (low is good)=4	406.56***	(20.669)	184.51***	(25.371)
Self-reported health (low is good)=5	733.36***	(91.716)	387.35***	(94.334)
Constant	-1393.5***	(88.319)	-1606.7***	(255.04)
Individual fixed effects	No		Yes	
City/province fixed effects	Yes		Yes	
Year fixed effects	Yes		Yes	
Observations	72925		72925	

Note. Robust standard errors in parenthesis. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table C.4: Self-reported health as predictor of private insurance enrollment

	(1) Indemnity		(2) Supplementary	
Self-reported health lagged by one year				
Good	0.178***	(0.0485)	0.0142	(0.0442)
Moderate	0.146**	(0.0489)	0.000490	(0.0448)
Bad	-0.00449	(0.0558)	-0.156**	(0.0554)
Very bad	-0.124	(0.103)	-0.244	(0.136)
Self-reported health				
Good	0.100	(0.0511)	-0.0368	(0.0455)
Moderate	0.0829	(0.0515)	-0.0122	(0.0461)
Bad	-0.0522	(0.0584)	-0.143*	(0.0567)
Very bad	-0.170	(0.106)	-0.369*	(0.143)
Female	0.387***	(0.0240)	0.329***	(0.0227)
Age	-0.0480***	(0.00117)	-0.0516***	(0.00117)
Married	1.148***	(0.0301)	0.979***	(0.0329)
Employed	0.392***	(0.0240)	0.259***	(0.0237)
Years of education	0.00917*	(0.00358)	0.0232***	(0.00388)
Logged household income	0.841***	(0.0213)	0.438***	(0.0197)
Logged number of household members	-0.860***	(0.0362)	-0.404***	(0.0353)
Childbirth	-0.184	(0.143)	0.0324	(0.101)
Chronic conditions	0.285***	(0.0287)	0.0725**	(0.0246)
Disabled	-0.248***	(0.0425)	-0.490***	(0.0609)
Physical limitation	-0.232***	(0.0476)	-0.305***	(0.0677)
Days per week engaged in vigorous physical activity	0.0200**	(0.00744)	0.00165	(0.00668)
Days per week engaged in moderate physical activity	0.0134**	(0.00521)	0.00203	(0.00506)
Days per week walked for more than ten minutes	-0.00503	(0.00404)	-0.00379	(0.00394)
Constant	-6.304***	(0.228)	-4.382***	(0.213)
Model	Logit		Logit	
City/province fixed effects	Yes		Yes	
Year fixed effects	Yes		Yes	
Observations	54278		54278	

Note. Self-reported health: Excellent is the baseline. Robust standard errors in parenthesis. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table C.5: Estimation of model (4.1) among marginal consumers

	(1) Health care expenditure		(2) Health care expenditure	
Supplementary insurance payments	0.861***	(0.0927)	0.799***	(0.0965)
Indemnity payments based on state	0.140***	(0.0269)	0.142***	(0.00424)
Indemnity payments based on utilization	0.646***	(0.119)	0.868	(0.456)
Enrolled in supplementary insurance	-23.34	(35.42)	326.7	(208.0)
Female	66.57	(35.41)		
Age	5.209	(2.871)	11.50	(28.47)
Married	72.34	(65.41)	531.7	(370.6)
Employed	-107.3*	(46.66)	149.1	(139.5)
Years of education	-21.29*	(10.75)	-23.08	(88.98)
Logged household income	149.9***	(41.34)	30.87	(93.60)
Logged number of household members	-137.9	(71.98)	298.9	(249.7)
Childbirth	400.8	(286.8)	-463.8*	(205.2)
Chronic conditions	228.5***	(41.21)	-227.1	(215.8)
Disabled	-261.8	(152.1)	8.272	(118.1)
Physical limitation	49.35	(224.5)	283.8	(356.2)
Days per week engaged in vigorous physical activity	4.220	(14.58)	-42.98	(28.78)
Days per week engaged in moderate physical activity	-6.657	(10.06)	15.27	(20.53)
Days per week walked for more than ten minutes	-8.862	(7.474)	0.346	(16.99)
Self-reported health (low is good)=2	-70.52	(61.53)	63.18	(74.94)
Self-reported health (low is good)=3	24.16	(64.73)	83.70	(91.41)
Self-reported health (low is good)=4	196.1*	(97.30)	393.1**	(138.7)
Self-reported health (low is good)=5	1281.0	(824.0)	439.3	(288.0)
Constant	-1006.6*	(439.1)	-719.8	(1891.8)
Individual fixed effects	No		Yes	
City/province fixed effects	Yes		Yes	
Year fixed effects	Yes		Yes	
Observations	2549		2549	

Note. Robust standard errors in parenthesis. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

BIBLIOGRAPHY

Daniel Adelman. An Efficient Frontier Approach to Scoring and Ranking Hospital Performance. *Operations Research*, 68(3):762–792, may 2020. ISSN 0030-364X. doi: 10.1287/opre.2019.1972. URL http://pubsonline.informs.org/doi/10.1287/opre.2019.1972.

Elodie Adida and Fernanda Bravo. Contracts for healthcare referral services: Coordination via outcome-based penalty contracts. *Management Science*, 65(3):1322–1341, 2019. ISSN 15265501. doi: 10.1287/mnsc.2017.3000.

Elodie Adida, Hamed Mamani, and Shima Nassiri. Bundled payment vs. fee-for-service: Impact of payment scheme on performance. *Management Science*, 63(5):1606–1624, 2017. ISSN 15265501. doi: 10.1287/mnsc.2016.2445.

AHRQ. Central Line-Associated Bloodstream Infections (CLABSI), 2021. URL https://www.ahrq.gov/topics/central-line-associated-bloodstream-infections-clabsi.html.

Linda H. Aiken, Sean P. Clarke, Douglas M. Sloane, Julie Sochalski, and Jeffrey H. Silber. Hospital Nurse Staffing and Patient Mortality, Nurse Burnout, and Job Dissatisfaction. *JAMA*, 288(16):1987–1993, oct 2002. ISSN 0098-7484. doi: 10.1001/jama.288.16.1987. URL http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.288.16.1987.

César Alameda and Carmen Suárez. Clinical outcomes in medical outliers admitted to hospital with heart failure. *European Journal of Internal Medicine*, 20(8):764–767, 2009. ISSN 09536205. doi: 10.1016/j.ejim.2009.09.010. URL http://dx.doi.org/10.1016/j.ejim.2009.09.010.

Dimitrios A. Andritsos and Christopher S. Tang. Incentive Programs for Reducing Readmissions when Patient Care is Co-Produced. *Production and Operations Management*, 27(6):999–1020, 2018. ISSN 19375956. doi: 10.1111/poms.12847.

Kenan Arifoğlu, Hang Ren, and Tolga Tezcan. Hospital readmissions reduction program does not provide the right incentives: Issues and remedies. *Management Science*, 67(4):2191–2210, 2021. ISSN 15265501. doi: 10.1287/mnsc.2020.3649.

Anil Aswani, Zuo Jun Max Shen, and Auyon Siddiq. Data-driven incentive design in the medicare shared savings program. *Operations Research*, 67(4):1002–1026, 2019. ISSN 15265463. doi: 10.1287/opre.2018.1821.

Anthony D. Bai, Siddhartha Srivastava, George A. Tomlinson, Christopher A. Smith, Chaim M. Bell, and Sudeep S. Gill. Mortality of hospitalised internal medicine patients bedspaced to non-internal medicine inpatient units: Retrospective cohort study. *BMJ Quality and Safety*, 27(1):11–20, 2018. ISSN 20445415. doi: 10.1136/bmjqs-2017-006925.

Patrick Bajari, C. Lanier Benkard, and Jonathan Levin. Estimating dynamic models of imperfect competition. *Econometrica*, 75(5):1331–1370, 2007. ISSN 00129682. doi: 10.1111/j.1468-0262.2007.00796.x.

Ann P. Bartel, Carri W. Chan, and Song-Hee Kim. Should Hospitals Keep Their Patients Longer? The Role of Inpatient Care in Reducing Postdischarge Mortality. *Management Science*, 66(6):2326–2346, jun 2020. ISSN 0025-1909. doi: 10.1287/mnsc.2019.3325. URL http://pubsonline.informs.org/doi/10.1287/mnsc.2019.3325.

Hamsa Bastani, Joel Goh, and Mohsen Bayati. Evidence of upcoding in pay-for-performance programs. *Management Science*, 65(3):1042–1060, 2019. ISSN 15265501. doi: 10.1287/mnsc.2017.2996.

Robert Batt, Hessam Bavafa, and Mohamad Soltani. Quality Improvement Spillovers: Evidence From the Hospital Readmissions Reduction Program. *SSRN Electronic Journal*, pages 1–45, 2020. ISSN 1556-5068. doi: 10.2139/ssrn.3132770.

RT Benson, JC Drew, and RB Galland. A Waiting List to Go Home: An Analysis of Delayed Discharges From Surgical Beds. *The Annals of The Royal College of Surgeons of England*, 88(7):650–652, nov 2006. ISSN 0035-8843. doi: 10.1308/003588406X149246. URL https://publishing.rcseng.ac.uk/doi/10.1308/003588406X149246.

Jillian A. Berry Jaeker and Anita L. Tucker. Past the Point of Speeding Up: The Negative Effects of Workload Saturation on Efficiency and Patient Severity. *Management Science*, 63(4):1042–1062, apr 2017. ISSN 0025-1909. doi: 10.1287/mnsc.2015.2387. URL http://pubsonline.informs.org/doi/10.1287/mnsc.2015.2387.

Dimitris Bertsimas and Jean Pauphilet. Hospital-wide Patient Flow Optimization. *Working paper*, 2020.

Mark J. Browne. Evidence of Adverse Selection in the Individual Health Insurance Market. *The Journal of Risk and Insurance*, 59(1):13, mar 1992. ISSN 00224367. doi: 10.2307/253214. URL https://www.jstor.org/stable/253214?origin=crossref.

Donald B. Chalfin, Stephen Trzeciak, Antonios Likourezos, Brigitte M. Baumann, and R. Phillip Dellinger. Impact of delayed transfer of critically ill patients from the emergency department to the intensive care unit. *Critical Care Medicine*, 35(6):1477–1483, jun 2007. ISSN 0090-3493. doi: 10.1097/01.CCM.0000266585.74905.5A. URL http://journals.lww.com/00003246-200706000-00004.

Carri W. Chan, Linda V. Green, Suparerk Lekwijit, Lijian Lu, and Gabriel Escobar. Assessing the Impact of Service Level When Customer Needs Are Uncertain: An Empirical Investigation of Hospital Step-Down Units. *Management Science*, 65(2):751–775, feb 2019. ISSN 0025-1909. doi: 10.1287/mnsc.2017.2974. URL http://pubsonline.informs.org/doi/10.1287/mnsc.2017.2974.

Suresh Chand, Herbert Moskowitz, John B. Norris, Steve Shade, and Deanna R. Willis. Improving patient flow at an outpatient clinic: study of sources of variability and improvement factors. *Health Care Management Science*, 12:325–340, sep 2009. ISSN 1386-9620. doi: 10.1007/s10729-008-9094-3. URL http://link.springer.com/10.1007/s10729-008-9094-3.

Christopher Chen and Nicos Savva. Unintended Consequences of Hospital Regulation: The Case of the Hospital Readmissions Reduction Program. *SSRN Electronic Journal*, 2018. ISSN 1556-5068. doi: 10.2139/ssrn.3236983.

Pierre André Chiappori and Bernard Salanié. Testing for asymmetric information in insurance markets. *Journal of Political Economy*, 108(1):56–78, 2000. ISSN 00223808. doi: 10.1086/262111.

Gichun Choi and Hyeonbok Lee. The Proper Role of National Health Insurance and Private Health Insurance. *Health and Welfare Policy Forum*, 248(6):30–42, 2017. doi: 10.23062/2017.06.4.

Simon Cowan. Competition in the water industry. *Oxford Review of Economic Policy*, 13 (1):83–92, 1997. ISSN 0266903X. doi: 10.1093/oxrep/13.1.83.

Maqbool Dada and William D. White. Evaluating financial risk in the Medicare Prospective Payment System. *Management Science*, 45(3):316–329, 1999. ISSN 00251909. doi: 10.1287/mnsc.45.3.316.

J G Dai and Pengyi Shi. Inpatient Overflow: An Approximate Dynamic Programming Approach. *Manufacturing & Service Operations Management*, 21(4):894–911, oct 2019. ISSN 1523-4614. doi: 10.1287/msom.2018.0730. URL http://pubsonline.informs.org/doi/10.1287/msom.2018.0730.

Jim G. Dai and Pengyi Shi. Recent Modeling and Analytical Advances in Hospital Inpatient Flow Management. *Production and Operations Management*, jan 2020. ISSN 1059-1478. doi: 10.1111/poms.13132. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/poms.13132.

Dag Morten Dalen. Yardstick competition and investment incentives. *Journal of Economics and Management Strategy*, 7(1):105–126, 1998. ISSN 10586407. doi: 10.1162/105864098567353.

Suzanne de Treville and John Antonakis. Could lean production job design be intrinsically motivating? Contextual, configurational, and levels-of-analysis issues. *Journal of Operations Management*, 24:99–123, jan 2006. ISSN 02726963. doi: 10.1016/j.jom.2005.04.001. URL http://doi.wiley.com/10.1016/j.jom.2005.04.001.

Gregory Dobson, Edieal Pinker, and R. Lawrence Van Horn. Division of Labor in Medical

Office Practices. *Manufacturing & Service Operations Management*, 11(3):525–537, jul 2009. ISSN 1523-4614. doi: 10.1287/msom.1080.0236. URL http://pubsonline.informs.org/doi/abs/10.1287/msom.1080.0236.

Jing Dong and Ohad Perry. Queueing Models for Patient-Flow Dynamics in Inpatient Wards. *Operations Research*, 68(1):250–275, jan 2020. ISSN 0030-364X. doi: 10.1287/opre.2019.1845. URL http://pubsonline.informs.org/doi/10.1287/opre.2019.1845.

Jing Dong, Pengyi Shi, Fanyin Zheng, and Xin Jin. Structural Estimation of Load Balancing Behavior in Inpatient Ward Network. *Working paper*, 2020.

David Dranove. Rate-Setting by Diagnosis Related Groups and Hospital Specialization. *The RAND Journal of Economics*, 18(3):417, 1987. ISSN 07416261. doi: 10.2307/2555606.

Randall P. Ellis and Thomas G. McGuire. Provider behavior under prospective reimbursement. Cost sharing and supply. *Journal of Health Economics*, 5(2):129–151, 1986. ISSN 01676296. doi: 10.1016/0167-6296(86)90002-0.

Richard Ericson and Ariel Pakes. Markov-Perfect Industry Dynamics: A Framework for Empirical Work. *The Review of Economic Studies*, 62(1):53, jan 1995. ISSN 00346527. doi: 10.2307/2297841. URL https://academic.oup.com/restud/article-lookup/doi/10.2307/2297841.

Antonio Estache, Marianela GonzÁlez, and Lourdes Trujillo. Efficiency gains from port reform and the potential for yardstick competition: Lessons from Mexico. *World Development*, 30(4):545–560, 2002. ISSN 0305750X. doi: 10.1016/S0305-750X(01)00129-2.

Marshall Fisher and Ananth Raman. Reducing the Cost of Demand Uncertainty Through Accurate Response to Early Sales. *Operations Research*, 44(1):87–99, feb 1996. ISSN 0030-364X. doi: 10.1287/opre.44.1.87. URL http://pubsonline.informs.org/doi/abs/10.1287/opre.44.1.87.

Deborah Gesensway. Having Problems Finding Your Patients?, 2010. URL https://www.todayshospitalist.com/having-problems-finding-your-patients/.

Malcolm Gladwell. The Moral-Hazard Myth, aug 2005. URL https://www-newyorker-com.proxy.library.upenn.edu/magazine/2005/08/29/the-moral-hazard-myth.

Pengfei Guo, Christopher S. Tang, Yulan Wang, and Ming Zhao. The impact of reimbursement policy on social welfare, revisit rate, and waiting time in a public healthcare system: Fee-for-service versus bundled payment. *Manufacturing and Service Operations Management*, 21(1):154–170, 2019. ISSN 15265498. doi: 10.1287/msom.2017.0690.

Martin B. Hackmann, Jonathan T. Kolstad, and Amanda E. Kowalski. Adverse selection and an individual mandate: When theory meets practice. *American Economic Review*, 105(3):1030–1066, 2015. ISSN 19447981. doi: 10.1257/aer.20130758.

Y. Haddadin, P. Annamaraju, and H. Regunath. *StatPearls*. StatPearls Publishing, Treasure Island, FL, 2021. URL https://www.ncbi.nlm.nih.gov/books/NBK430891/.

Jerry A. Hausman. Exact Consumer ' s Surplus and Deadweight Loss. *The American Review*, 71(4):662–676, 1981. ISSN 00028282.

Young Rock Hong, Oliver Nguyen, Sandhya Yadav, Emma Etzold, Jihee Song, R. Paul Duncan, and Kea Turner. Early Performance of Hospital Value-based Purchasing Program in Medicare: A Systematic Review. *Medical Care*, 58(8):734–743, 2020. ISSN 15371948. doi: 10.1097/MLR.0000000000001354.

Navid Izady and Israa Mohamed. A Clustered Overflow Configuration of Inpatient Beds in Hospitals. *Manufacturing & Service Operations Management*, 23(1):139–154, jan 2021. ISSN 1523-4614. doi: 10.1287/msom.2019.0820. URL http://pubsonline.informs.org/doi/10.1287/msom.2019.0820.

Houyuan Jiang, Zhan Pang, and Sergei Savin. Performance-based contracts for outpatient medical services. *Manufacturing and Service Operations Management*, 14(4):654–669, 2012. ISSN 15234614. doi: 10.1287/msom.1120.0402.

Hee-Chung Kang, Yunseop Oh, Hyeyeon Baek, Solip Ha, Soun Kim, Eunwon Suh, Jaeseok Hong, Jongheon Park, and Haegon Cho. Assessment of the Innovation Outcome in the Korea Health Care System: 2017 Korea Healthcare Quality Report. Technical report, Korea Institute for Health and Social Affairs, 2017.

Diwas S. Kc and Christian Terwiesch. Impact of Workload on Service Time and Patient Safety: An Econometric Analysis of Hospital Operations. *Management Science*, 55(9):1486–1498, sep 2009. ISSN 0025-1909. doi: 10.1287/mnsc.1090.1037. URL http://pubsonline.informs.org/doi/abs/10.1287/mnsc.1090.1037.

Diwas Singh KC and Christian Terwiesch. The Effects of Focus on Performance: Evidence from California Hospitals. *Management Science*, 57(11):1897–1912, nov 2011. ISSN 0025-1909. doi: 10.1287/mnsc.1110.1401. URL http://pubsonline.informs.org/doi/abs/10.1287/mnsc.1110.1401.

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, jan 2015. ISSN 0025-1909. doi: 10.1287/mnsc.2014.2057. URL http://pubsonline.informs.org/doi/abs/10.1287/mnsc.2014.2057.

Song-Hee Kim, Fanyin Zheng, and Joan Brown. Identifying the Bottleneck Unit: Impact of Congestion Spillover in Hospital Inpatient Unit Network. *Working paper*, pages 1–35, 2020. ISSN 1556-5068. doi: 10.2139/ssrn.3667970. URL https://www.ssrn.com/abstract=

3667970.

Rachel Kohn, Michael O. Harhay, Brian Bayes, Hummy Song, Scott D. Halpern, Meeta Prasad Kerlin, and S. Ryan Greysen. Influence of bedspacing on outcomes of hospitalised medicine service patients: a retrospective cohort study. *BMJ Quality & Safety*, apr 2020a. ISSN 2044-5415. doi: 10.1136/bmjqs-2019-010675. URL https://qualitysafety.bmj.com/lookup/doi/10.1136/bmjqs-2019-010675.

Rachel Kohn, Michael O. Harhay, Gary E. Weissman, George L. Anesi, Brian Bayes, Hummy Song, Scott D. Halpern, S. Ryan Greysen, and Meeta Prasad Kerlin. The Association of Geographic Dispersion with Outcomes among Hospitalized Pulmonary Service Patients. *Annals of the American Thoracic Society*, 17(2):249–252, feb 2020b. ISSN 2329-6933. doi: 10.1513/AnnalsATS.201906-471RL. URL https://www.atsjournals.org/doi/10.1513/AnnalsATS.201906-471RL.

Ludwig Kuntz, Roman Mennicken, and Stefan Scholtes. Stress on the Ward: Evidence of Safety Tipping Points in Hospitals. *Management Science*, 61(4):754–771, apr 2015. ISSN 0025-1909. doi: 10.1287/mnsc.2014.1917. URL http://pubsonline.informs.org/doi/10.1287/mnsc.2014.1917.

Ludwig Kuntz, Stefan Scholtes, and Sandra Sülz. Separate and Concentrate: Accounting for Patient Complexity in General Hospitals. *Management Science*, 65(6):2482–2501, jun 2019. ISSN 0025-1909. doi: 10.1287/mnsc.2018.3064. URL http://pubsonline.informs.org/doi/10.1287/mnsc.2018.3064.

Timo Kuosmanen and Andrew L. Johnson. Conditional yardstick competition in energy regulation. *Energy Journal*, 41:67–92, 2020. ISSN 01956574. doi: 10.5547/01956574.41.SI1.TKUO.

Soonman Kwon, Tae-jin Lee, and Chang-yup Kim. *Republic of Korea Health System Review*. Health Systems in Transition Vol. 5 No. 4 2015. World Health Organization, 2015. ISBN 9789290617105. URL https://apps.who.int/iris/handle/10665/208215.

Donald K.K. Lee and Stefanos A. Zenios. An evidence-based incentive system for Medicare's end-stage renal disease program. *Management Science*, 58(6):1092–1105, 2012. ISSN 00251909. doi: 10.1287/mnsc.1110.1471.

Hau L. Lee and Christopher S. Tang. Variability Reduction Through Operations Reversal. *Management Science*, 44(2):162–172, feb 1998. ISSN 0025-1909. doi: 10.1287/mnsc.44.2.162. URL http://pubsonline.informs.org/doi/abs/10.1287/mnsc.44.2.162.

Hau L. Lee, V. Padmanabhan, and Seungjin Whang. Information Distortion in a Supply Chain: The Bullwhip Effect. *Management Science*, 50(12_supplement):1875–1886, dec 2004. ISSN 0025-1909. doi: 10.1287/mnsc.1040.0266. URL http://pubsonline.informs.org/doi/abs/10.1287/mnsc.1040.0266.

Seung Jun Lee, Sriram Venkataraman, Gregory R. Heim, Aleda V. Roth, and Jon Chilingerian. Impact of the value-based purchasing program on hospital operations outcomes: An econometric analysis. *Journal of Operations Management*, 66(1-2):151–175, 2020. ISSN 02726963. doi: 10.1002/joom.1057.

Yassine Lefouili. Does competition spur innovation? The case of yardstick competition. *Economics Letters*, 137:135–139, 2015. ISSN 01651765. doi: 10.1016/j.econlet.2015.10.033.

Eugene Litvak, Peter I. Buerhaus, Frank Davidoff, Michael C. Long, Michael L. McManus, and Donald M. Berwick. Managing Unnecessary Variability in Patient Demand to Reduce Nursing Stress and Improve Patient Safety. *The Joint Commission Journal on Quality and Patient Safety*, 31(6):330–338, jun 2005. ISSN 15537250. doi: 10.1016/S1553-7250(05)31044-0. URL https://linkinghub.elsevier.com/retrieve/pii/S1553725005310440.

W. G. Manning, J. P. Newhouse, N. Duan, E. B. Keeler, A. Leibowitz, and M. S. Marquis. Health insurance and the demand for medical care: evidence from a randomized experiment. *American Economic Review*, 77(3):251–277, 1987. ISSN 00028282.

Rui Cunha Marques. A yardstick competition model for Portuguese water and sewerage services regulation. *Utilities Policy*, 14(3):175–184, 2006. ISSN 09571787. doi: 10.1016/j.jup.2006.03.004.

Kusum S. Mathews and Elisa F. Long. A Conceptual Framework for Improving Critical Care Patient Flow and Bed Use. *Annals of the American Thoracic Society*, 12(6):886–894, jun 2015. ISSN 2329-6933. doi: 10.1513/AnnalsATS.201409-419OC. URL http://www.atsjournals.org/doi/10.1513/AnnalsATS.201409-419OC.

Lesley Meng, Robert Batt, and Christian Terwiesch. The Impact of Facility Layout on Worker Behavior: An Empirical Study of Nurses in the Emergency Department. *Working paper*, 2020. ISSN 1556-5068. doi: 10.2139/ssrn.3143883.

F. Mizutani. Empirical analysis of yardstick competition in the Japanese railway industry. *International Journal of Transport Economics*, 24(3):367–392, 1997.

National Center for Health Statistics. *Health, United States, 2016: With Chartbook on Long-term Trends in Health*. Hyattsville, MD, 2017.

John A. Nyman. The economics of moral hazard revisited. *Journal of Health Economics*, 18(6):811–824, 1999. ISSN 01676296. doi: 10.1016/S0167-6296(99)00015-6.

Ariel Pakes and Paul McGuire. Computing Markov-Perfect Nash Equilibria: Numerical Implications of a Dynamic Differentiated Product Model. *The RAND Journal of Economics*, 25(4):555, 1994. ISSN 07416261. doi: 10.2307/2555975. URL http://doi.wiley.com/10.2307/2555975.

Mark Pauly. The Economics of Moral Hazard: Comment. *The American Economic Review*, 58(3):531–537, 1968.

Martin Pesendorfer and Philipp Schmidt-Dengler. Identification and Estimation of Dynamic Games. may 2003. URL http://www.nber.org/papers/w9726.pdf.

Elaine Rabin, Keith Kocher, Mark McClelland, Jesse Pines, Ula Hwang, Niels Rathlev, Brent Asplin, N. Seth Trueger, and Ellen Weber. Solutions To Emergency Department 'Boarding' And Crowding Are Underused And May Need To Be Legislated. *Health Affairs*, 31(8):1757–1766, aug 2012. ISSN 0278-2715. doi: 10.1377/hlthaff.2011.0786. URL http://www.healthaffairs.org/doi/10.1377/hlthaff.2011.0786.

Sampath Rajagopalan and Chunyang Tong. Payment Models to Coordinate Healthcare Providers with Partial Attribution of Outcome Costs. *Manufacturing & Service Operations Management*, (June), 2021. ISSN 1523-4614. doi: 10.1287/msom.2020.0928.

Ray Reagans, Linda Argote, and Daria Brooks. Individual Experience and Experience Working Together: Predicting Learning Rates from Knowing Who Knows What and Knowing How to Work Together. *Management Science*, 51(6):869–881, jun 2005. ISSN 0025-1909. doi: 10.1287/mnsc.1050.0366. URL http://pubsonline.informs.org/doi/abs/10.1287/mnsc.1050.0366.

P. M. Robinson. Root-N-Consistent Semiparametric Regression. *Econometrica*, 56(4):931–954, jul 1988. ISSN 00129682. doi: 10.2307/1912705. URL https://search.proquest.com/docview/220297257?accountid=12834https://www.jstor.org/stable/1912705?origin=crossref.

John Rust. Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher. *Econometrica*, 55(5):999, sep 1987. ISSN 00129682. doi: 10.2307/1911259. URL https://www.jstor.org/stable/1911259?origin=crossref.

Andrew M. Ryan, Sam Krinsky, Kristin A. Maurer, and Justin B. Dimick. Changes in Hospital Quality Associated with Hospital Value-Based Purchasing. *New England Journal of Medicine*, 376(24):2358–2366, 2017. ISSN 0028-4793. doi: 10.1056/nejmsa1613412.

Nicos Savva, Tolga Tezcan, and Özlem Yıldız. Can yardstick competition reduce waiting times? *Management Science*, 65(7):3196–3215, 2019. ISSN 15265501. doi: 10.1287/mnsc.2018.3089.

John W. Sawkins. Yardstick competition in the English and Welsh water industry Fiction or reality? *Utilities Policy*, 5(1):27–36, 1995. ISSN 09571787. doi: 10.1016/0957-1787(95)00011-N.

Pengyi Shi, Mabel C. Chou, J. 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, jan 2016. ISSN 0025-1909. doi: 10.1287/mnsc.2014.2112. URL http://pubsonline.informs.org/doi/10.1287/mnsc.2014.2112.

Gicheol Shin. Discussions on the Empirical Study on Health Care Consumptions by Private Health Insurance Policyholders. *Health and Social Welfare Review*, 35(4):5–34, dec 2015. ISSN 1226-072X. doi: 10.15709/hswr.2015.35.4.5. URL http://www.kci.go.kr/kciportal/landing/article.kci?arti{_}id=ART002069285.

Andrei Shleifer. A Theory of Yardstick Competition. *The RAND Journal of Economics*, pages 319–327, 1985.

Kut C. So and Christopher S. Tang. Modeling the impact of an outcome-oriented reimbursement policy on clinic, patients, and pharmaceutical firms. *Management Science*, 46 (7):875–892, 2000. ISSN 00251909. doi: 10.1287/mnsc.46.7.875.12040.

Hummy Song, Anita L. Tucker, Ryan Graue, Sarah Moravick, and Julius J. Yang. Capacity Pooling in Hospitals: The Hidden Consequences of Off-Service Placement. *Management Science*, 66(9):3825–3842, sep 2020. ISSN 0025-1909. doi: 10.1287/mnsc.2019.3395. URL http://pubsonline.informs.org/doi/10.1287/mnsc.2019.3395.

Olan A. Soremekun, Christian Terwiesch, and Jesse M. Pines. Emergency Medicine: An Operations Management View. *Academic Emergency Medicine*, 18(12):1262–1268, dec 2011. ISSN 10696563. doi: 10.1111/j.1553-2712.2011.01226.x. URL http://doi.wiley.com/10.1111/j.1553-2712.2011.01226.x.

S. Sriram, Pradeep K. Chintagunta, and Puneet Manchanda. Service Quality Variability and Termination Behavior. *Management Science*, 61(11):2739–2759, nov 2015. ISSN 0025-1909. doi: 10.1287/mnsc.2014.2105. URL http://pubsonline.informs.org/doi/10.1287/mnsc.2014.2105.

Neophytos Stylianou, Robin Fackrell, and Christos Vasilakis. Are medical outliers associated with worse patient outcomes? A retrospective study within a regional NHS hospital using routine data. *BMJ Open*, 7:e015676, may 2017. ISSN 2044-6055. doi: 10.1136/bmjopen-2016-015676. URL https://bmjopen.bmj.com/lookup/doi/10.1136/bmjopen-2016-015676.

Thomas P. Tangerås. Collusion-proof yardstick competition. *Journal of Public Economics*, 83(2):231–254, 2002. ISSN 00472727. doi: 10.1016/S0047-2727(00)00173-0.

Thomas P. Tangeras. Yardstick competition and quality. *Journal of Economics and Management Strategy*, 18(2):589–613, 2009. ISSN 10586407. doi: 10.1111/j.1530-9134.2009.00223.x.

Guihua Wang, Jun Li, Wallace J. Hopp, Franco L. Fazzalari, and Steven F. Bolling. Using Patient-Specific Quality Information to Unlock Hidden Healthcare Capabilities. *Manufac-*

turing & Service Operations Management, 21(3):582–601, jul 2019. ISSN 1523-4614. doi: 10.1287/msom.2018.0709. URL http://pubsonline.informs.org/doi/10.1287/msom.2018.0709.

Jingui Xie, Weifen Zhuang, Marcus Ang, Mabel C. Chou, Li Luo, and David D. Yao. Analytics for Hospital Resource Planning—Two Case Studies. *Production and Operations Management*, feb 2020. ISSN 1059-1478. doi: 10.1111/poms.13158. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/poms.13158.

Linli Xu, Jorge M. Silva-Risso, and Kenneth C. Wilbur. Dynamic Quality Ladder Model Predictions in Nonrandom Holdout Samples. *Management Science*, 64(7):3187–3207, jul 2018. ISSN 0025-1909. doi: 10.1287/mnsc.2017.2780. URL http://pubsonline.informs.org/doi/10.1287/mnsc.2017.2780.

S. Ylvinger. The operation of Swedish motor-vehicle inspections: efficiency and some problems concerning regulation. *Transportation*, 25:23–36, 1998. ISSN 0091-6749. URL https://doi.org/10.1023/A:1004918613194.

Dennis J. Zhang, Itai Gurvich, Jan A. Van Mieghem, Eric Park, Robert S. Young, and Mark V. Williams. Hospital readmissions reduction program: An economic and operational analysis. *Management Science*, 62(11):3351–3371, 2016. ISSN 15265501. doi: 10.1287/mnsc.2015.2280.