

Alexander, Diane

**Working Paper**

## How do doctors respond to incentives? Unintended consequences of paying doctors to reduce costs

Working Paper, No. 2017-09

**Provided in Cooperation with:**

Federal Reserve Bank of Chicago

*Suggested Citation:* Alexander, Diane (2017) : How do doctors respond to incentives? Unintended consequences of paying doctors to reduce costs, Working Paper, No. 2017-09, Federal Reserve Bank of Chicago, Chicago, IL

This Version is available at:

<https://hdl.handle.net/10419/200599>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



Federal Reserve Bank of Chicago

## **How do Doctors Respond to Incentives? Unintended Consequences of Paying Doctors to Reduce Costs**

*Diane Alexander*

April 10, 2017

WP 2017-09

*\*Working papers are not edited, and all opinions and errors are the responsibility of the author(s). The views expressed do not necessarily reflect the views of the Federal Reserve Bank of Chicago or the Federal Reserve System.*

# How do Doctors Respond to Incentives? Unintended Consequences of Paying Doctors to Reduce Costs\*

Diane Alexander<sup>†</sup>

April 10, 2017

## Abstract

Billions of dollars have been spent on pilot programs searching for ways to reduce health-care costs. I study one such program, where hospitals pay doctors bonuses for reducing the total hospital costs of admitted Medicare patients (a “bundled payment”). Doctors respond to the bonuses by becoming more likely to admit patients whose treatment can generate high bonuses, and sorting healthier patients into participating hospitals. Conditional on patient health, however, doctors do not reduce costs or change procedure use. These results highlight the ability of doctors to game incentive schemes, and the risks of basing nationwide healthcare reforms on pilot programs.

---

\*I thank Janet Currie, Ilyana Kuziemko, Hank Farber, Jon Gruber, Tom Vogl, Anne Case, and seminar participants at Princeton University, the University of Virginia, the University of Maryland, MIT Sloan, Northwestern Kellogg, the Federal Reserve Bank of Chicago, and the University of Notre Dame. In addition, I am indebted to participants of the NBER Summer Institute Health Care workshop for their helpful comments. I also thank the staff at the New Jersey Department of Health for their help accessing the data, and the Center for Health and Wellbeing for financial support.

<sup>†</sup>Federal Reserve Bank of Chicago

# 1 Introduction

Lowering the growth in health care costs has long been a top U.S. public policy goal. Yet while many ideas exist for how to reduce costs, there is no consensus on which path is most promising (Gruber, 2008, 2010). Because of this uncertainty, the Patient Protection and Affordable Care Act (ACA) earmarked billions of dollars for pilot programs.<sup>1</sup> The ACA’s strategy is to try “virtually every cost-control reform proposed by doctors, economists, and health policy experts and [include] the means for these reforms to be assessed quickly and scaled up if they’re successful,” thus ensuring “that effective change will occur” (Orszag and Emanuel, 2010).

A large set of these pilot programs focus on changing the financial incentives of doctors—motivated by the idea that the current system of paying doctors separately for each service provided (“fee-for-service”) encourages them to perform unnecessary procedures. These pilot programs purport to study how doctors respond to different payment schemes, an important open question in the literature. However, I demonstrate that the small-scale nature of pilot programs leaves them susceptible to gaming and selection bias, and thus the information they generate may not be informative for a nationwide reform.

I analyze the effects of the New Jersey Gainsharing Demonstration, a pilot program where hospitals paid doctors bonuses for reducing the total hospital costs of treatment for Medicare admissions. The bonuses were designed to experiment with bundled payments—an incentive scheme where doctors are paid one fee for treating a patient, rather than separately for each service provided. The bonuses are supposed to lower hospital costs by reducing the incentive to provide treatments with low marginal benefits. I find no evidence that the bonuses reduced costs; instead, doctors responded by changing whether and where patients were admitted to hospital.

My primary empirical strategy leverages the fact that many doctors in New Jersey treat patients in more than one hospital, working as independent contractors. I exploit this institutional feature to measure the effect of the bonuses by comparing changes in a doctor’s behavior at a participating hospital to the same doctor in a non-participating hospital. I am able to estimate this within doctor specification due to a unique data set constructed from confidential hospital discharge records, which allow me to follow both patients and doctors over time, and across all hospitals in New Jersey.

---

<sup>1</sup>The Center for Medicare and Medicaid Innovation was established by Section 3021 of the Affordable Care Act (ACA). The Innovation Center is tasked with testing innovative health care payment and service delivery models with the potential to improve the quality of care and reduce Medicare, Medicaid, and CHIP expenditures. The ACA appropriated \$10 billion for the Innovation Center from FY 2011 to FY 2019 (<http://www.hhs.gov/about/budget/fy2015/budget-in-brief/cms/innovation-programs/index.html>).

Doctors respond to the cost-reduction bonuses by changing the patterns of admission across the patient types defined by the bonus formula. Under the program, patients are divided into types by diagnosis and severity of illness categories, and a maximum bonus is assigned to each type. Doctors are then paid a fraction of this maximum bonus, depending on how close the total treatment cost is to pre-program cost benchmarks. I find that the introduction of the bonuses is associated with an increase the likelihood that patients in high-bonus types are admitted, relative to baseline. Conversely, patients in low-bonus types are less likely to be admitted. As capacity constraints and program rules limit the ability of doctors to increase overall admission rates, doctors instead reallocate admission across patients.

Simultaneously, doctors divert healthier patients into participating hospitals, as they are cheaper to treat on average. After the bonus program was implemented, the mix of patients who were admitted at participating hospitals became ex-ante healthier. Patients admitted to participating hospitals had fewer chronic conditions and lower scores on co-morbidity indices based on previous visits, conditional on their type. As healthier patients are cheaper to treat, doctors receive higher bonuses for treating these patients, on average. Defining the bonuses within diagnosis and severity level cells was meant to serve as a type of risk-adjustment. However, I find that doctors are able to identify low-cost patients even within these groups, and sort patients across hospitals in order to increase their expected bonus payments.

Yet, conditional on admission and patient health, the bonuses did not reduce costs or change procedure use. I look at many measures of services performed: length of stay, the use of diagnostic imaging procedures labeled as overused by doctors (CT scans, MRIs, and other diagnostic imaging procedures), and total costs. I find no evidence that doctors change costs or procedure use in response to the program, relative to their behavior at non-participating hospitals. The bonuses create two conflicting forces which may explain why the program did not decrease costs, conditional on patient health. First, there is the intended effect: less care is provided if a patient is admitted under the bonus program than if they were admitted in a hospital with no bonuses. On the other hand, the bonus program causes doctors to admit some patients who otherwise would not have been admitted, and admitted patients receive more care.

It is concerning that doctors changed the composition of admitted patients to maximize their bonuses, as the program was not designed to optimize admission rates across patient types. Hospital admission is an important outcome, both in terms of costs and patient health. For the patient, it can be the difference between intense and prolonged monitoring, and being sent home after treatment. For Medicare, admission means an order of magnitude higher charges. And while the Gainsharing Demonstration explicitly forbade increasing

overall admission rates due to the bonus program, it is unclear whether this can be enforced in the long run. Any increases in overall admission rates would be extremely costly to both Medicare and the patients themselves.

While sorting healthier patients into participating hospitals may seem relatively benign, this behavior can severely bias policy evaluations and result in ineffective programs being taken to scale. In an early evaluation of the Gainsharing Demonstration, the Agency for Healthcare Research and Policy published an article reporting that the bonuses reduced costs per admission by eight percent (AHRQ, 2014). The apparent success of the first wave of the program led to its expansion. However, the initial evaluation only compared the costs of admitted patients at participating hospitals, before and after the program was implemented. I replicate this exercise and show that a simple pre- versus post- comparison of admitted patients is misleading, and that the apparent cost savings disappear in a more careful evaluation.

What, if anything, can the New Jersey Gainsharing Demonstration tell us about how these incentives would fare if implemented nationwide? Would shutting down the sorting margin result in a more powerful incentive to reduce costs? I examine this hypothesis in two sub-samples where doctors cannot sort patients between hospitals: patients who were admitted through the emergency room, and patients whose doctors only work in one hospital. There is no evidence in either case that shutting down the sorting channel leads to a larger cost reduction. Notably, doctors who only work in one hospital act just like their multi-hospital counterparts: manipulating the admission margin, but not reducing costs.

One critique of using a doctor-level difference-in-difference specification is that some doctors may respond to incentives in one hospital by changing their practice style at all hospitals in which they work. Through the lens of a doctor-level difference-in-difference identification strategy, this response would look like a null effect. If enough doctors respond to the bonuses by changing their practice style, they could still decrease costs on net. I use an alternative strategy based on doctor-level program exposure to measure the bonuses' effect on total hospital costs incurred. Consistent with the main estimation strategy, there is no evidence that the bonuses are associated with lower costs; if anything, costs appear to rise with program exposure.

**Related Literature** This paper contributes to three main strands of literature. First, it is directly related to the literature on how doctors respond to financial incentives. There is a large body of work studying how reimbursement levels influence procedure choice, mostly focusing on the decision to perform one particular procedure (Alexander, 2015; Clemens and Gottlieb, 2014; Coey, 2013; Dranove and Wehner, 1994; Gruber and Owings, 1996; Gruber

et al., 1999; Grant, 2009; Hadley et al., 2001, 2009; Keeler and Fok, 1996; Yip, 1998).<sup>2</sup> These papers generally find that doctors supply more services when payment increases, as well as when the payment of a competing procedure decreases. An implication of this research is that reforms which lower the profit for performing “unnecessary” procedures could be very effective at lowering costs.

Current cost-reduction proposals, however, generally involve changing the entire payment system, which could change doctor behavior on margins other than just procedure choice. To this end, a much smaller branch of the literature has studied how doctors respond to different types of payment systems—for example, fee-for-service versus capitated payments (Ho and Pakes, 2014; Dickstein, 2014).<sup>3</sup> Unfortunately, studying the effect of payment structure on doctor decision-making is hampered both by data availability, and the fact that doctors practicing under different payment schemes may differ on unobservable characteristics. Therefore, how much and on what margins doctors will respond to payment reform policies remains an open question.

Second, doctors sending healthier patients to participating hospitals is similar to evidence that Managed Care plans are able to select healthier patients into their plans (Duggan, 2004; Duggan and Hayford, 2013; Leibowitz et al., 1992; Brown et al., 2011). There is much less work, however, on the ability of doctors to identify patients with low expected costs. Doctors selecting patients according to their underlying health has been studied in the context of “report card” policies—public disclosures of the patient health outcomes of individual doctors. The evidence on report cards, however, is mixed; Dranove et al. (2003) find that the introduction of report cards cause cardiac surgeons to select healthier patients, while Kolstad (2013) finds little evidence of selection. Especially with the recent popularity of cost reduction strategies that target doctor pay, it is important to know whether doctors are able to identify low-cost patients to treat.

Third, the problems and limitations of pilot programs have been widely studied in economics, particularly in development, education, and environmental economics (Duflo, 2004; Cullen et al., 2013; Allcott, 2015). However, these lessons have generally not been applied to U.S. health care reform. The Centers for Medicare and Medicaid Services (CMS) has been running pilot programs (or “demonstrations”) since the 1960s, and the Affordable Care Act appropriated \$10 billion for the Center for Medicare and Medicaid Innovation, which tests “innovative health care payment and service delivery models”. Furthermore, the results of

---

<sup>2</sup>Most of these papers focus on C-sections, though other procedures such as coronary artery bypass grafting and breast conserving surgery have also been studied.

<sup>3</sup>A closely related literature looks at the reaction of *hospitals* to the introduction of prospective payment (Cutler, 1990, 1995; Ellis and McGuire, 1996; Dafny, 2005). These papers find that hospitals respond by changing treatment intensity and coding practices in response to DRG specific price changes.

these pilot programs help direct the annual spending of Medicare, a 600 billion dollar per year program. In this paper, I point out that even when there is evidence that such programs are effective, it may be due to gaming rather than true improvements in efficiency.

**Roadmap** The rest of the paper is organized as follows. Section 2 describes the bonus program, and the specific incentives it created for doctors. Section 3 develops a model of doctor decision-making. The model shows that the bonuses incentivize doctors to change who is admitted, and to sort patients between hospitals. The effect of the bonuses on resource use, however, is ambiguous. In the remainder of the paper, I measure the impact of the bonuses empirically. Section 4 describes my data and identification strategy, and results are presented in section 5 and 6. Section 7 provides several extensions and robustness exercises, and section 8 concludes.

## 2 Institutional Background

The exact employment relationship between doctors and the hospitals they work within is complicated, and varies from place to place. For the most, doctors treating patients in hospitals are independent contractors, rather than hospital employees. Below, I briefly describe the institutional setting in which these doctors make treatment decisions, how hospitals and doctors are paid, and what changed under the Gainsharing Demonstration.

### 2.1 How Doctors Treat Patients within Hospitals

Doctors treating patients in hospitals have three main decisions to make: where to send the patient, whether they should be admitted, and the course of treatment. Consider first the question of where to send a patient. In this decision, doctors are limited to hospitals where they have prearranged relationships which allow them to see patients—so-called admitting or surgical privileges.<sup>4</sup> Doctors often have such privileges at more than one hospital, and thus must decide where to send each patient. In the New Jersey discharge data, the average doctor is seen to treat patients at two different hospitals—this institutional feature is important for my main identification strategy, which compares the behavior of doctors working in a hospital that offers the bonuses to the same doctor working in one that does not.

When treating a patient in a hospital setting, doctors also have to decide whether to admit a patient and treat them, or treat the patient in the hospital and then discharge

---

<sup>4</sup>Even emergency room doctors are usually not employed by the hospital, but are provided by separate business.



them. The technical definition of admission is simply that a doctor has written an order to that effect. In practice, admitted patients generally stay at least overnight and occupy a bed. When considering whether to admit a patient, doctors must weigh the benefits against the costs; admitted patients are intensely monitored, and receive more care. On the other hand, admission is costly for the patient, both in terms of time and money. In addition, admitted patients spend more time in the hospital, and thus face a higher risk of contracting a hospital acquired infections, which are often resistant to treatment. Admission is a real choice; in nearly all diagnosis groups there are both patients treated with and without being admitted.

Simultaneously, the doctor decides on a course of diagnostic tests and treatment. Diagnostic tests help determine the patient’s clinical condition, and can inform the admission decision. Treatment itself can also inform the admission decision; for example, Chan (2015) cites the response to bronchodilators for suspected asthma. While the doctor legally in charge of a patient generally makes these decisions, care is also provided by other doctors, physician assistants, and nurses who share the on-the-ground responsibilities of treatment. Thus, while there is one doctor of record for each patient who determines and is responsible for treatment, many of the minute-to-minute treatment decisions are made by other practitioners.

## 2.2 How Doctors and Hospitals are Paid

For the most part, doctors in the US are paid under the fee-for-service system, and traditional Medicare is no exception. Under this payment scheme, doctors are paid separately for each service provided to the patient. Conversely, hospitals in the US are paid either a fixed amount per visit according to a broad diagnosis category, or a per diem for each day spent in the hospital (Reinhardt, 2006).<sup>5</sup> Medicare, which makes up approximately a third of the average hospital’s net revenue, pays hospitals a fixed sum based on the patient’s diagnosis (called diagnosis related groups, or “DRGs”)—no matter how expensive the patient is to treat. Thus for Medicare patients, hospitals would like doctors to treat patients as cheaply as possible.

The financial incentives of doctors and hospitals over how much care to provide are fundamentally at odds, pushing doctors to do more and hospitals to do less. While hospitals can theoretically constrain doctors’ resource use through the threat of revoking their privi-

---

<sup>5</sup>Medicaid pays hospitals either a flat amount per visit based on diagnosis, or with per diem payments (a lump sum for each day spent in the hospital). Private insurers pay hospitals based on either DRGs, per diems, or discounts negotiated off list charges. Payments from Medicare and private insurers each make up approximately third of hospital revenue (Reinhardt, 2006).

leges, in reality this is difficult. Doctors are afforded a lengthy due process to protect them from competitive forces that could override quality or patient safety. Furthermore, hospitals benefit from having doctors with privileges on staff, as these same privileges are what bring people into the hospital in the first place. Hospitals would like to use pay incentives to align the incentives of doctors with their own, but it is difficult in the current legal environment. Federal law constrains the ability of hospitals and doctors to participate in cost reduction programs, with the rationale that hospitals will pressure doctors into giving too little care, which would be bad for patient welfare.<sup>6</sup> Medicare demonstration projects, however, are typically granted waivers to these statutes.

## 2.3 The Gainsharing Demonstration

The New Jersey Gainsharing Demonstration was designed by the New Jersey Hospital Association to reduce hospital costs by aligning the incentives of doctors with those of hospitals. Many argue that paying doctors under fee-for-service system incentivizes additional care on the margin, and causes doctors to provide treatments with low or zero marginal benefits. Under the program, doctors are still paid separately for each service provided, but can also receive bonuses for lowering the total hospital costs incurred while treating admitted Medicare patients. These bonuses are paid by hospitals to doctors, and are supposed to reduce hospital costs by lowering the use of unnecessary procedures. Doctors treating admitted Medicare patients at participating hospitals are eligible to receive one bonus per visit, where the maximum bonus they can receive varies by the patient's diagnosis and severity of illness.

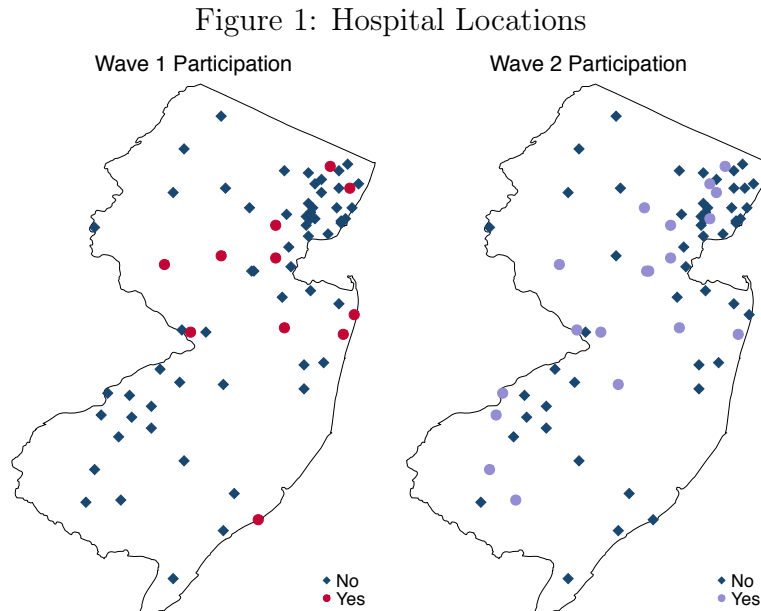
The Gainsharing Demonstration took place in two waves, which both applied only to doctors treating admitted Medicare patients.<sup>7</sup> The initial phase took place in twelve New Jersey hospitals from July 1st, 2009 to July 1st, 2012. Eight of the original twelve hospitals opted to extend the program through March 31st, 2013. Based on the reported success of the Gainsharing Demonstration, the New Jersey Hospital Association applied for and secured Federal approval for a second, larger demonstration program under the ACA's Bundled Payments for Care Improvement initiative (AHRQ, 2014). On April 1st, 2013 the program

---

<sup>6</sup>The civil money penalty (CMP) set forth in section 1128A(b)(1) of the Social Security Act prohibits any hospital or critical access hospital from knowingly making a payment directly or indirectly to a doctor as an inducement to reduce or limit services to Medicare or Medicaid beneficiaries under the doctor's care. In addition, gainsharing arrangements may also implicate the anti-kickback statute (section 1128B(b) of the Social Security Act) and the doctor self-referral prohibitions of the Act (section 1876 of the Social Security Act) (Office of Inspector General, 1999).

<sup>7</sup> During its first incarnation, it was called the New Jersey Gainsharing Demonstration. Later, it was rechristened and expanded as a part of the Bundled Payment for Care Improvement Initiative (BPCI) under the CMS Innovation Center, which was charged by the PPACA to support the development and testing of innovative health care payment and service delivery models. (For ease of exposition, I will call both waves the Gainsharing Demonstration throughout the paper, as the payment incentives were nearly identical.)

was renamed the BPCI Model 1 program, and was expanded to 23 hospitals (for simplicity, I refer to both the first and second wave as the Gainsharing Demonstration throughout the paper). Figure 1 shows that the participating hospitals in each wave are scattered around the state, and are thoroughly interspersed with non-participating hospitals.



Notes: Blue diamonds are hospitals that never participated, red circles are hospitals that took up the bonuses in the first wave, and purple circles are hospitals that joined in the second wave.

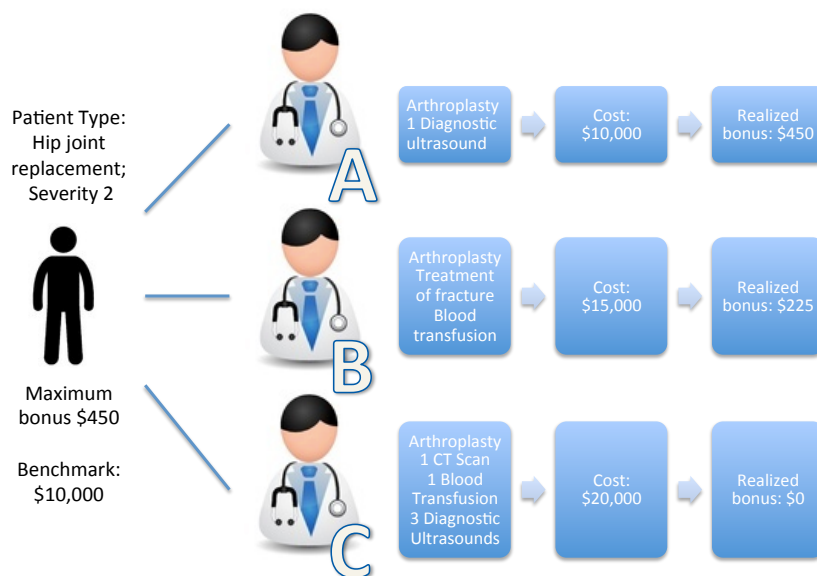
When a hospital joins the Demonstration, doctors working in the hospital have the option to sign up for the program. While I do not have data on which or how many doctors signed up, anecdotal evidence suggests take-up was high. There is no reason for an eligible doctor to abstain, as there is no change in the process or form of payment, no additional paperwork, and no risk; doctors are only rewarded for improvement, and not punished for stagnation or increasing costs. While many providers are ultimately involved in patient care, only the responsible doctor is eligible to receive a cost reduction bonus under the Gainsharing Demonstration. For medical cases, this is the attending doctor, and for surgical cases, it is the surgeon. . As doctors could only receive bonuses when treating admitted patients, language was included in the Demonstration that total admissions could not rise under the program, though it was unclear how this would be enforced.

### 2.3.1 Bonus calculation

The bonus a doctor receives through the Gainsharing Demonstration for treating an eligible (admitted and covered by Medicare) patient is calculated in three steps. First, patients are

divided into types based on their diagnosis and how sick they are, using 3M’s All Patient Refined Diagnosis Related Group (APR-DRG) system (for example, one type would be “hip joint replacement, severity of illness level two”). Second, a maximum bonus is assigned to each patient type. All doctors face the same maximum bonus for treating patients of the same type. Third, this maximum bonus is scaled according to whether and how much the doctor reduces hospital costs for their patient relative to pre-program hospital costs for their patient’s type in New Jersey. A hypothetical bonus calculation example is presented in Figure 2. In this example, three doctors treat three patients with the same type, but receive different bonuses based on the costs of the treatment they provide.

Figure 2: Hypothetical Bonus Calculation

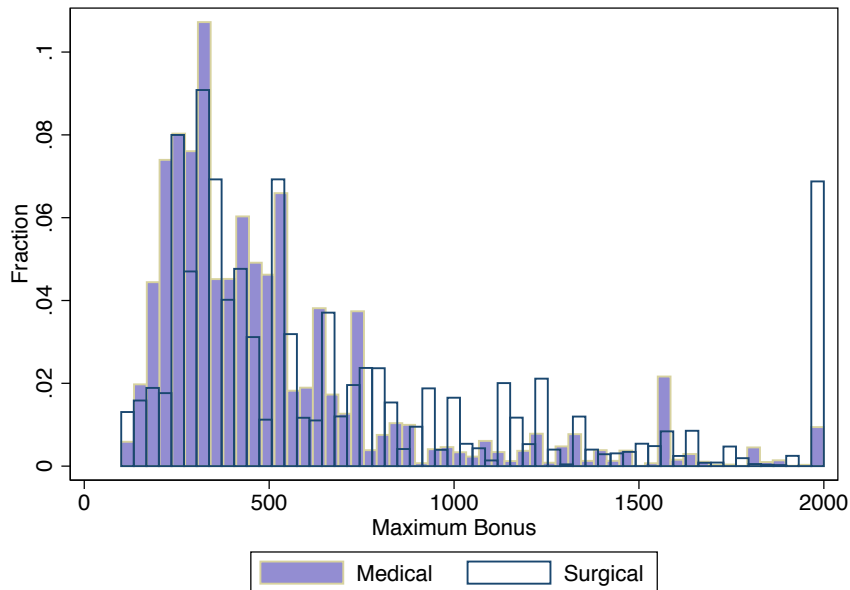


The maximum bonuses are calculated using hospital cost data from before the program started (the base year was 2007 for the original demonstration and 2011 for the expansion). The maximum bonus for treating a patient type is defined as one tenth of the average deviation from the 25th percentile of the hospital cost distribution in the state of New Jersey for that patient type in the base year. To this end, a third party calculated four maximum bonus amounts for each diagnosis (APR-DRG), depending on the severity of the patient’s illness (SOI). The four severity of illness categories capture the fact that the same diagnosis (e.g. “peptic ulcer and gastritis”) may be more or less serious depending on a patient’s age and comorbidities.<sup>8</sup> I recreate these maximum bonuses using list charges from hospital billing records deflated by Medicare’s hospital level cost-to-charge ratio (more details

<sup>8</sup>As patient types are partially determined by the types and numbers of co-morbidities recorded by the doctor, there is a potential for “up-coding”—doctors changing a patient’s diagnosis to increase expected profit. I discuss this in more detail in section 5.

on bonus calculation can be found in the appendix). An example of maximum bonuses for two particular APR-DRGs is given in Table A.1, and the distribution of maximum bonuses is shown in Figure 3.

Figure 3: Distribution of Maximum Bonuses



Notes: Each observation is a Medicare beneficiary’s inpatient visit to a general medical/surgical hospital in New Jersey from 2006-2013, excluding visits that went through the emergency room.

The maximum bonus, rather than the amount actually received, is the important number when considering the impact of the Gainsharing program on doctor behavior. The maximum bonus represents the most a doctor can hope to earn, *ex ante*, for any given patient. A reduction in costs of the same dollar amount for different patient types translates into different realized bonuses, depending on the maximum bonuses. Thus, the size of the maximum bonus reflects how valuable a patient is for participating doctors.

While the formulas for calculating the maximum bonuses are opaque, the doctors had a lot information about the sizes of the bonuses that were available under the Gainsharing Demonstration. Doctors were given quarterly “dashboards”, which gave them real time feedback on the performance and explicitly told them the amount of unearned incentive that they were leaving on the table (see Figure A.1). Therefore, it is reasonable to expect that doctors to quickly become familiar with the quirks of the program.

The rationale behind the formulas used to calculate maximum bonuses is that high cost variance within a diagnosis is a red flag, suggesting that there are high cost patients who could be getting the same treatment as low cost patients. The bonuses are designed to make reducing the treatment costs of patients in diagnoses with high cost variance especially

profitable for doctors. However, waste generated by unnecessary treatment is just one explanation for the underlying cause of cost variation. Alternatively, high cost variance within a group of patients could be due disease pathophysiology, rather than doctor behavior.

Consider again Figure 2: either the three doctors are treating essentially the same patient, or they are treating patients with underlying medical variation. In the first scenario, higher spending by doctor C represents waste. In the second, spending variation reflects underlying variation in the progression of a patient’s disease. If the latter is true, diagnoses with high cost variance may be exactly the diagnoses where it is relatively simple to find patients with much lower than average expected costs, making sorting particularly attractive.

### 2.3.2 Characteristics of Participating Hospitals

The hospitals that formed the demonstration and its expansion are similar to other New Jersey hospitals, on average. A cap of twelve participating hospitals for the original demonstration was mandated by Medicare, despite considerable interest from additional hospitals. In response, the New Jersey Hospital Association chose the first twelve participants to represent New Jersey hospitals as a whole.

Table 1: Hospital Characteristics

Participation	Wave 1			Wave 2		
	No	Yes	Diff	No	Yes	Diff
# of Hospitals (Gen. Medical/Surgical)	53	12		42	23	
Nongovernment Not-for-Profit	0.83	0.83	0.00	0.77	0.9	-0.13
Bed Size Code	5.40	6.00	-0.60	5.40	5.60	-0.20
ER Visits	51,194	54,469	-3,275	48,701	55,411	-6,710
Hospitals in a Network	0.54	0.58	-0.04	0.52	0.59	-0.07
CBSA Type: Metro (Pop. of 50,000+)	0.17	0.17	0.00	0.17	0.17	0.00
Medicare Discharges	6,393	8,449	-2,056*	6,236	7,400	-1,164
Medicare Days	39,007	50,011	-11,004	37,999	44,585	-6,586
Medicaid Discharges	2,445	1,950	495	2,365	2,341	24
Medicaid Days	12,085	8,575	3,510	11,558	11,295	263
"Grade A"	0.40	0.67	-0.27*	0.45	0.44	0.01

Notes: American Hospital Association Annual Survey (2008). Medicaid/Medicare days are the total number of hospital days used by beneficiaries. Grade A refers to a hospital report card, reported by the Leapfrog Group (<http://www.hospitalsafetyscore.org/>). \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

As can be seen in Table 1, this appears to have been successful. The main difference between participating and non-participating hospitals—especially in the first wave—is that hospitals participating in the program have more Medicare patients on average. Hospitals

with more Medicare patients have the most to gain from a program designed to reduce the costs of treating this population. Hospitals that participated in the first wave were also more likely to receive a grade of A on a hospital quality report card. In the second wave, however, more hospitals took up the program, and these differences disappeared.

Despite the fact that participating and non-participating hospitals are similar on observable characteristics, the selection of hospitals into the bonus program is clearly non-random. Larger hospitals with more Medicare patients are more likely to participate, and these hospitals may be on different cost trajectories than non-participating hospitals. All of the identifying variation used in the main analysis is within doctor, however, which sidesteps many of the difficulties posed by differential trends at the hospital level. Instead of comparing hospitals that do and do not participate, I compare doctors working under the Gainsharing program to themselves working in the status quo payment environment.

### 3 Conceptual Framework

To formalize how the bonuses should affect doctor decision-making, I present a stylized model of the incentives and choices faced by doctors working in a hospital setting. I consider a doctor who works in two hospitals, and must decide whether a patient is admitted, where to send the patient, and how much care to provide. First, I describe the outcome when neither hospital offers a cost reduction bonus. Next, I introduce the cost reduction bonuses to one of the hospitals in the model. Finally, I compare how the doctor’s decisions change as a result of the introduction of the bonuses.

#### 3.1 The Set Up

The model consists of one doctor treating a population of patients with mass one, where all patients are within a single diagnosis-severity of illness type. I assume that the type is exogenously defined, though I will examine the validity of this assumption empirically. For each patient, the doctor must make three decisions: whether a patient is admitted,  $A \in \{0, 1\}$ , which hospital they attend,  $H \in \{0, 1\}$ , and how much care is provided,  $q \in \mathbb{R}^+$ . When neither hospital offers a bonus, the two hospitals are identical. Patients vary only by their sickness level  $\beta$ , which is uniformly distributed from zero to  $\bar{\beta}$ .

Doctors are utility maximizers, and choose  $H$ ,  $A$ , and  $q$  to maximize a weighted average of their profit from treating the patient and the patient’s utility from treatment, the weight placed on profit is  $\lambda$ . Doctors are paid a reimbursement rate,  $a$ , for each unit of care,  $q$ , provided to the patient. The payment,  $a$ , does not depend on the hospital choice or

whether the patient is admitted. Thus, the doctor's profit from treating a patient is  $aq$ .

In addition to profit, doctors also care about their patient's utility from treatment. A doctor's concern for their patient's welfare can be understood as altruism on behalf of their patients, or as the doctor acting to preserve their reputation. The utility a patient derives from medical care is:

$$\begin{cases} \beta q - \frac{b}{2}q^2 & \text{if } A = 0 \\ \beta q - \frac{b}{2}q^2 + \gamma q - C & \text{if } A = 1 \end{cases} \quad (1)$$

The patient's utility from medical treatment is concave in  $q$ , with sicker patients (those with a higher  $\beta$ ) benefiting more from medical care. The key assumption is that patients have a bliss point in  $q$ . Care provided past this preferred  $q$  need not necessarily become physically harmful, but can be interpreted as patients facing co-insurance and the opportunity cost of their time.

A patient's utility from treatment depends additionally on whether or not they are admitted. If a patient is admitted to the hospital, there are two opposing effects. On one hand, being admitted makes treatment more beneficial (represented in the model by  $\gamma$ ). There are many benefits to being admitted; admitted patients receive more care, and are intensely monitored. On the other hand, the care received by admitted patients is very expensive, and requires a much longer stay in the hospital. The additional care is costly in monetary terms, in terms of a patients' time, and because it translates into a greater probability of contracting a hospital acquired infection. Thus, patients also face a fixed cost of admission,  $C$ ; patients dislike being admitted to the hospital, all else equal. When making the decision to admit a patient, a doctor trades off the costs and benefits for their patient, as well as the difference in their compensation.

When doctors are indifferent between hospitals, I assume they randomly assign patients such that they have an equal probability of going to each hospital.<sup>9,10</sup>

### 3.2 No Bonuses

The two hospitals are identical in the case with no bonuses, and thus the hospital choice drops out—doctors behave the same in each hospital. Doctors are utility maximizers, and choose  $q$  and  $A$  to maximize a weighted average of their profit from treating the patient and the patient's utility from treatment:

---

<sup>9</sup>The randomization can be interpreted as patients having a slight preference for the closest hospital, and patients being evenly distributed across space.

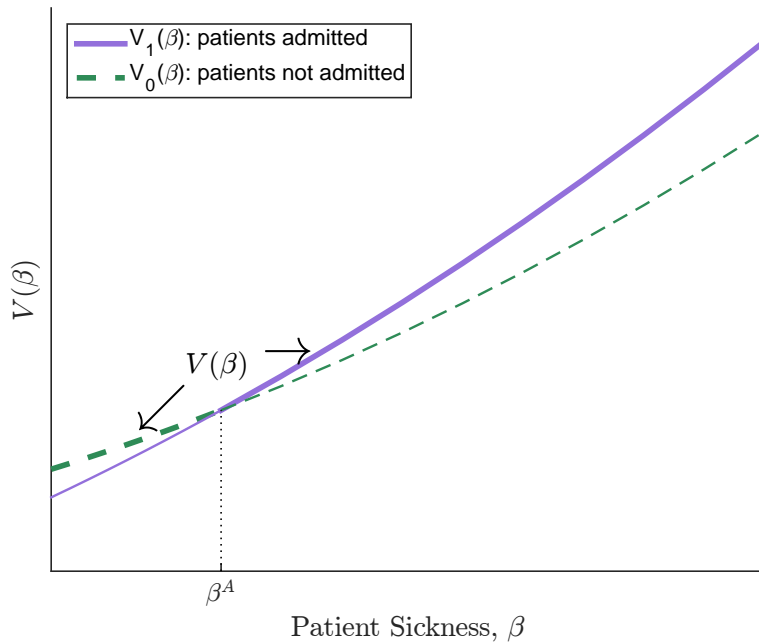
<sup>10</sup>Doctors could assign patients such that any proportion goes to each hospital; I use 50-50 to keep examples simple.



$$\begin{aligned}
\max_{q,A} U(q, A; \beta) &= \underbrace{\lambda [aq]}_{\text{profit}} + (1 - \lambda) \underbrace{\left[ \beta q + (\gamma q - C) * \mathbb{1}\{A = 1\} - \frac{b}{2} q^2 \right]}_{\text{patient's utility from treatment}} \\
&= \max \left\{ \underbrace{\lambda [aq^*(\beta)] + (1 - \lambda) \left[ (\beta + \gamma) q^*(\beta) - C - \frac{b}{2} q^*(\beta)^2 \right]}_{\mathbf{V}_1(\beta) = U(q^*(\beta); \beta, A=1)}, \right. \\
&\quad \left. \underbrace{\lambda [aq^*(\beta)] + (1 - \lambda) \left[ \beta q^*(\beta) - \frac{b}{2} q^*(\beta)^2 \right]}_{\mathbf{V}_0(\beta) = U(q^*(\beta); \beta, A=0)} \right\}
\end{aligned}$$

The intuition is fairly straightforward. Doctors would like to provide as much care  $q$  as possible to maximize their profits, but are constrained by patient preferences. Relatively healthy patients (low  $\beta$ ) dislike admission, while for sicker patients (high  $\beta$ ), admission is beneficial. Since doctors take into account patient's preferences, there is a sickness threshold  $\beta^A$  which defines the optimal admission rule.

Figure 4: Doctor's Utility as a Function of  $\beta$ : without Bonus  
Without Bonus



Notes: The bold line sections show the optimal quantity of care provided as a function of  $\beta$ , without bonuses.

**Proposition 1:** Under some parameter conditions, there exists a  $\beta^A$  such that all patients with  $\beta < \beta^A$  are not admitted, and all patients with  $\beta \geq \beta^A$  are admitted.

The optimal decision rule for admission is depicted in Figure 4, which plots the value function of a doctor under two scenarios: all patients being admitted ( $V_1(\beta)$ ), and no patients being admitted ( $V_0(\beta)$ ). Doctors always admit patients when the  $V_1(\beta) \geq V_0(\beta)$ , and never admit patients when  $V_0(\beta) > V_1(\beta)$ .  $\beta^A$  is defined as the sickness level where  $V_0(\beta) = V_1(\beta)$ . Thus, the value function  $V(\beta)$  is the upper envelope of  $V_0(\beta)$  and  $V_1(\beta)$ , where the sickest patients are admitted and the healthiest patients are not admitted. As doctors randomize when they are indifferent between hospitals,  $\frac{\bar{\beta} - \beta^A}{2}$  patients are admitted at each hospital. A formal proof is presented in the Mathematical Appendix.

### 3.3 With Bonuses

Next, I consider what happens when cost reduction bonuses of the form used in the Gain-sharing Demonstration are introduced at hospital 1. Adding the bonuses only changes the framework described above in one way—doctors’ profits change at the bonus hospital:

$$\begin{cases} aq + \max\{\alpha_0 - \alpha_1 q, 0\} & \text{if } H = 1 \text{ and } A = 1 \\ aq & \text{else} \end{cases} \quad (2)$$

If an admitted patient is treated at the bonus hospital, the doctor is now eligible to receive a cost reduction bonus:  $\max\{\alpha_0 - \alpha_1 q, 0\}$ . The bonus is decreasing in the amount of care provided,  $q$ , but is never negative. The maximum bonus for the diagnosis-severity of illness group is  $\alpha_0$ , and  $\alpha_1$  represents how quickly the bonus decays as  $q$  increases. Everything else remains the same, including the number of patients admitted to the bonus hospital,  $\beta' = \frac{\bar{\beta} - \beta^A}{2}$ .<sup>11</sup> Doctors are constrained by the number of patients admitted at the participating in the absence of the bonus program, as the program included language restricting doctors from increasing overall admission. Even if the rules had not mentioned admission levels, holding admission fixed is equivalent to introducing capacity constraints—assuming hospital capacity does not change in response to the program. Doctors can, however, change which patients are admitted and where they are treated. Past research has shown that patients typically accept their doctors’ recommendations (Manning et al., 1987). Since all patients affected by the program are covered by Medicare, and all hospitals accept Medicare, it seems reasonable to assume most patients would agree to use whichever hospital is recommended by their doctor.

---

<sup>11</sup>The capacity constraint  $\beta'$  is just a number; doctors can admit any patients they want, and are not constrained to pick patients in an interval of  $\beta$ .

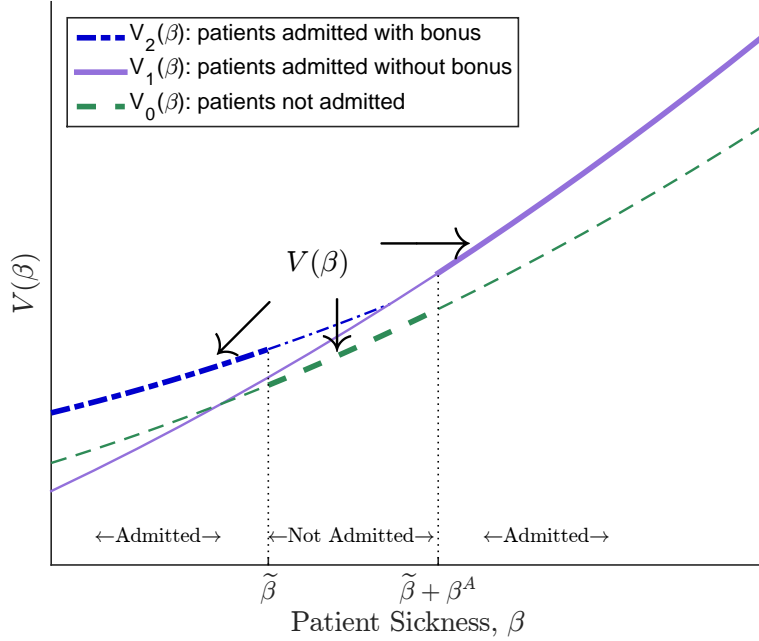
Doctors now choose  $A \in \{0, 1\}$ ,  $H \in \{0, 1\}$ , and  $q$  to maximize the utility function

$$\begin{aligned}
\max_{q, H, A} U(q, H, A; \beta) &= \underbrace{\lambda [aq + \max \{\alpha_0 - \alpha_1 q, 0\} * \mathbb{1} \{H = 1, A = 1\}]}_{\text{profit}} \\
&\quad + \underbrace{(1 - \lambda) \left[ \beta q + (\gamma q - C) * \mathbb{1} \{A = 1\} - \frac{b}{2} q^2 \right]}_{\text{patient's utility from treatment}} \\
&= \max \left\{ \underbrace{\lambda [aq^*(\beta) + \alpha_0 - \alpha_1 q^*(\beta)] + (1 - \lambda) \left[ (\beta + \gamma) q^*(\beta) - C - \frac{b}{2} q^*(\beta)^2 \right]}_{\mathbf{V}_2(\beta) = U(q^*(\beta); \beta, H=1, A=1)}, \right. \\
&\quad \underbrace{\lambda [aq^*(\beta)] + (1 - \lambda) \left[ (\beta + \gamma) q^*(\beta) - C - \frac{b}{2} q^*(\beta)^2 \right]}_{\mathbf{V}_1(\beta) = U(q^*(\beta); \beta, H=0, A=1)}, \\
&\quad \left. \underbrace{\lambda [aq^*(\beta)] + (1 - \lambda) \left[ \beta q^*(\beta) - \frac{b}{2} q^*(\beta)^2 \right]}_{\mathbf{V}_0(\beta) = U(q^*(\beta); \beta, A=0)} \right\}
\end{aligned}$$

subject to the capacity constraint that only  $\beta' = \frac{\bar{\beta} - \beta^A}{2}$  patients can be admitted at each hospital. The expression is the same as in the case without the bonus, with the addition of  $V_2(\beta)$ : the value function if doctors receive the cost reduction bonus.

Whether or not there are bonuses, the admitted patients are always those with the largest (positive) difference between the utility a doctor receives from admitting them and not admitting them. Before the bonuses are introduced, this difference is largest for the sickest patient ( $\beta = \bar{\beta}$ ), and is decreasing in  $\beta$ . The introduction of the bonuses at hospital 1, however, eliminates this monotonicity. The cost reduction bonuses increase the doctor's profit from admitting healthy (low  $\beta$ ) patients, up until the point where a patient is sick enough that quantity of care chosen is too high to generate a bonus (represented by the blue dash-dotted line in Figure 5). After the introduction of the bonus, the patients whose admission generates the biggest utility gain are at the extremes: the lowest  $\beta$  patients because of the bonus, and the highest  $\beta$  patients because these patients have the highest utility from treatment.

Figure 5: Doctor's Utility as a Function of  $\beta$ : with Bonus



Notes: The bold line sections show the optimal quantity of care provided as a function of  $\beta$ , with bonuses.

**Proposition 2:** Under some parameter restrictions, there exists a  $\tilde{\beta}$  such that patients with  $\beta \in [0, \tilde{\beta}]$  are admitted at the bonus hospital, patients with  $\beta \in [\tilde{\beta}, \tilde{\beta} + \beta^A]$  are not admitted, and the remaining patients with  $\beta \in [\tilde{\beta} + \beta^A, \bar{\beta}]$  are admitted at either the bonus or non-bonus hospital.

After the bonuses are introduced, doctors would like to admit all patients (see Figure 5; the upper envelope contains segments of  $V_2(\beta)$  and  $V_1(\beta)$ , but not  $V_0(\beta)$ ). Not all patients can be admitted, however, as doctors are limited by the original hospital capacity—only  $\beta' = \frac{\bar{\beta} - \beta^A}{2}$  patients can be admitted at each hospital. The introduction of the bonuses has no impact on the treatment of the sickest patients—doctors will continue to admit them. For the healthiest patients, however, the bonus is large enough that doctors will now admit them, despite the fact that these patients dislike admission. Doctors will admit low  $\beta$  patients at the bonus hospital up until  $\tilde{\beta}$ . They will also admit the sickest  $\bar{\beta} - (\tilde{\beta} + \beta^A)$  patients, randomizing over hospital choice such that they admit  $\beta'$  total patients at each hospital. The patients with  $\beta$ s in the middle of the distribution will not be admitted. This optimal decision rule is shown in Figure 5. The exact form of  $\tilde{\beta}$ , as well as the conditions necessary for an interior solution, are detailed in the Mathematical Appendix.

The cost reduction bonuses introduce two distortions. First, the bonuses increase the probability of admission for the healthiest patients and decrease the probability of admission

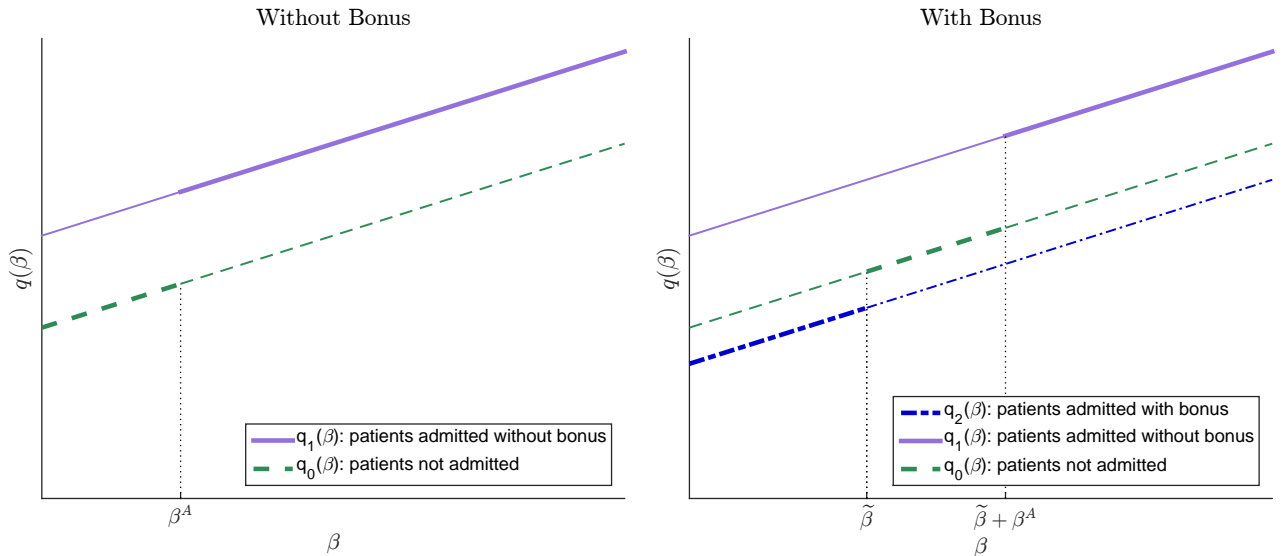
for sicker patients. Many patients with  $\beta < \tilde{\beta}$  are not admitted without the bonus (the “pre-period”), and all are admitted in the when the bonus is introduced (the “post-period”). On the other hand, many “medium sick” patients with  $\beta \in [\tilde{\beta}, \tilde{\beta} + \beta^A]$  are admitted in the pre-period, and are not admitted in the post-period. Second, the bonuses cause sorting. After their introduction, doctors send the healthiest patients exclusively to the bonus hospital. Previously, the non-bonus hospital would have received some of the healthier patients, whereas now they only get patients with  $\beta > \tilde{\beta} + \beta^A$ .

The bonuses’ affect on the quantity of care provided to bonus generating patients, however, is not clear. If a patient is admitted both with and without the bonuses, then  $q$  clearly decreases. If a patient is only admitted under the bonus program, on the other hand, then the change in  $q$  is ambiguous. Intuitively, there are two conflicting forces. The first is downward pressure on  $q$  from the bonus (represented by  $\alpha_1$ ). The second is upward pressure on  $q$  from admission (represented by  $\gamma$ ).

**Proposition 3:** The direction of the change in  $q$  conditional on  $\beta$  from the pre- to the post-period for bonus-generating patients ( $\beta \in [0, \tilde{\beta}]$ ) is ambiguous.

Whether the quantity of care provided for the bonus generating patients is higher or lower than the counterfactual of neither hospital offering a bonus is determined by the relative size of  $\gamma$  and  $\alpha_1$ . For more details, see the Mathematical Appendix.

Figure 6: Optimal Quantity of Care as a Function of  $\beta$



Notes: The bold line sections show the optimal quantity of care provided as a function of  $\beta$ , both with and without bonuses.

Finally, the model predicts the results of the naive evaluation. After the bonuses are introduced, the average  $q$  for admitted patients falls at the participating hospital. The

average  $q$  falls because the composition of patients at the participating hospital has changed, not because costs have decreased conditional on patient health ( $\beta$ ). A simple comparison of average costs with and without the bonuses, however, would find that costs went down at the participating hospital (see Figure 6).

## 4 Data and Empirical Strategy

According to the conceptual framework outlined above, the introduction of the bonus program will cause doctors to change their decisions over admission—both in terms of whether and where patients are admitted. The bonuses may also impact the quantity of services provided, though the direction and magnitude are ambiguous. The relative sizes of these three effects, and whether the program ultimately decreases costs, are empirical questions which I address in the remainder of the paper.

### 4.1 Data Sources

The primary data are the New Jersey Uniform Billing Records, which cover all hospital discharges in New Jersey from 2006 to 2013. Each record in the confidential file includes the patient’s name and the medical license number of the attending doctor and surgeon (if the case was surgical). From this raw data, I create a panel by matching patient records across visits by sex, date of birth, and first and last names.<sup>12</sup> I also create doctor identifiers using the recorded license numbers of doctors and surgeons. The final file includes codes for patients and doctors, allowing me to track them over time and across all hospitals in New Jersey. The ability to follow both patients and doctors is often lacking in medical records, and is an important strength of this paper. The discharge data also include admission and discharge dates, all diagnoses and procedure codes, payer information, patient demographic information, and list charges. To these data, I add information on hospitals from the American Hospital Association (AHA) annual survey, and Medicare’s cost-to-charge ratio series.

I supplement the billing records with the bonus amounts that doctors could have received for treating each patient during the program. While I do not have access to the actual maximum bonuses used, I recreate them based on the formula provided by the New Jersey Hospital Association. The first step is to define the patient types, which is done by passing the billing records through 3M’s All Patient Refined - Diagnosis Related Group (APR-DRG) software. For each record, the software creates a diagnosis group, a severity of illness (SOI)

---

<sup>12</sup>The Levenshtein edit distance is used to match names, because of problems with typos and misspellings (stata command strgroup).

category, and designates the visit as medical or surgical. I then combine this information with cost data from the billing records and the bonus formula used in the Gainsharing Demonstration to reconstruct the maximum bonuses that a doctor could earn by treating each patient.

My outcome variables are all constructed from the billing records. To measure how the bonuses affect admission, I create a counterfactual measure for whether a patient would have been admitted in the absence of the program. To measure sorting, I construct measures of each patient’s latent health in the year running up to their visit. Resource use and costs measures are taken directly from the billing records—I look at the effect of the bonuses on costs, length of stay, and the use of diagnostic procedures. Each of these outcome variables are described in more detail in section 5.

## 4.2 Estimation Sample

The main sample consists of visits where an admitted patient was covered by Medicare, as these are the cases which can generate bonuses for doctors. I restrict the sample to patients seen in general medical and surgical hospitals that were open throughout the sample period. This restriction mainly excludes psychiatric and rehabilitation facilities, which were not targeted by the program. Visits to doctors with very few admitted patients were also dropped, as these doctors likely did not have enough patients to qualify for the bonus program. Most importantly, my main analysis omits cases where patients are admitted through the emergency room. I exclude these patients in order to highlight sorting, as doctors have no hospital choice decision when deciding whether to admit patients from the emergency room. However, these patients provide a useful placebo check, and I examine them in a separate analysis at the end.

The main sample consists of approximately 400,000 each of medical and surgical visits, which were conducted by 3,515 doctors in 73 hospitals (see Table 2). The patients are predominately white, with an average age of 74, and a high disease burden. The doctors worked at 2.1 hospitals on average, with a third working in both a participating and non-participating hospital. The average maximum bonus a doctor could earn for treating a surgical patient was \$697, and for a medical patient was \$513. While few doctors receive the whole maximum bonus, even taking home half of these amounts would be a windfall (for comparison, in 2012 Medicare paid doctors \$675.99 to repair a knee ligament (Smith, ed, 2012)).<sup>13</sup>

---

<sup>13</sup>Medicare facility charge for repair of knee ligament (CPT 27405), 2012.

Table 2: Main Sample Characteristics

Patients		Doctors		Outcomes	
N	815,014	N	3,515	Charlson index	2.26
N (medical)	389,604	Avg. # of patients	232	Total chronic	3.52
N (surgical)	425,410	Avg. # of hospitals	2.1	Viral Infections	0.01
Avg. age	74	Med. # of hospitals	2.0	Kidney disease	0.10
% white	81	Ever in policy hosp	68%	Asthma	0.03
% black	11	Ever in other hosp	65%	Length of stay	6.74
% woman	54	Ever in both types	33%	CT scans	0.03
% in policy hosp	17	Avg. max bonus (med.)	\$513	MRIs	0.01
APR-DRGs (med.)	162	Avg. max bonus (surg.)	\$697	Diag. imaging	0.15
APR-DRGs (surg.)	117				

Notes: admitted Medicare patients in general medical/surgical hospitals (2006-2013), excluding those admitted through the emergency room. APR-DRG stands for All Patient Refined-Diagnosis Related Group. % in policy hosp is fraction treated at participating hospitals while the program was in effect.

### 4.3 Empirical Strategy

The main challenge for identifying the effect of the cost-reduction bonuses on doctor decision-making is that participating hospitals are different—and likely on different trajectories with respect to costs—from hospitals that did not take up the program. If costs in participating hospitals are trending differently from costs in non-participating hospitals, comparing the change in outcomes before and after the program was introduced at participating hospitals to non-participating hospitals (a hospital-level difference-in-difference estimator) would be inappropriate. I address this concern by looking within doctor. In this case, the identifying variation comes from choices made by the same doctor working at multiple hospitals. The identifying assumption is now that in the absence of the program, a doctor’s behavior would have been on the same trend across all hospitals in which she works.

The regressions take the form of a difference-in-difference specification with doctor fixed effects:

$$Outcome_{idht} = \beta_0 + \beta_1 Policy_{ht} + \beta_2 X_{it} + \lambda_t + \lambda_h + \lambda_d + \epsilon_{idht} \quad (3)$$

where  $i$  stands for individual,  $d$  for doctor,  $h$  for hospital, and  $t$  for time (here the.  $Policy_{dht}$  is an indicator for whether the visit occurred in a participating hospital when the bonus program was in effect, and the coefficient of interest is  $\beta_1$ . The patient characteristics included in  $X_{it}$  vary slightly by specification, but in general contain age, race, and sex, and dummies



for patient type (APR-DRG and severity of illness pairs). Hospital, quarter-by-year, and doctor fixed effects are also included in all regressions ( $\lambda_h$ ,  $\lambda_t$  and  $\lambda_d$ ).

When I look at the effect of the program on which patients are admitted to the hospital, I further want to know whether patients are differentially affected, depending on the size of the maximum bonus attached to their type. In this case, I interact the policy variable with the bonus size:

$$\begin{aligned} Outcome_{idht} = & \beta_0 + \beta_1 Policy_{dht} + \beta_2 HighBonus_{idht} + \beta_3 Policy * HighBonus_{idht} + \\ & \beta_4 X_{idht} + \lambda_t + \lambda_h + \lambda_d + \epsilon_{idht} \end{aligned} \tag{4}$$

where  $HighBonus_{idht}$  is defined as a maximum bonus at or above the median amount across patients. In all specifications, standard errors are clustered at the hospital level.

The intuition behind this identification strategy is to take doctors who work in both a hospital that eventually takes up the bonuses, and one that does not. If the doctor's behavior in the bonus hospital changes in comparison to the same doctor at the other hospital, when the bonus program goes into effect, it is attributed to the bonuses. One weakness of this identification strategy, however, is that it will not be able to detect any responses to the bonuses that occur in both hospitals. For example, if a doctor responds to the program by changing her practice style in all hospitals in which she works, this change will not be attributed to the bonuses. For now, I rule out this type of behavior. In section 6, I address this issue with a second identification strategy, which examines total resource use as a function of program exposure.

## 5 Results

All results are presented separately for medical and surgical patients, as there are important differences between these groups. For one, surgical patients have higher admission rates (with a few APR-DRGs at nearly full admission), so there is less room to manipulate the admission margin in response to bonuses. Resource use is also higher on average for surgical cases, which is important when considering the impact of the bonuses on length of stay and diagnostics. In addition, the consequences for the patient of changing admission and the quantity of services may be different for medical and surgical cases, which could lead to distinct program effects across the two groups.

## 5.1 The Admission Margin

### 5.1.1 Outcomes

To measure the effect of the bonuses on admission decisions, I would ideally know whether each admitted patient who was treated under the bonus program would have been admitted if the bonus program did not exist. While I cannot know this counterfactual, I estimate it from my data in two steps. First, I take data from before the program started, which includes both Medicare patients that were and were not admitted. Using this pre-program data, I regress admission on a large set of observable characteristics. The model does a good job of predicting admission for Medicare patients, with a pseudo- $R^2$  of 0.49.<sup>14</sup>

Next, I use the results of this regression to predict whether or not each patient in my main sample (all of whom are actually admitted) would have been admitted in the pre-period. I call this variable the baseline admission probability, as it answers the question: would a patient with these characteristics have been admitted in the pre-period? If the introduction of the bonus program is associated with a decrease in the baseline admission probability of admitted patients, it is consistent with the bonuses inducing doctors to admit some patients who would not have been otherwise.

### 5.1.2 Results

Table 3 shows the effect of the bonuses on the average baseline admission probability. On average, the introduction of the bonuses had no effect, which is consistent both with program rules and binding capacity constraints. While the bonuses are associated with a small decrease in the average baseline admission probability of both medical and surgical patients, it is not statistically significant (columns 1 and 3 of Table 3, for medical and surgical patients respectively). This null result, however, conceals important heterogeneity with respect to the size of the bonus. While doctors were barred from admitting all of their patients in response to the bonuses, they could change the composition of admitted patients across patient types (APR-DRG and severity of illness pairs).

For patients in high bonus types, the introduction of the bonuses is associated with a significant decrease in the baseline admission probability—implying that some patients in high bonus types would not have been admitted in the absence of the program. For these patients, the effect of the bonuses is the sum of the coefficients on the policy and its interaction with high bonus. The policy is associated with a statistically significant decrease of 0.043 percentage points for medical patients (Table 3, column 2) and 0.016 for surgical

---

<sup>14</sup>Details of this procedure are in the appendix.

patients (Table 3, column 4), implying that doctors are more likely to admit patients in high bonus types when the policy is in effect (5% and 2% of the mean, respectively).

Table 3: Bonuses Change Which Patients are Admitted

	Medical Patients		Surgical Patients	
	(1)	(2)	(3)	(4)
	Baseline	Baseline	Baseline	Baseline
	Admission	Admission	Admission	Admission
Policy	-0.002 (0.005)	0.026*** (0.005)	-0.008 (0.005)	0.011*** (0.004)
High bonus		-0.045** (0.019)		-0.001 (0.013)
Policy * high bonus		-0.069*** (0.005)		-0.027*** (0.005)
Mean dep. var.	0.906	0.906	0.952	0.952
Clusters	73	73	73	73
F-test: $\beta_1 + \beta_3 = 0$		0.000		0.012
Observations	385,845	385,845	405,400	405,400

Notes: Quarter-by-year, doctor, hospital, and diagnosis by severity of illness fixed effects also included. Standard errors clustered at the hospital level. "F-test:  $\beta_1 + \beta_3 = 0$ " reports the p-value from this F-test, where  $\beta_1$  is the coefficient on the policy, and  $\beta_3$  is the coefficient on the policy variable interacted with an indicator for a high bonus type. \*p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01

Conversely, the bonuses are associated with a small increase in the baseline admission probability of low bonus patients, suggesting that some low bonus patients were not admitted, that would have been admitted in the absence of the program. The effect of the policy on the baseline admission of low bonus patients is simply the coefficient on the policy variable in columns 2 and 4 of Table 3, which is statistically significant and positive in both cases. Low bonus medical patients saw an 0.026 percentage point increase (3% of mean) in the baseline admission rates, and low bonus surgical patients saw an increase of 0.011 percentage points (1% of mean). This increase in the baseline admission probability for low bonus patients is consistent with doctors making room for more high bonus patients by not admitting some patients with low bonus types.

The effect of the bonuses on admission is much larger for medical patients, compared to surgical patients. The differential responsiveness is likely due to a combination of two effects. First, admission rates are lower on average for medical patients, leaving more room for discretion. Some surgical APR-DRGs have 100% admission rates, and thus admission cannot increase. Second, surgical procedures tend to be more uniform than treating medical cases, and have stronger protocols and norms about whether admission is necessary. Thus, it

is likely that doctors treating medical patients have more discretion over admission decisions, on average.

## 5.2 The Hospital Sorting Margin

### 5.2.1 Outcomes

In the conceptual framework presented in section 3, doctors responded to the Gainsharing Demonstration in part by sending the healthiest (lowest  $\beta$ ) patients within a patient type to participating hospitals, as these patients required fewer resources and therefore generated the biggest bonuses. Do doctors sort healthier patients into participating hospitals, conditional on patient type? To answer this question, I need measures of the latent health of patients that are known to the doctor (or at least correlated with information known to the doctor), but not used in the bonus formula. To create such measures, I exploit the time-series dimension of the data, and use information from past patient visits.

My preferred measure of latent patient health is the Charlson Co-morbidity Index, which is computed based on 17 conditions, each weighted by the associated risk of death.<sup>15</sup> This index has been widely validated, and has shown to be strongly predictive of hospital resource utilization (Charlson et al., 2008). In order to measure a patient’s latent health (rather than the acute event that brought them to the hospital), I construct a “leave-out” version of the Charlson Co-morbidity Index. The leave-out index is constructed using data on each patient’s previous visits to the hospital, excluding the current visit.

While the Charlson Co-morbidity Index is a useful summary measure of latent patient health, it is by definition incomplete. The index only captures a handful of conditions, all of which are very serious. A patient with a high disease burden, but where each individual condition is less serious, may not score highly on the index but still be expensive to treat. Thus, I also look separately at whether patients were treated in the hospital for other medical conditions over the past year: asthma, viral infections, and chronic kidney disease, as well as the number of visits for chronic conditions. Finally, I also look at the total costs generated by hospital visits over the past year, as patients with better latent health should be cheaper to treat at all points in time.

---

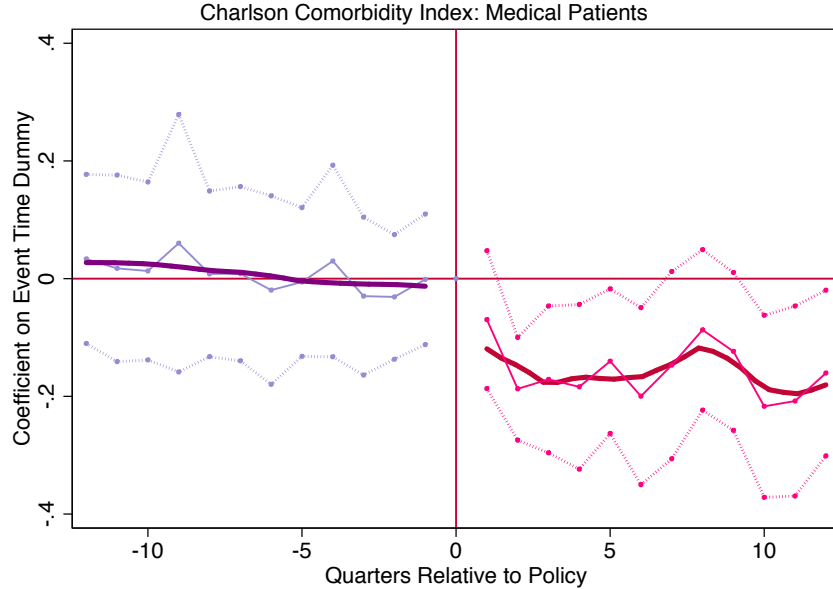
<sup>15</sup>The Charlson Co-morbidity Index is a weighted sum over the following 17 conditions, where weights are in parentheses: acute myocardial infarction (1), congestive heart failure (1), peripheral vascular disease (1), cerebrovascular disease (1), dementia (1), chronic pulmonary disease (1), rheumatologic disease (connective tissue disease) (1), peptic ulcer disease (1), mild liver disease (1), diabetes without complications (1), diabetes with chronic complications (1), hemiplegia or paraplegia (2), renal disease (2), cancer (2), moderate or severe liver disease (3), metastatic carcinoma (6), AIDS/HIV (6)

### 5.2.2 Results

Taking the Charlson Co-morbidity Index as my preferred measure of patient health, I find that doctors admit healthier patients to bonus eligible hospitals in response to the program. Figure 7 displays the effect of the bonus policy on the average Charlson Co-morbidity Index of patients in event time (for medical patients), where the implementation of the policy is normalized to  $t = 1$ . The event time specification is identical to equation 3, except the binary policy variable is replaced with quarterly event time dummies denoting the number of quarters before and after a hospital took up the policy.<sup>16</sup>

The coefficients on the event time dummies are plotted in Figure 7, which shows that medical patients admitted at participating hospitals under the Gainsharing Demonstration quickly became healthier based on their past medical history, conditional on type. In the quarters before the bonuses were introduced, doctors did not systematically send healthier patients to hospitals that eventually would take up the policy. After the policy comes into effect, however, there is a clear drop in the average co-morbidity burden of patients, as doctors sort their healthier patients into participating hospitals.

Figure 7: Healthier Patients Sent to Participating Hospitals



Notes: Each observation is a Medicare beneficiary's inpatient visit to a general medical/surgical hospital in New Jersey from 2006-2013, excluding visits that went through the emergency room.

The sorting result depicted in Figure 7 is presented in regression form in column 1 of Table 4. The bonuses are associated with a decrease in the average Charlson Co-morbidity Index of

<sup>16</sup>The model is fully saturated; hospitals which never participated are assigned an event time of -12.

medical patients of 0.092. To put the magnitude of this change in perspective, decreasing the index by a tenth is associated with decreases in in-hospital mortality of 3-6 percent across seven OECD countries (Quan et al., 2011).<sup>17</sup> The Charlson Co-morbidity Index has been shown to be strongly predictive of resource utilization (Charlson et al., 2008), suggesting that these healthier patients are indeed cheaper to treat. The Charlson Index of surgical patients does not respond to the program, probably because the diagnoses included in the index are more closely tied to medical conditions than surgical ones (Table A.2 lists the top 10 diagnosis groups for medical and surgical patients in the sample).

Table 4: Doctors Sort Healthier Patients into Participating Hospitals

Panel A: Medical Patients						
	(1)	(2)	(3)	(4)	(5)	(6)
	Charlson	Tot. Costs	Tot. Chronic	Asthma	Viral Inf.	Chron. Kidney
Policy	-0.092** (0.042)	-519.725 (540.878)	-0.061 (0.046)	-0.003* (0.002)	-0.002* (0.001)	-0.006** (0.003)
Mean dep. var.	2.785	17,917.969	4.168	0.040	0.019	0.115
Clusters	73	73	73	73	73	73
Observations	389,604	389,604	389,604	389,604	389,604	389,604
Panel B: Surgical Patients						
	(1)	(2)	(3)	(4)	(5)	(6)
	Charlson	Tot. Costs	Tot. Chronic	Asthma	Viral Inf.	Chron. Kidney
Policy	0.003 (0.030)	86.710 (212.931)	0.015 (0.056)	-0.003* (0.001)	-0.001 (0.001)	-0.000 (0.002)
Mean dep. var.	1.775	9,047.469	2.918	0.027	0.009	0.085
Clusters	73	73	73	73	73	73
Observations	425,410	425,410	425,410	425,410	425,410	425,410

Notes: Quarter-by-year, doctor, hospital, and diagnosis by severity of illness fixed effects also included, as well as dummies for age categories, sex, and race. All diagnoses from index visit are excluded. Tot. Chronic refers to the number of body systems affected by chronic conditions. Standard errors clustered at the hospital level. \*p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01

In addition, the number of chronic conditions recorded, as well as the probability patients were seen for asthma, viral infections, or chronic kidney disease in the past year all decreased for medical patients admitted in participating hospitals (Columns 3-5 of Table 4). These patients also have accumulated fewer hospital costs over the past year (Column 2 of Table 4), though the estimate is not statistically significant. The effects for analogous surgical

<sup>17</sup>Calculation assumes mortality decreases linearly between a Charlson score of 3 and 2.

patients are much noisier, though they are less likely to have been seen for asthma over the past year.

At least for medical patients, table 4 shows that conditional on type, patients admitted by doctors at participating hospitals are healthier than patients admitted by the same doctors at non-participating hospitals—exactly as suggested by the model. The mechanism appears to be straightforward: conditional on type, healthier patients are cheaper to treat, and cheaper patients earn higher bonuses. By sorting patients across hospitals, doctors can earn a bonus without changing treatment conditional on admission and patient health.

Table 5: Doctors Sort Healthier Patients into Participating Hospitals.: High Admission APR-DRGs

Panel A: Medical Patients						
	(1)	(2)	(3)	(4)	(5)	(6)
	Charlson	Tot. Costs	Tot. Chronic	Asthma	Viral Inf.	Chron. Kidney
Policy	-0.112* (0.060)	-133.886 (768.247)	-0.024 (0.061)	-0.002 (0.002)	-0.001 (0.002)	-0.008* (0.004)
Mean dep. var.	2.785	17,917.969	4.168	0.040	0.019	0.115
Clusters	73	73	73	73	73	73
Observations	142,493	142,493	142,493	142,493	142,493	142,493
Panel B: Surgical Patients						
	(1)	(2)	(3)	(4)	(5)	(6)
	Charlson	Tot. Costs	Tot. Chronic	Asthma	Viral Inf.	Chron. Kidney
Policy	0.000 (0.026)	168.296 (191.144)	0.003 (0.053)	-0.003* (0.002)	-0.002** (0.001)	-0.001 (0.002)
Mean dep. var.	1.775	9,047.469	2.918	0.027	0.009	0.085
Clusters	73	73	73	73	73	73
Observations	158,770	158,770	158,770	158,770	158,770	158,770

Notes: Quarter-by-year, doctor, hospital, and diagnosis by severity of illness fixed effects also included, as well as dummies for age categories, sex, and race. CCI stands for Charlson Co-morbidity Index, which is calculated based on information in previous visits. Tot. Chronic refers to the number of body systems affected by chronic conditions. Standard errors clustered at the hospital level. \*p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01

One might wonder whether the results in Table 4 might be driven by changes in the admission margin—patients that are only admitted because of the program are likely healthier on average—rather than pure sorting. Table 5 addresses this concern by repeating the analysis on the subsample of patients in diagnoses that are nearly always admitted (the top third of the sample of medical and surgical patients, based on the average admission rate in

each APR-DRG), as it is unlikely that the composition of this sample would be affected by changing admission thresholds. The results are less precisely estimated, the point estimates are strikingly similar. If anything, there is stronger evidence of sorting for surgical patients in Table 5, despite the fact that these APR-DRGs have over 99 percent admission rates. Thus, it appears that doctors are responding to the bonuses by both changing admission thresholds and sorting patients across hospitals.

### 5.2.3 Up-Coding

Both the model and the sorting results assume that the diagnosis and severity of illness margins are unaffected by the Gainsharing Demonstration. One might be worried about this assumption, as during the 1990s many hospitals were accused of up-coding—exaggerating a patient’s diagnosis to extract a higher reimbursement from Medicare. Silverman and Skinner (2004) found, for example, that between 1989 and 1996, the percentage point share of the most generous diagnosis groups (DRGs) for pneumonia and respiratory infections rose precipitously.

The diagnosis groups used by Medicare are particularly susceptible to up-coding, as there are often multiple DRGs for each diagnosis, where the most severe version pays a much higher amount. For example, there are separate Medicare diagnosis groups (MS-DRGs) for diabetes with complications (638), diabetes with major complications (637), and diabetes without complications (639), where the more severe codes are reimbursed at higher rates. In the diagnosis groups used for the bonus calculations, however, this feature is lacking. In order to upcode at the diagnosis level doctors would have to change the diagnosis conceptually, which seems unlikely (e.g., changing a diagnosis from “diabetes” (APR-DRG 420) to “malnutrition, failure to thrive, and other nutritional disorders” (APR-DRG 421)).

It is also possible that doctors could respond to the Gainsharing Demonstration by trying to move their patients into higher severity of illness bins.<sup>18</sup> Influencing the severity of illness (SOI) designation should be difficult, as it is imputed by software and not recorded by the doctor. The only way doctors can affect the severity of illness is to change which secondary diagnoses are recorded on a patient’s chart. While the link between any one co-morbidity and the designation generated by the software is not clear, adding additional diagnoses to all patients could lead to higher average SOI designations. If doctors managed to inflate the severity of illness of all patients in response to the program, the average “true sickness level” of the patients in each cell would decrease—the sickest patients in the first severity bin would be shifted into the next bin, and so on up the chain. Up-coding, therefore, could

---

<sup>18</sup>Though while a “with complications” designation always leads to a higher payout in the Medicare DRG system, a higher SOI level does not necessarily lead to a higher bonus.



generate similar patterns in the data as sorting.

Table 6: Effect of Program on Population-Level Severity of Illness

	SOI: Medical Patients		SOI: Surgical Patients	
	(1)	(2)	(3)	(4)
	Full Sample	High Admission	Full Sample	High Admission
Policy	-0.026 (0.021)	-0.021 (0.029)	0.011 (0.014)	0.011 (0.020)
Mean dep. var.	2.363	2.614	1.996	2.065
Clusters	73	73	73	73
Observations	389,604	142,493	425,410	158,770

Notes: Quarter-by-year, doctor, hospital, and APR-DRG fixed effects also included, as well as dummies for age categories, sex, and race. Standard errors clustered at the hospital level.  
\*p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01

There is no association in the data between the bonuses and the average severity of illness within APR-DRG cells, however, suggesting that up-coding is not a concern in this context. The regressions reported in Table 6 use the same empirical strategy outlined in equation 3, but with severity of illness as the dependent variable and APR-DRG rather than APR-DRG by severity of illness fixed effects. The introduction of the bonus program appears to have no effect the average severity of illness. Not only are the point estimates insignificant, they are very small, and not even consistently positive.

One interpretation of this null result is that doctors did not have sufficient information about how the software translated co-morbidities into severity of illness levels to up-code. Another is that the proximity of the payer (the hospital) to the recipient (the doctor) in the Gainsharing Demonstration differs substantially from earlier settings where up-coding has been found. Even if doctors are able to influence the severity codes, it may be much harder to up-code patients when working within the walls of the entity making the payment, in comparison to a distant third party such as Medicare. Either way, it does not appear that changes in the composition of APR-DRG cells as a result of up-coding are likely to be driving the results in Tables 4 and 5.

## 5.3 The Quantity of Services Margin

### 5.3.1 Outcomes

Finally, I ask whether the bonuses reduced average hospital costs, conditional on latent patient health. To examine whether the bonuses changed procedure use or costs, I again

use a variety of measures. The first two are summary measures of resource use. Length of stay is defined as the number of nights spent in the hospital, and is often used to proxy for the intensity of care provided during the visit. The hospital total costs incurred during a visit are estimated using the total list charges reported in the discharge data, deflating them by Medicare’s hospital-year level cost-to-charge ratio, and then converting them to real 2010 dollars. Deflating the list charges is an important step, as list charges are closer to bargaining tools than a measure of the costs to the hospital of providing a service. The Medicare cost-to-charge ratio is explicitly designed to translate list charges into an estimate of the resource cost of inpatient care.

In addition to summary measures of resource use, I look specifically at the use of diagnostic imaging to proxy for the use of unnecessary procedures. While it is difficult to pinpoint any specific test as unnecessary, there is widespread agreement that diagnostic imaging is overused (Hillman and Goldsmith, 2010; Abaluck et al., 2015).<sup>19</sup> If the bonuses are associated with a reduction in use of expensive diagnostic imaging procedures such as magnetic resonance imaging (MRI) and computed tomography (also called CT or CAT scans), it would be consistent with the bonuses lowering the use of unnecessary procedures. The bonuses could also cause doctors to substitute expensive tests for cheaper tests; in particular, I look at whether the bonuses increase the use of diagnostic ultrasounds, which are cheap and radiation-free imaging tests.<sup>20</sup>

### 5.3.2 Results

Table 7 shows that despite the fact that the program was explicitly designed to reduce costs, the bonuses have no effect on costs or resource use, conditional on admission and patient health. The program is not associated with significant decreases in length of stay, diagnostic imaging tests, or costs. Even taking the point estimates at face value, the magnitudes are small, and the signs are not consistently negative. In addition, there is no evidence of substitution between high-tech (MRIs and CT scans) and low-tech (diagnostic ultrasounds) imaging, consistent with the disappointing results of the Medicare Imaging Demonstration, which tried to reduce inappropriate use of high-tech imaging through decision support software (Timbie et al., 2014).

---

<sup>19</sup>For example, over half of the procedures labeled by doctors as unnecessary in the *Choosing Wisely* campaign (<http://www.choosingwisely.org/>) are directly related to diagnostic imaging (Rao and Levin, 2012)

<sup>20</sup>Unnecessary diagnostic imaging not only contributes to high health care costs—it may also harm patients. False positives can lead to additional treatments with much higher health risks. With CT scans there is also a risk that patients will react to the contrast material, which is rare but serious (Lessler et al., 2010). In addition, radiation exposure may increase later cancer risk (Smith-Bindman, 2010).

Table 7: Bonuses Do Not Reduce Costs or Change Procedure Use

## Panel A: Medical Patients

	(1)	(2)	(3)	(4)	(5)	(6)
	Length of Stay	CT Scan	MRI	Diag. Ultra	Any Imaging	Total Costs
Policy	-0.025 (0.173)	0.000 (0.003)	-0.001 (0.002)	-0.001 (0.003)	-0.002 (0.006)	485 (438)
Mean dep. var.	6.979	0.032	0.016	0.036	0.112	11,459
Clusters	73	73	73	73	73	58
Observations	389,604	389,604	389,604	389,604	389,604	355,306

## Panel B: Surgical Patients

	(1)	(2)	(3)	(4)	(5)	(6)
	Length of Stay	CT Scan	MRI	Diag. Ultra	Any Imaging	Total Costs
Policy	-0.056 (0.115)	-0.003 (0.003)	-0.000 (0.001)	0.001 (0.006)	0.007 (0.010)	922 (558)
Mean dep. var.	6.516	0.025	0.008	0.061	0.186	18,557
Clusters	73	73	73	73	73	58
Observations	425,410	425,410	425,410	425,410	425,410	371,262

Notes: Quarter-by-year, doctor, hospital, and diagnosis by severity of illness fixed effects also included, as well as dummies for age categories, sex, and race, and the variables measuring underlying health from Table 4. Standard errors clustered at the hospital level. The sample is smaller when looking at costs, as the cost-to-charge ratio is not available for all hospitals. \*p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01

The regressions reported in Table 7 include controls for all latent health measures examined in the previous section, it is likely that I cannot completely control for differences in underlying health status. Given the fact that healthier patients were sorted into participating hospitals sorting, these patients may have required fewer resources from the start. Thus, the small decreases reported in some measures in Table 7 should be considered an upper bound on the true effect.

Given these null results, how did the initial evaluation conclude that the program succeeded in decreasing costs? In the first column of Table 8, I replicate the initial evaluation of the first wave of the program. As in the initial study, I only include hospitals that eventually take up the program, with no controls for latent health or doctor fixed effects. Here, the policy appears to decrease costs, and this decrease is statistically significant. In column 2, however, I show that clustering standard errors at the hospital level already renders the decrease in costs insignificant. In columns 3-5, I add health controls, comparison hospitals, and doctor fixed effects, and show that the sign flips from negative to positive. It is possible

to conclude that the bonus program lowered costs, but this conclusion does not hold in to a more thorough investigation.

Table 8: Replicating the Initial Evaluation

Panel A: Medical Patients					
	(1)	(2)	(3)	(4)	(5)
	Total Costs	Total Costs	Total Costs	Total Costs	Total Costs
Policy	-431** (218)	-431 (864)	-350 (858)	448 (638)	611 (614)
Health controls	-	-	x	x	x
Comparison hospitals	-	-	-	x	x
Doctor F.E.	-	-	-	-	x
Mean	10910.3	10910.3	10910.3	11280.9	11280.9
Clusters	.	11	11	58	58
<i>N</i>	85374	85374	85374	293052	293052

Panel B: Surgical Patients					
	(1)	(2)	(3)	(4)	(5)
	Total Costs	Total Costs	Total Costs	Total Costs	Total Costs
Policy	-608** (294)	-608 (936)	-559 (916)	1465 (893)	1349* (782)
Health controls	-	-	x	x	x
Comparison hospitals	-	-	-	x	x
Doctor F.E.	-	-	-	-	x
Mean	17289.1	17289.1	17289.1	18343.6	18343.6
Clusters	.	11	11	58	58
<i>N</i>	81816	81816	81816	309289	309289

Notes: Quarter-by-year, hospital, and diagnosis by severity of illness fixed effects also included, as well as dummies for age categories, sex, and race. Health controls are the variables measuring underlying health from Table 4. Comparison hospitals are those where the Gainsharing policy was not implemented. \*p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01

## 6 Alternative Strategy: Doctor-Level Program Exposure

The previous section suggests that many doctors respond to the bonuses by changing their behavior in participating hospitals, relative to non-participating hospitals: they manipulate admission and sort patients to maximize their bonuses, but do not reduce costs. However, if some doctors respond to the bonuses by reducing costs at both hospitals, the within-doctor strategy will not pick this up. In addition, policymakers may want to know what effect the bonuses had on total costs and procedure use.

In order to isolate the effect of the bonuses on total costs and procedure use, I use an alternative identification strategy based on doctor-level program exposure. The exposure variable is defined by the fraction of a doctor’s Medicare patients that would have been affected by the program if the distribution of patients across hospitals was fixed in the pre-period (2006-2008). Program exposure is zero in the pre-period, when no doctors are working under the bonus scheme, and then rises to the pre-program fraction of a doctor’s caseload treated at participating hospitals. For a doctor who only works in participating hospitals, the exposure variable is zero before the program and one when the program goes into effect. For a doctor whose caseload in the pre-period is split evenly between two hospitals, one of which participates, the exposure variable goes from zero to one-half. Unlike the within-doctor identification strategy, the exposure variable captures the fact that some doctors only admit patients to participating hospitals, others are not exposed at all, and many doctors are in between. And by construction, the exposure measure reflects only ex ante exposure, and will not be affected by doctors sorting patients in response to the bonus program.

To analyze the effect of doctor-level program exposure on total costs and procedure use, I collapse data on all Medicare patients (including both those that were admitted and those that were not) to the doctor-quarter level, and regress exposure on the same cost and quantity measures as in section 5.3:

$$outcome_{dt} = \beta_0 + \beta_1 exposure_{dt} + \lambda_d + \lambda_t + \epsilon_{dt} \quad (5)$$

where  $\lambda_d$  and  $\lambda_t$  are doctor and quarter fixed effects. The coefficient of interest is  $\beta_1$ , which I interpret as the effect of program exposure on total costs and procedure use, net of sorting.

There is no evidence of any cost-saving response to the Gainsharing Demonstration in response to program exposure—if anything, exposure is associated with higher costs. Columns 1 and 2 of Table 9 support the (at least short-run) effectiveness of program rules prohibiting increases in the number of admitted patients—doctors with more exposure do not increase overall admissions. However, the exposure to the Gainsharing Demonstration is associated with higher costs (column 3 shows the effect on the total costs incurred over a quarter, and column 4 the average costs incurred per patient). In addition, there is no evidence that program exposure decreases the number of imaging tests performed. Instead, exposure is associated with a marginally significant increase in the number of CT scans performed (column 5), and an insignificant increase in the overall number of diagnostic imaging tests (column 8). Finally, patients of doctors with more program exposure do not spend fewer days in the hospital (column 9).

Table 9: Simulated Share Treated on Costs and Procedure Use

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Admitted	Adm.	Total	Avg.	CT		Diag.	Diag.	Hosp.
	Patients	Rate	Costs	Costs	Scans	MRIs	Ultra.	Imaging	Days
Program exposure	0.065 (0.354)	-0.000 (0.005)	1,612. (3,981)	301*** (114)	0.096* (0.058)	-0.012 (0.022)	-0.037 (0.037)	0.068 (0.104)	-0.769 (2.175)
Mean dep. var.	18.232	0.617	221,501	8,963	0.999	0.380	0.882	3.257	113.519
Clusters	3,348	3,348	3,222	3,222	3,348	3,348	3,348	3,348	3,348
Observations	98,599	98,599	90,002	90,002	98,599	98,599	98,599	98,599	98,599

Notes: Sample includes all Medicare patients seen by doctors in the main analysis. Doctor and quarter-by-year fixed effects included. Total admissions, total costs, CT scans, MRIs, diagnostic ultrasounds, and diagnostic imaging are totals at the doctor-quarter level. Hospital days refers to the total number of days per quarter patients stayed in the hospital. Admission rate and average costs are averaged across patients at the doctor-quarter level. \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$

The results of the doctor-level exposure analysis concord with the results of the within-doctor identification strategy, and appear to rule out systematic changes in practice style in response to the bonuses. One interpretation of Table 9 is that some doctors respond to the program by performing additional tests in order to justify admission for patients who otherwise would be sent home. While the estimates are too noisy to pinpoint the source of the cost increase, there is no evidence from either identification strategy that the Gainsharing Demonstration resulted in lower costs.

## 7 Placebo Tests and Robustness

### 7.1 Placebo tests

To confirm that the main results in section 5 are not spurious, I conduct two placebo tests. First, I randomly assign New Jersey hospitals to participate in the program, holding constant both the number of participating hospitals and the timing. I repeat the main within-doctor regressions using randomly assigned participation, and plot the CDFs of the resulting coefficients (based on 100 repetitions) in Figure A.6. Second, I hold fixed the true hospital participation, but randomly assign start dates for the program, and repeat the main within-doctor regressions. The CDFs of the coefficients from this second simulation (again based on 100 repetitions) are presented in Figure A.6.

The coefficients from the true regressions are represented by a red vertical line, and the 90th percentile by a red horizontal line. In nearly all cases, the true coefficients are well

above the 90th percentile—in a few cases, the true coefficients are larger than any coefficient generated under the simulation. The results of the simulations suggest that it is extremely unlikely that the findings in section 5 are due to chance.

## **7.2 Other patient groups**

### **7.2.1 Emergency Room Medicare Patients**

While patients admitted through the emergency room were excluded from the main analysis, they provide both a useful placebo test for sorting. Patients admitted through the emergency room (ER) cannot be sorted in response to the bonuses, as emergency room doctors cannot send a healthier than average ER patient to a different hospital. In Table A.5, I show that the sorting results pass this placebo test—the introduction of the bonuses is not associated with a change in underlying patient health for patients admitted through the emergency room.

The incentives for manipulating admission are also weaker for Medicare patients admitted through the emergency room. While it varies from hospital to hospital, an attending in the emergency department generally decides to admit a patient, and then a different doctor is responsible for the patient after admission. The way the Gainsharing Program is designed, a bonus from treating an ER patient would most likely go to the second doctor, and thus the bonuses should not influence admission. Consistent with this intuition, there is also no effect of the bonus program on baseline admission for patients admitted through the ER (see Table A.4).

Exactly because there is no sorting margin and a limited admission margin, the bonuses could have a large effects on costs and procedure use for patients admitted through the emergency room. In Table A.6, however, I show that the effects of the bonuses on costs and quantity measures for these patients are similar to those in the main sample. While the point estimates for diagnostic ultrasounds and MRIs and CT scans hint at possible substitution towards cheaper diagnostic tests, they are not precisely measured, and there is no evidence of lower costs.

### **7.2.2 Patients Not Covered by Medicare**

Despite the fact that only Medicare patients are included in the bonus program, it is possible that the program could spill over into the treatment of other patients—particularly “near Medicare” patients. “Near Medicare” patients, aged 50 to 64, have many of the same health

problems as the Medicare population, but are too young to qualify for Medicare coverage.<sup>21</sup> The bonus program could spill over into the care of near Medicare by crowding out care for near Medicare patients, as their treatment cannot generate a bonus.

I find no evidence that the program caused doctors to crowd out younger patients (see Tables A.7 through A.9). If anything, the results for near Medicare patients point in the same direction as the main results, though are roughly 80% smaller in magnitude. If anything, these results are consistent with doctors not being able to perfectly target Medicare patients when responding to the bonuses, rather than doctors crowding out the care of near Medicare patients.

### 7.2.3 Patients of Single Hospital Doctors

Finally, I consider separately the sub-sample hospital visits treated by doctors who only practice at one hospital. These “single hospital doctors” provide an interesting glimpse into what we might expect if the program was scaled up. If all hospitals participated in the program, the sorting margin would be shut down, but doctors could still respond on both admissions and costs. One might expect that the incentives to lower costs would be more powerful if the sorting margin was shut down.

Using single hospital doctors in a hospital-level difference in difference specification, I find admission manipulation based on bonus size, but not that the bonuses reduce costs (see Tables A.10 through A.12).<sup>22</sup> And in yet another placebo check on the sorting results, there is no evidence that the patients of doctors who work in one hospital become healthier after the bonuses are introduced. Thus, there is no evidence that the bonuses are more effective at reducing costs when the sorting margin is shut down—perhaps because they are able to more easily increase their bonuses through manipulating which patients are admitted.

## 8 Conclusion

In this paper, I show that a pilot program that paid doctors bonuses for reducing costs was unsuccessful; doctors changed which patients were admitted and sorted healthier patients into participating hospitals, but did not reduce costs. The results of this program provide two main takeaways for policy. First, doctors are able to identify high and low cost patients

---

<sup>21</sup>Medicare patients who are younger than 65 are excluded from this analysis. Patients younger than 65 can be enrolled in Medicare if they are receiving Social Security Disability Insurance, or have Lou Gehrig’s or kidney failure.

<sup>22</sup>The single hospital doctor analysis uses a hospital-level difference in difference strategy, which is the same as the main specifications, but without doctor fixed effects.



within narrowly defined bins, and use this information to their advantage. Adjusting payments for patient severity is a ubiquitous feature of health care compensation systems, and policy makers should be wary of doctors manipulating these margins. Second, and more broadly, pilot programs may not be well suited for comparing the effectiveness of different health policy reforms. Not only did the sorting and selection behavior of providers undermine the pilot program—incomplete evaluations called the program a success and led to its expansion.

The idea that general equilibrium effects may differ substantially from partial equilibrium effects is an old idea, but has not been taken to heart in US health care policy—Medicare spends billions of dollars on pilot programs. Yet for these programs to be informative, they must contain the same incentives as if the program was expanded nationwide. This is difficult to accomplish with pilot programs, which generally only cover small parts of tightly connected health care markets. If doctors or other providers are able to sort patients in and out of pilot programs in response to changing incentives, they cannot tell us which reforms will be more effective when taken to scale.

In 2015, Department of Health and Human Services (HHS) committed to moving 50 percent of Medicare payments (and 90 percent of fee-for-service Medicare payments) into “value-based” payment models by 2018. Many alternative payment models will be used to meet the 2018 goal—such as “Accountable Care Organizations (ACOs), advanced primary care medical home models, new models of bundling payments for episodes of care, and integrated care demonstrations for beneficiaries that are Medicare-Medicaid enrollees” (HHS, 2015). The early evidence on the effectiveness of these alternative payment models comes largely from pilot programs and other partial equilibrium settings. The results of the Gain-sharing Demonstration call into question both the way in which pilot programs are evaluated, and the ability of pilot programs to provide this evidence.

## References

- Abaluck, Jason, Leila Agha, Christopher Kabrhel, Ali Raja, and Arjun Venkatesh**, “Negative Tests and the Efficiency of Medical Care: What Determines Heterogeneity in Imaging Behavior?” Technical Report, NBER Working Paper No. 19956 2015.
- Agency for Healthcare Research and Quality**, “Hospital Gain-Sharing Program Offers Incentives to Physicians Based on Their Efficiency, Producing Significant Cost Savings Without Decline in Quality,” Technical Report, AHRQ Innovations Exchange 2014.
- Alexander, Diane**, “Does Physician Pay Affect Procedure Choice and Patient Health? Evidence from Medicaid C-section Use,” Technical Report, Princeton 2015.

- Allcott, Hunt**, “Site Selection Bias in Program Evaluation,” *The Quarterly Journal of Economics*, 2015, *130*(3), 1117–1165.
- Brown, Jason, Mark Duggan, Ilyana Kuziemko, and William Woolston**, “Will Risk-Adjustment Decrease Health Care Costs? New Evidence from the Medicare Advantage Programs,” Technical Report 2011.
- Chan, David**, “The Efficiency of Slacking Off: Evidence from the Emergency Room,” Technical Report, NBER WP 21002 2015.
- Charlson, Mary, Robert Charlson, Janey Peterson, Spyridon Marinopoulos, William Briggs, and James Hollenberg**, “The Charlson Comorbidity Index is Adapted to Predict Costs of Chronic Disease in Primary Care Patients,” *Journal of Clinical Epidemiology*, 2008, *61*(12), 1234–1240.
- Clemens, Jeffrey and Joshua Gottlieb**, “Do Physicians’ Financial Incentives Affect Medical Treatment and Patient Health?,” *The American Economic Review*, 2014, *104*(4), 1320–1349.
- Coey, Dominic**, “Physician Incentives and Treatment Choices in Heart Attack Management,” Technical Report, SIEPR Discussion Paper No. 12-027 2013.
- Cullen, Julie Berry, Steven Levitt, Erin Robertson, and Sally Sadoff**, “What Can Be Done to Improve Struggling High Schools?,” *The Journal of Economic Perspectives*, 2013, *27*(2), 133–152.
- Cutler, David**, “Empirical Evidence on Hospital Delivery under Prospective Payments,” Technical Report, MIT 1990.
- , “The Incidence of Adverse Medical Outcomes under Prospective Payment,” *Econometrica*, 1995, *63*(1), 29–50.
- Dafny, Leemore**, “How Do Hospitals Respond to Price Changes?,” *American Economic Review*, 2005, *95*(5), 1525–1547.
- Department of Health and Human Services**, “Better Care. Smarter Spending. Healthier People: Paying Providers for Value, Not Volume,” Technical Report 2015.
- Dickstein, Michael**, “Physician vs. Patient Incentives in Prescription Drug Choice,” Technical Report 2014.
- Dranove, David and Paul Wehner**, “Physician-Induced Demand for Childbirths,” *The Journal of Health Economics*, 1994, *13*(1), 61–73.
- , **Daniel Kessler, Mark McClellan, and Mark Satterthwaite**, “Is More Information Better? The Effects of “Report Cards” on Health Care Providers,” *Journal of Political Economy*, 2003, *111*(3), 555–588.
- Duflo, Esther**, “Scaling up and Evaluation,” Technical Report, Annual World Bank Conference on Development Economics 2004.

- Duggan, Mark**, “Does Contracting Out Increase the Efficiency of Government Programs? Evidence from Medicaid HMOs,” *Journal of Public Economics*, 2004, *88*(12), 2549–2572.
- **and Tamara Hayford**, “Has the Shift to Managed Care Reduced Medicaid Expenditures? Evidence from State and Local-Level Mandates,” *Journal of Policy Analysis and Management*, 2013, *32*(3), 505–535.
- Ellis, Randall and Thomas McGuire**, “Hospital Response to Prospective Payment: Moral Hazard, Selection, and Practice-Style Effects,” *The Journal of Health Economics*, 1996, *15*, 257–277.
- Grant, Darren**, “Physician Financial Incentives and Cesarean Delivery: New Conclusions from the Healthcare Cost and Utilization Project,” *Journal of Health Economics*, 2009, *28*(1), 244–250.
- Gruber, Jonathan**, “Covering the Uninsured in the United States,” *Journal of Economic Literature*, 2008, *46*(3), 51–68.
- , “The Cost Implications of Health Care Reform,” *The New England Journal of Medicine*, 2010, *362*, 2050–2051.
- **and Maria Owings**, “Physician Financial Incentives and Cesarean Section Delivery,” *RAND Journal of Economics*, 1996, *27*(1), 99–123.
- , **John Kim, and Dina Mayzlin**, “Physician Fees and Procedure Intensity: the Case of Cesarean Delivery,” *Journal of Health Economics*, 1999, *18*(4), 473–490.
- Hadley, Jack, James Reschovsky, Catherine Corey, and Stephen Zuckerman**, “Medicare Fees and the Volume of Physicians’ Services,” *Inquiry*, 2009, *46*(4), 372–390.
- , **Jean Mitchell, and Jeanne Mandelblatt**, “Medicare Fees and Small Area Variations in Breast-Conserving Surgery among Elderly Women,” *Medical Care Research and Review*, 2001, *58*(3), 334–360.
- Hillman, Bruce and Jeff Goldsmith**, “The Uncritical Use of High-Tech Medical Imaging,” *The New England Journal of Medicine*, 2010, *363*, 4–6.
- Ho, Kate and Ariel Pakes**, “Hospital Choices, Hospital Prices and Financial Incentives to Physicians,” Technical Report, NBER Working Paper No. 19333 2014.
- Keeler, Emmett and Thomas Fok**, “Equalizing Physician Fees Had Little Effect on Cesarean Rates,” *Medical Care Research and Review*, 1996, *53*, 465–471.
- Kolstad, Jonathan**, “Information and Quality When Motivation Is Intrinsic: Evidence from Surgeon Report Cards,” *American Economic Review*, 2013, *103*(7), 2875–2910.
- Leibowitz, Arleen, Joan Buchanan, and Joyce Mann**, “A Randomized Trial to Evaluate the Effectiveness of a Medicaid HMO,” *Journal of Health Economics*, 1992, *11*(3), 235–257.

- Lessler, Adam, Joshua Isserman, Rajan Agarwal, Harold Palevsky, and Jesse Pines**, “Testing Low-Risk Patients for Suspected Pulmonary Embolism: A Decision Analysis,” *Annals of Emergency Medicine*, 2010, *55*(4), 316–326.
- Manning, Willard, Joseph Newhouse, Naihua Duan, Emmett Keeler, and Arleen Leibowitz**, “Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment,” *American Economic Review*, 1987, *77*(3), 251–277.
- Office of Inspector General**, “Special Advisory Bulletin: Gainsharing Arrangements and CMPs for Hospital Payments to Physicians to Reduce or Limit Services to Beneficiaries,” Technical Report 1999.
- Orszag, Peter and Ezekiel Emanuel**, “Health Care Reform and Cost Control,” *New England Journal of Medicine*, 2010, *363*(7), 601–603.
- Quan, Hude, Bing Li, and Chantal Couris**, “Updating and Validating the Charlson Comorbidity Index and Score for Risk Adjustment in Hospital Discharge Abstracts Using Data from 6 Countries,” *American Journal of Epidemiology*, 2011, *173*(6), 676–682.
- Rao, Vijay and David Levin**, “The Overuse of Diagnostic Imaging and the Choosing Wisely Initiative,” *Annals of Internal Medicine*, 2012, *157* (8), 574–576.
- Reinhardt, Uwe**, “The Pricing of U.S. Hospital Services: Chaos Behind a Veil of Secrecy,” *Health Affairs*, 2006, *25*, 57–69.
- Silverman, Elaine and Jonathan Skinner**, “Medicare Upcoding and Hospital Ownership,” *Journal of Health Economics*, 2004, *23*, 369–389.
- Smith-Bindman, Rebecca**, “Is Computed Tomography Safe?,” *New England Journal of Medicine*, 2010, *363* (1), 1–4. PMID: 20573919.
- Smith, Sherry, ed.**, *Medicare RBRVS 2012: The Physician Guide*, American Medical Association, 2012.
- Timbie, Justin, Peter Hussey, Lane Burgette, Neil Wenger, Afshin Rastegar, Ian Brantley, Dmitry Khodyakov, Kristin Leuschner, Beverly Weidmer, and Katherine Kahn**, “Medicare Imaging Demonstration Final Evaluation: Report to Congress,” Technical Report, RAND Corporation 2014.
- Yip, Winnie**, “Physician Responses to Medical Fee Reductions: Changes in the Volume and Intensity of Supply of Coronary Artery Bypass Graft (CABG) Surgeries in the Medicare and Private Sectors,” *Journal of Health Economics*, 1998, *17*, 675–699.

# For Online Publication

## 9 Appendix

### 9.1 Bonus Calculation Details:

#### Maximum Bonus:

The maximum bonus is calculated using cost data from 2007, before the program started. Within each diagnosis and severity of illness level pair, the maximum bonus is ten percent of the average deviation of costs from the 25<sup>th</sup> percentile of costs:

$$0.1 * \left( \frac{1}{n} \sum_{i=1}^n (c_i - c_{25^{th}pctile}) \right) \quad (6)$$

where  $c_i$  is the cost of care for a patient in 2007 (before the program), and  $c_{25^{th}pctile}$  is the 25<sup>th</sup> percentile of the cost distribution for the particular diagnosis/severity pair in 2007. The maximum bonus then is constrained to be between \$100 and \$2000.

I calculate these maximum bonuses using all inpatients over 55 at general medical and surgical hospitals in 2007. From the hospital discharge records, I know the total list charges for each visit, as well as the APR-DRG and SOI. I deflate the list charges using the hospital level Medicare cost-to-charge ratios, and use the above formula. The resulting maximum bonuses should be very similar to those used in the Gainsharing Demonstration, as these same records and cost ratios to calculate their bonuses, unless different information was by the hospital.

#### Realized Bonus:

The realized bonus is composed of two parts: a performance incentive and an improvement incentive. The performance incentive depends on how much a doctor reduced costs of a particular patient relative to the pool of patients of that type before the program started. The improvement incentive depends on how much a doctor reduced costs of a particular patient relative to their own costs for that type of patient before the program started. For the first year of the program, the weight was 1/3 for the performance incentive and 2/3 for the improvement incentive. Some hospitals changed these weights to favor the performance incentive over the course of the program.

**Realized Bonus Formula for Surgical Patients** The rate year cost is the cost of the index visit, while the 25th, 75th, and 90th percentiles refer to those percentiles of the cost distribution of all patients of a particular type in 2007. The base year cost refers to the costs of the doctor’s own patients of the particular type in 2007.

$$\underbrace{\frac{1}{3} * MaxBonus * \frac{90th\ pctile - rate\ yr\ cost}{90th\ pctile - 25th\ pctile}}_{Performance\ Incentive} + \underbrace{\frac{2}{3} * MaxBonus * \frac{base\ yr\ cost - rate\ yr\ cost}{75th\ pctile - 25th\ pctile}}_{Improvement\ Incentive} \quad (7)$$

**Realized Bonus Formula for Medical Patients** The performance incentive is the same, but the improvement incentive is calculated using length of stay rather than costs.

$$\underbrace{\frac{1}{3} * MaxBonus * \frac{90th\ pctile - rate\ yr\ cost}{90th\ pctile - 25th\ pctile}}_{Performance\ Incentive} + \underbrace{\frac{2}{3} * \frac{MaxBonus}{Best\ practice\ LOS} * (base\ yr\ LOS - rate\ yr\ LOS)}_{Improvement\ Incentive} \quad (8)$$

Due to the fact that the maximum bonus is what matters ex ante, when treatment decisions are being made, I focus on the maximum bonus throughout the paper. In principle I could calculate realized bonuses as well, but I do not for two reasons. First, due to typos and problems with string matching, there is measurement error in my assignment of patients to doctors. This doesn’t matter for the creation of maximum bonuses, but if I assigned particularly expensive visit to the wrong doctor in the base year, this would throw off the calculation of the improvement incentive. The second reason is that the documents detailing the bonus calculation are extremely vague as to what base year cost or base year length of stay is used.

## 9.2 Baseline Admission Probability Calculation Details

In order to get a measure of whether a patient would have been admitted in the absence of the bonus program, I estimate the “baseline admission probability” for each patient in my main sample of admitted Medicare patients. This baseline admission probability is calculated in two steps.

1. I take raw data from before the program started (2006-2008), which includes all Medicare patients who went to any New Jersey hospital, whether or not they were admitted. Using this pre-program data, I use a logistic specification to regress admission on a large set of medically relevant observable characteristics: the Charlson Comorbidity Index of the visit; the Charlson Comorbidity Index based on past visits; sex; the number

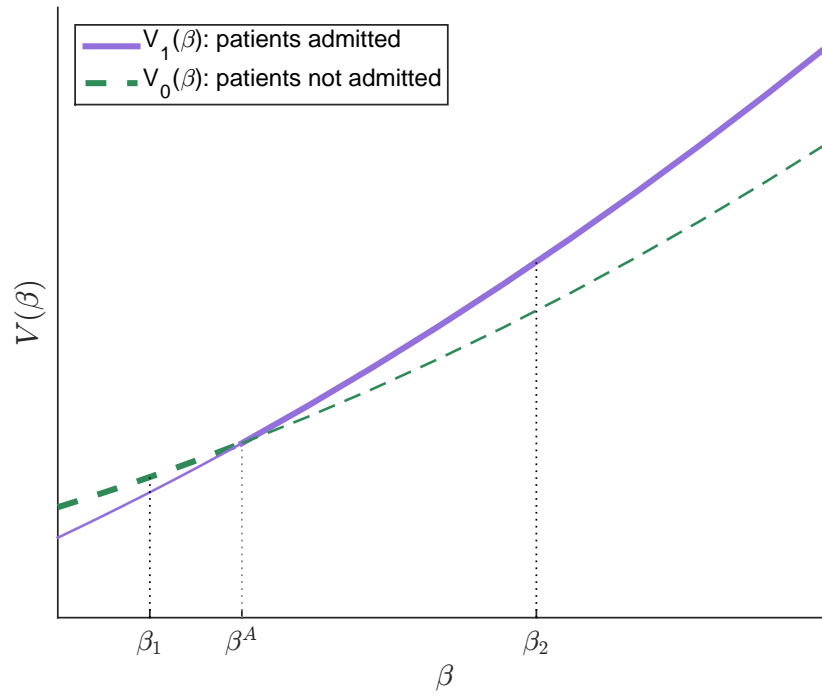
of body systems affected with chronic conditions, dummy variables for age categories, dummy variables for APR-DRGs, indicators for the following serious co-morbidities: cancer, chronic heart disease, asthma, diabetes, arthritis, atherosclerosis, chronic kidney disease, and HIV/AIDs, and quarter-by-year dummies. This model does a good job of predicting admission; the pseudo  $R^2$  of is 0.49.

2. I use the results of the above admission regression to predict whether or not each patient in my main sample (all of whom are actually admitted) would have been admitted in the pre-period, based on their medical characteristics. I call this predicted value the baseline admission probability, as it answers the question: would a patient with these medical characteristics have been admitted in the pre-period? If the introduction of the bonus program is associated with a decrease in the baseline admission probability of admitted patients, it is consistent with the bonuses inducing doctors to admit some patients who would not have been otherwise.

The baseline admission probability from step 2 is used as the outcome in Table 3.

### 9.2.1 Appendix Figures

Figure A.2: Doctor's Utility as a Function of  $\beta$ : without Bonuses



The bold line sections show the optimal decision rule as a function of  $\beta$ .



Figure A.1: Physician Dashboard

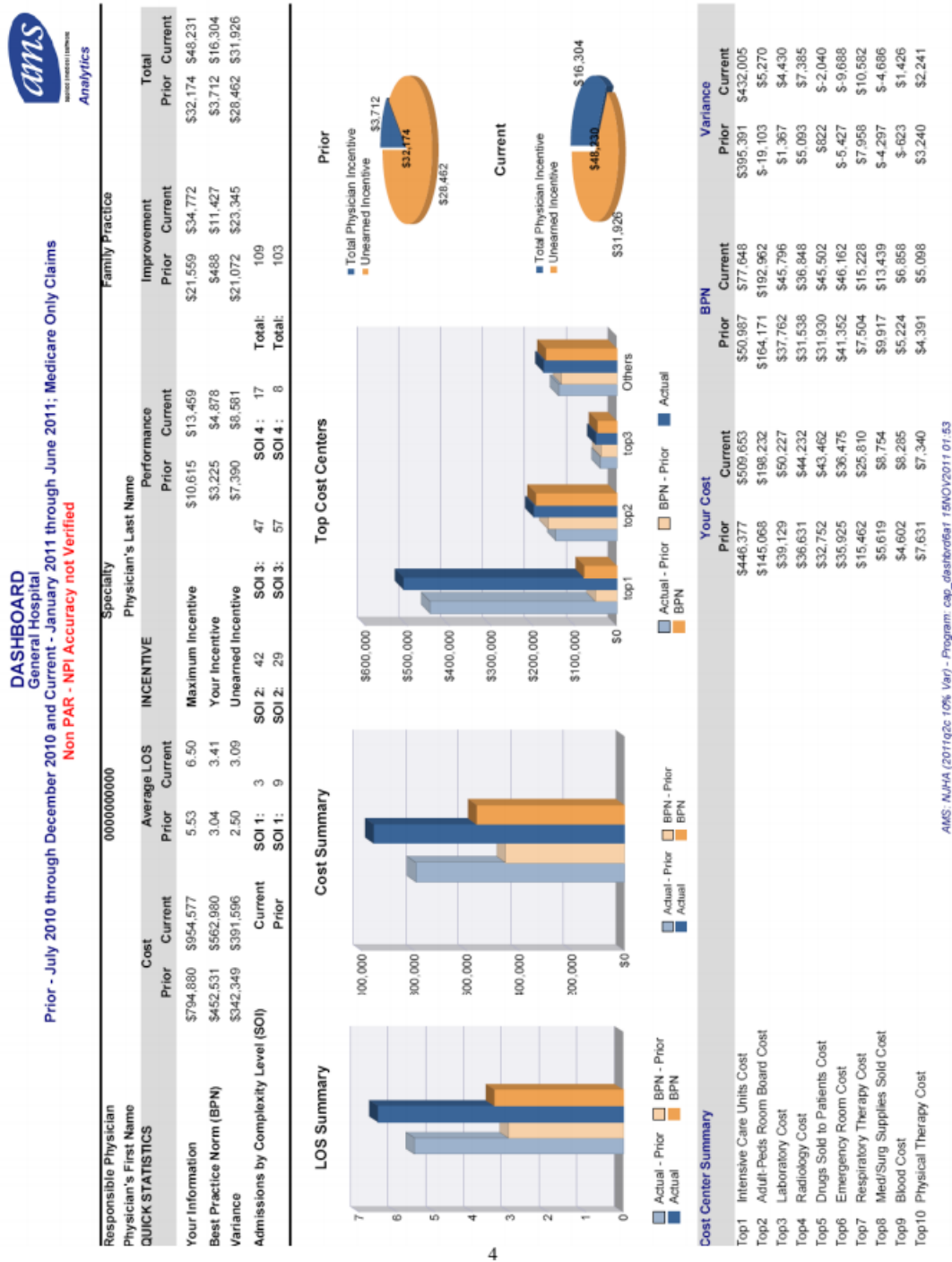
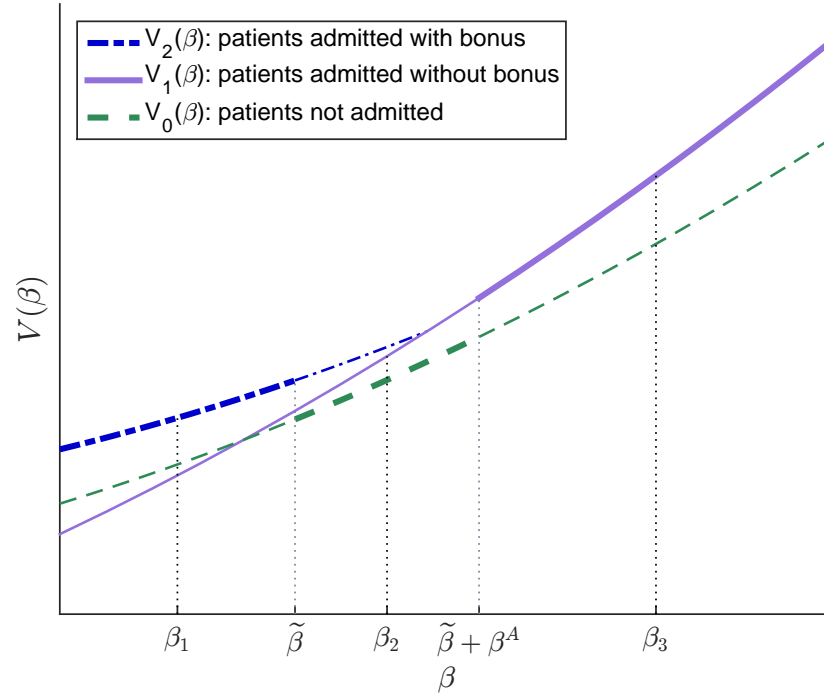
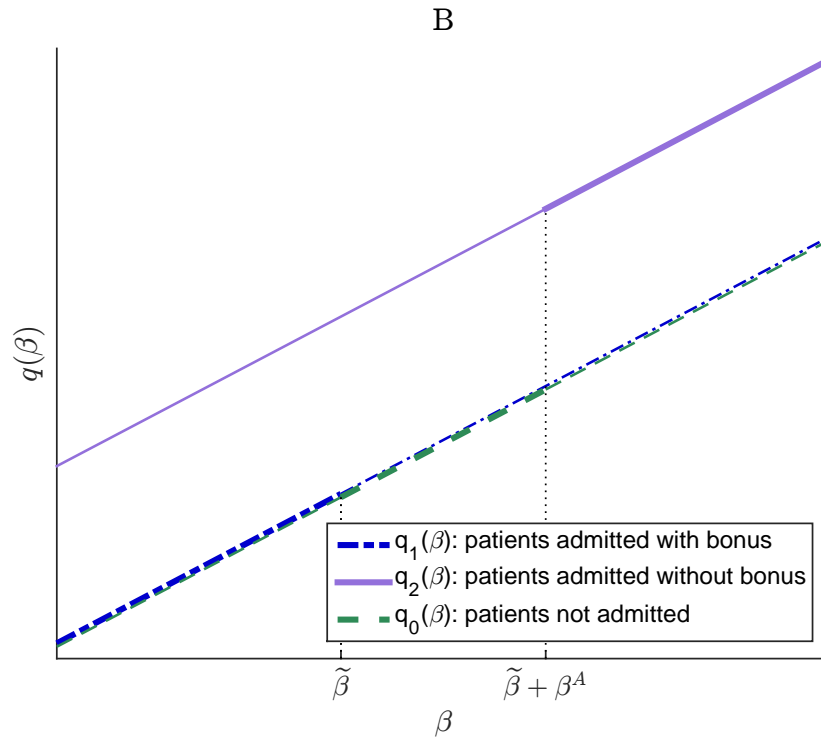
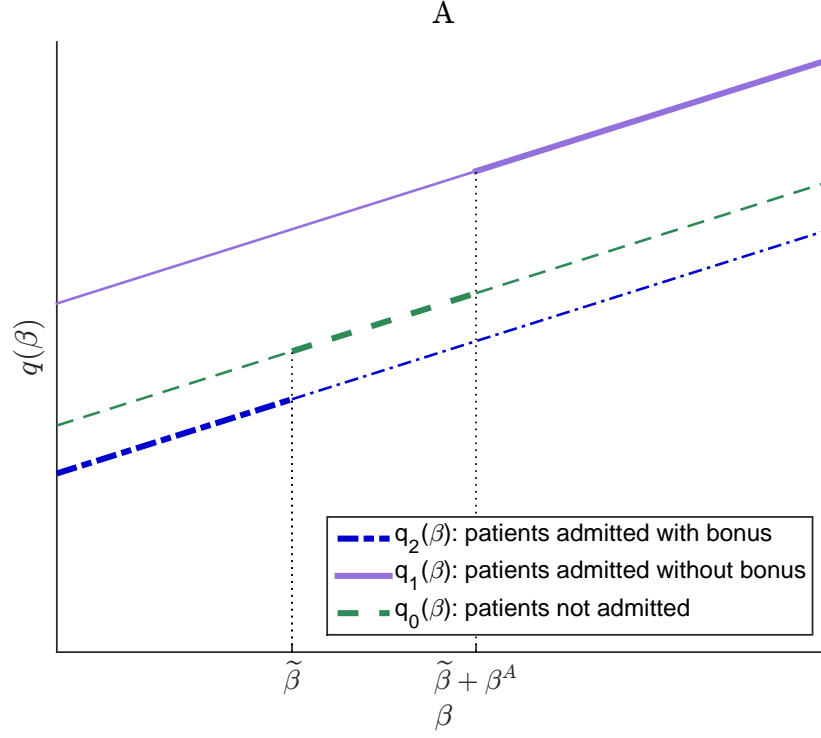


Figure A.3: Doctor's Utility as a Function of  $\beta$ : with Bonuses



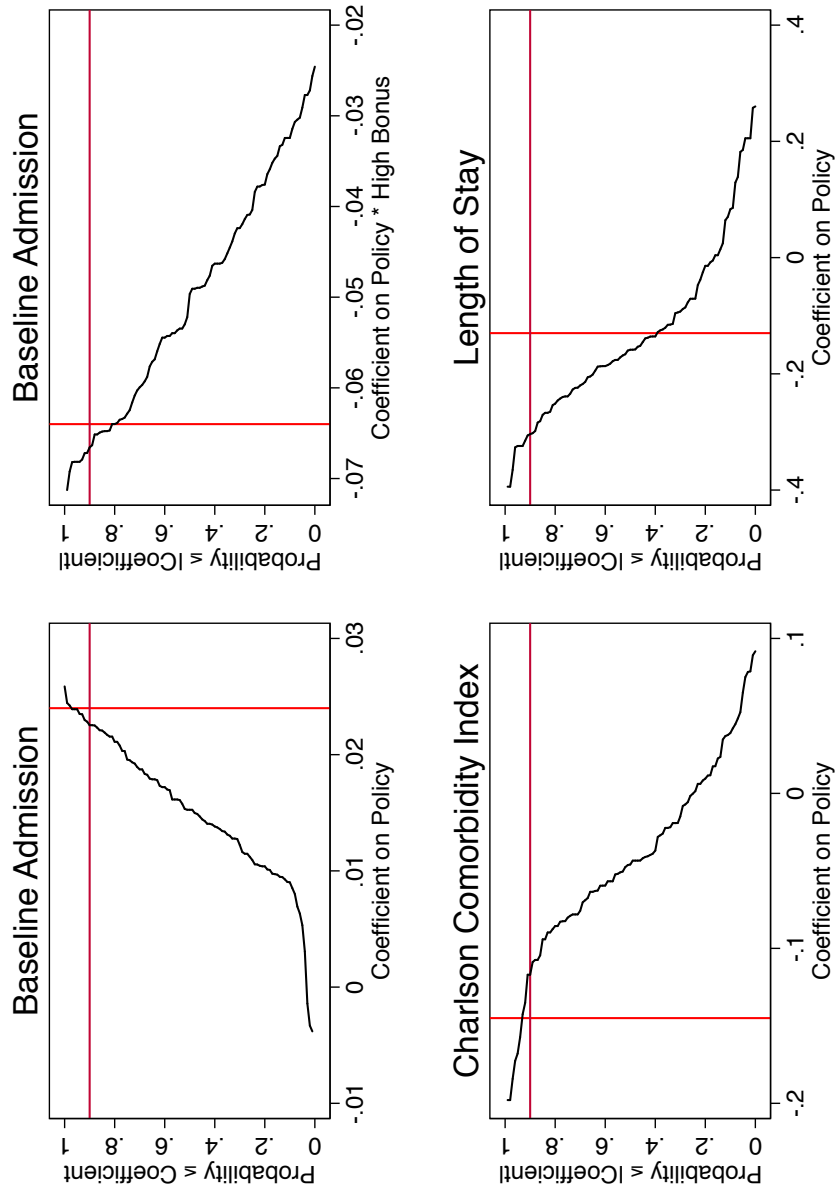
The bold line sections show the optimal decision rule as a function of  $\beta$ .

Figure A.4: Optimal Quantity of Care as a Function of  $\beta$ : with Bonuses



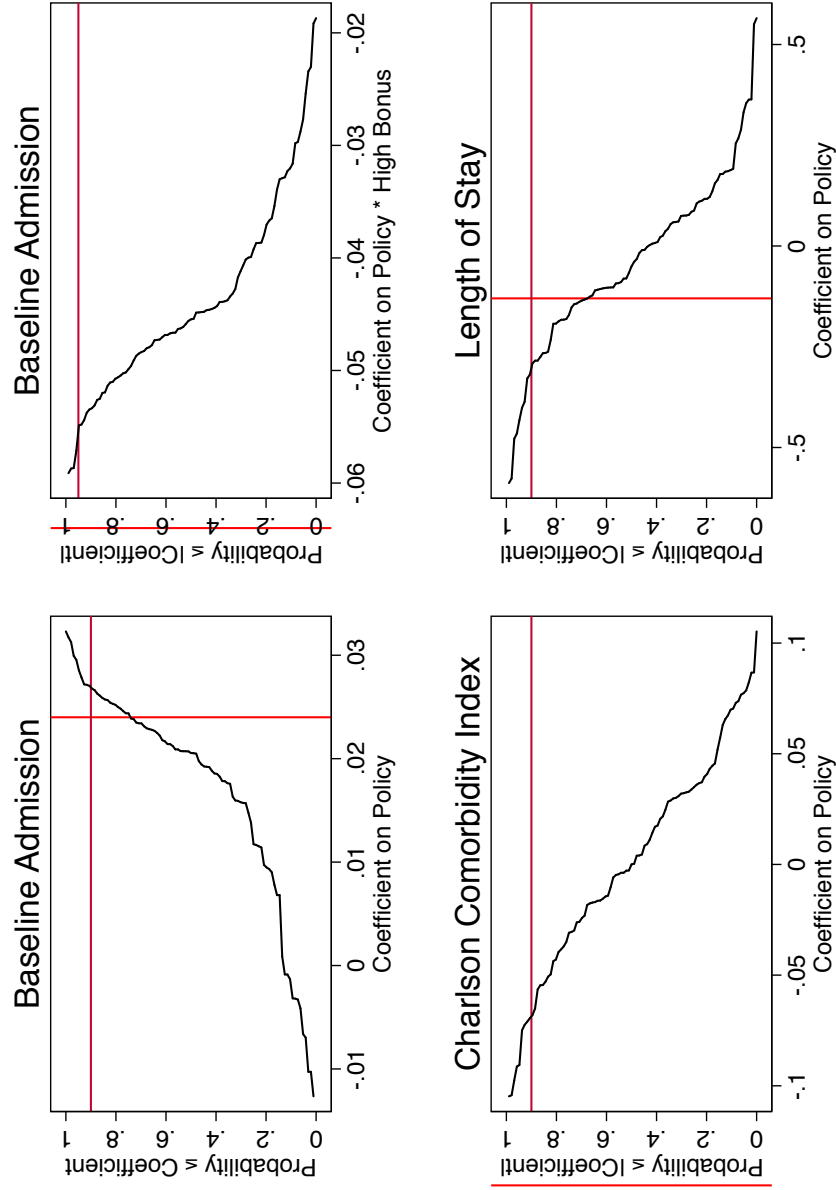
The bold line sections show the quantity of care provided along the optimal decision rule. Figure A.4A shows the optimal quantity of care under one set of parameters; Figure A.4B shows the optimal quantity of care under another set of parameters.

Figure A.5: Placebo Test: Randomize Dates



Based on 100 repetitions, with the starting dates of the Gainsharing Demonstration and the BPCI Model 1 program randomly assigned (two random dates are chosen; the first is assigned to the Gainsharing Demonstration, and the second to the BPCI Model 1 program). The vertical bar denotes the estimate in the real data, and the horizontal bar is placed at the 90<sup>th</sup> percentile.

Figure A.6: Placebo Test: Randomize Program Participation



Based on 100 repetitions, with the participation into the the Gainssharing Demonstration and the BPCI Model 1 program randomly assigned (holding fixed the number of hospitals participating in each at the true value). The vertical bar denotes the estimate in the real data, and the horizontal bar is placed at the 90<sup>th</sup> percentile.

### 9.2.2 Appendix Tables

Table A.1: All Patient Refined-Diagnosis Related Group (APR-DRG) Examples

	Severity of Illness (SOI)	Maximum Bonus	Number of Patients
Peptic ulcer and gastritis	1	\$189	632
	2	\$280	1,871
	3	\$510	1,552
	4	\$1,403	317
Hip joint replacement	1	\$308	15,711
	2	\$433	12,341
	3	\$911	1,439
	4	\$1,669	557

Notes: APR-DRG and SOI from 3M's grouping software; maximum incentive calculated according to gain-sharing formula. Number of patients are for admitted Medicare patients in main sample, which excludes those who went through the emergency room.

Table A.2: Top 10 Diagnosis Groups for Medical and Surgical Patients

APR-DRG	Frequency
<i>Medical patients</i>	
194: heart failure	24,702
720: septicemia & disseminated infections	21,410
140: chronic obstructive pulmonary disease	14,831
139: other pneumonia	13,085
860: rehabilitation	12,693
201: cardiac arrhythmia & conduction disorders	11,699
192: cardiac catheterization w circ disorder	10,083
460: renal failure	9,873
463: kidney and urinary tract infections	9,371
190: acute myocardial infarction	9,025
Total	389,604
<i>Surgical patients</i>	
302: knee joint replacement	45,995
175: percutaneous cardiovascular procedures w/o ami	39,691
301: hip joint replacement	30,048
221: major small & large bowel procedures	25,663
173: acute and subacute endocarditis	24,559
24: extracranial vascular procedures	12,408
171: perm cardiac pacemaker implant w/o ami, hf, shock	11,140
174: percutaneous cardiovascular procedures w ami	10,555
308: hip & femur procedures for non-trauma	10,415
163: cardiac valve procedures w cardiac catheterization	8,378
Total	425,410

Notes: APR-DRG and SOI from 3M's grouping software. Number of patients are for admitted Medicare patients who did not go through the emergency room.

Table A.3: Baseline Admission Probability: Predicting Admission

APR-DRG	Frequency
<i>Medical patients</i>	
194: heart failure	24,702
720: septicemia & disseminated infections	21,410
140: chronic obstructive pulmonary disease	14,831
139: other pneumonia	13,085
860: rehabilitation	12,693
201: cardiac arrhythmia & conduction disorders	11,699
192: cardiac catheterization w circ disorder	10,083
460: renal failure	9,873
463: kidney and urinary tract infections	9,371
190: acute myocardial infarction	9,025
Total	389,604
<i>Surgical patients</i>	
302: knee joint replacement	45,995
175: percutaneous cardiovascular procedures w/o ami	39,691
301: hip joint replacement	30,048
221: major small & large bowel procedures	25,663
173: acute and subacute endocarditis	24,559
24: extracranial vascular procedures	12,408
171: perm cardiac pacemaker implant w/o ami, hf, shock	11,140
174: percutaneous cardiovascular procedures w ami	10,555
308: hip & femur procedures for non-trauma	10,415
163: cardiac valve procedures w cardiac catheterization	8,378
Total	425,410

Notes: Sample is all Medicare patients in 2006. Dummy variables for APR-DRGs are also included.



## ER Medicare Patients

Table A.4: Effect of Program on Baseline Admission Probability: Medicare ER Patients

	Medical Patients		Surgical Patients	
	(1)	(2)	(3)	(4)
	Baseline Admission	Baseline Admission	Baseline Admission	Baseline Admission
Policy	0.003 (0.004)	0.002 (0.004)	0.000 (0.002)	-0.001 (0.002)
High bonus		0.317*** (0.011)		0.497*** (0.053)
Policy * high bonus		0.004 (0.003)		0.001 (0.002)
Mean	0.752	0.752	0.969	0.969
Clusters	78	78	76	76
N	1,585,753	1,585,753	69,742	69,742

Notes: Quarter-by-year, doctor, hospital, and diagnosis by severity of illness fixed effects also included. Standard errors clustered at the hospital level. \*p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01

Table A.5: Effect of Program on Ex-Ante Patient Health: Medicare ER Patients

	Medical Patients		Surgical Patients	
	CCI Past	Tot. Chronic	CCI Past	Tot. Chronic
Policy	-0.04 (0.02)	-0.02 (0.04)	-0.05 (0.04)	-0.06 (0.06)
Mean	2.562	5.159	1.918	4.504
Clusters	78	78	76	76
N	1,592,186	1,592,186	95,542	95,542

Notes: Quarter-by-year, doctor, hospital, and diagnosis by severity of illness fixed effects also included, as well as dummies for age categories, sex, and race. CCI stands for Charlson Co-morbidity Index, which is calculated based on information in previous visits. Tot. Chronic refers to the number of body systems affected by chronic conditions. Standard errors clustered at the hospital level. \*p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01

Table A.6: Effect of Program on Costs and Procedure Use: Medicare ER Patients

## Panel A: Medical Patients

	(1)	(2)	(3)	(4)	(5)	(6)
	Length of Stay	CT Scan	MRI	Any Imaging	Diag. Ultra	Total Costs
Policy	-0.026 (0.084)	-0.003 (0.005)	-0.003 (0.003)	0.004 (0.007)	0.009* (0.005)	471.649 (387.532)
Mean	5.211	0.0797	0.0306	0.135	0.0364	8,300.5
Clusters	78	78	78	78	78	65
<i>N</i>	1,592,186	1,592,186	1,592,186	1,592,186	1,592,186	1,341,584

## Panel B: Surgical Patients

	(1)	(2)	(3)	(4)	(5)	(6)
	Length of Stay	CT Scan	MRI	Any Imaging	Diag. Ultra	Total Costs
Policy	-0.249 (0.236)	-0.003 (0.006)	-0.002 (0.005)	0.002 (0.010)	0.007 (0.007)	249.986 (584.883)
Mean	7.655	0.0746	0.0268	0.202	0.0485	17,000.0
Clusters	76	76	76	76	76	65
<i>N</i>	95,542	95,542	95,542	95,542	95,542	82,959

Notes: Quarter-by-year, doctor, hospital, and diagnosis by severity of illness fixed effects also included, as well as dummies for age categories, sex, and race, and the variables measuring underlying health from Table 4. Standard errors clustered at the hospital level. \*p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01

## Near Medicare Patients (Ages 50-64)

Table A.7: Effect of Program on Baseline Admission Probability: Near Medicare Patient

	Medical Patients		Surgical Patients	
	(1)	(2)	(3)	(4)
	Baseline	Baseline	Baseline	Baseline
	Admission	Admission	Admission	Admission
Policy	0.000 (0.003)	0.004 (0.004)	-0.004 (0.005)	-0.001 (0.004)
High bonus		0.028 (0.028)		0.609*** (0.020)
Policy * high bonus		-0.011** (0.004)		-0.006** (0.002)
Mean	0.699	0.699	0.845	0.845
Clusters	79	79	78	78
N	183,259	183,259	247,488	247,488

Notes: Quarter-by-year, doctor, hospital, and diagnosis by severity of illness fixed effects also included. Standard errors clustered at the hospital level. \*p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01

Table A.8: Effect of Program on Ex-Ante Patient Health: Near Medicare Patients

	Medical Patients		Surgical Patients	
	CCI Past	Tot. Chronic	CCI Past	Tot. Chronic
Policy	-0.06 (0.04)	-0.05 (0.03)	-0.02 (0.02)	-0.01 (0.03)
Mean	1.866	4.162	0.885	3.531
Clusters	79	79	78	78
N	187,054	187,054	256,141	256,141

Notes: Quarter-by-year, doctor, hospital, and diagnosis by severity of illness fixed effects also included, as well as dummies for age categories, sex, and race. CCI stands for Charlson Co-morbidity Index, which is calculated based on information in previous visits. Tot. Chronic refers to the number of body systems affected by chronic conditions. Standard errors clustered at the hospital level. \*p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01

Table A.9: Effect of Program Costs and Procedure Use: Near Medicare Patients

## Panel A: Medical Patients

	(1)	(2)	(3)	(4)	(5)	(6)
	Length of Stay	CT Scan	MRI	Any Imaging	Diag. Ultra	Total Costs
Policy	-0.020 (0.162)	-0.000 (0.004)	-0.001 (0.003)	0.000 (0.007)	0.003 (0.004)	612.815 (527.039)
Mean	6.181	0.0344	0.0190	0.115	0.0422	10,459.2
Clusters	79	79	79	79	79	66
<i>N</i>	187,054	187,054	187,054	187,054	187,054	173,591

## Panel B: Surgical Patients

	(1)	(2)	(3)	(4)	(5)	(6)
	Length of Stay	CT Scan	MRI	Any Imaging	Diag. Ultra	Total Costs
Policy	0.056 (0.103)	-0.003 (0.003)	-0.000 (0.001)	0.008 (0.009)	-0.002 (0.004)	1,032.178** (458.950)
Mean	5.052	0.0180	0.00716	0.135	0.0458	16,148.8
Clusters	78	78	78	78	78	66
<i>N</i>	256,141	256,141	256,141	256,141	256,141	230,910

Notes: Quarter-by-year, doctor, hospital, and diagnosis by severity of illness fixed effects also included, as well as dummies for age categories, sex, and race, and the variables measuring underlying health from Table 4. Standard errors clustered at the hospital level. \* $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

## Single Hospital Doctors

Table A.10: Effect of Program on Baseline Admission Probability: Single Hospital Doctors

	Medical Patients		Surgical Patients	
	(1)	(2)	(3)	(4)
	Baseline Admission	Baseline Admission	Baseline Admission	Baseline Admission
Policy	0.001 (0.009)	0.031*** (0.008)	-0.005 (0.007)	0.010 (0.009)
High bonus		-0.080** (0.030)		-0.058*** (0.007)
Policy * high bonus		-0.077*** (0.007)		-0.024** (0.011)
Mean	0.911	0.911	0.947	0.947
Clusters	70	70	71	71
<i>N</i>	173,619	173,619	120,363	120,363

Notes: Quarter-by-year, hospital, and diagnosis by severity of illness fixed effects also included. Standard errors clustered at the hospital level. \*p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01

Table A.11: Effect of Program on Ex-Ante Patient Health: Single Hospital Doctors

	Medical Patients		Surgical Patients	
	CCI Past	Tot. Chronic	CCI Past	Tot. Chronic
Policy	-0.05 (0.06)	0.02 (0.07)	-0.02 (0.04)	0.02 (0.06)
Mean	2.653	4.135	1.603	2.737
Clusters	70	70	71	71
<i>N</i>	175,190	175,190	129,225	129,225

Notes: Quarter-by-year, hospital, and diagnosis by severity of illness fixed effects also included, as well as dummies for age categories, sex, and race. CCI stands for Charlson Co-morbidity Index, which is calculated based on information in previous visits. Tot. Chronic refers to the number of body systems affected by chronic conditions. Standard errors clustered at the hospital level. \*p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01

Table A.12: Effect of Program Costs and Procedure Use: Single Hospital Doctors

## Panel A: Medical Patients

	(1)	(2)	(3)	(4)	(5)	(6)
	Length of Stay	CT Scan	MRI	Diag. Ultra	Any Imaging	Total Costs
Policy	0.091 (0.304)	0.004 (0.003)	-0.001 (0.004)	-0.001 (0.004)	0.001 (0.007)	619.496 (660.191)
Mean	7.174	0.0194	0.0130	0.0334	0.0890	11118.3
Clusters	70	70	70	70	70	57
<i>N</i>	175,190	175,190	175,190	175,190	175,190	169,118

## Panel B: Surgical Patients

	(1)	(2)	(3)	(4)	(5)	(6)
	Length of Stay	CT Scan	MRI	Diag. Ultra	Any Imaging	Total Costs
Policy	0.127 (0.124)	-0.002 (0.004)	-0.000 (0.001)	0.011 (0.010)	0.020* (0.011)	1,730.635** (738.486)
Mean	6.871	0.0199	0.00710	0.0720	0.173	19727.8
Clusters	71	71	71	71	71	58
<i>N</i>	129,225	129,225	129,225	129,225	129,225	120,393

Notes: Quarter-by-year, hospital, and diagnosis by severity of illness fixed effects also included, as well as dummies for age categories, sex, and race, and the variables measuring underlying health from Table 4. Standard errors clustered at the hospital level. \*p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01

### 9.2.3 Mathematical Appendix: Model of doctor Decision-making

Doctors make three decisions: whether a patient is admitted,  $A \in \{0, 1\}$ , whether to admit a patient to a bonus hospital or a regular hospital,  $H \in \{0, 1\}$ , and how much care to provide,  $q$ . Patients vary only by their sickness level  $\beta \sim U([0, \bar{\beta}])$ . Doctors choose  $A$ ,  $H$ , and  $q$  to maximize a weighted average of their profits and the patient's utility from receiving treatment,

$$\begin{aligned} \max_{A, H, q} U(A, H, q; \beta) = & \underbrace{\lambda [aq + \max\{\alpha_0 - \alpha_1 q, 0\} * \mathbb{1}\{H = 1, A = 1\}]}_{\text{Doctor's profits}} \\ & + (1 - \lambda) \underbrace{\left[ \beta q + (\gamma q - C) * \mathbb{1}\{A = 1\} - \frac{b}{2} q^2 \right]}_{\text{Patient's utility of treatment}} \end{aligned} \quad (9)$$

Doctors' profits in the normal hospital are the amount of services provided,  $q$ , multiplied by a reimbursement rate,  $a$ . If  $A = 1$  and  $H = 1$ , doctors may also receive a bonus:  $\max\{\alpha_0 - \alpha_1 q, 0\}$ . The patient's utility function for medical care is concave in  $q$ . Sicker patients and admitted patients get more benefit from any treatment,  $q$ . Patients also care about admission. Care provided when a patient is admitted is more beneficial ( $\gamma$ ), but there is a fixed cost to the patient of admission,  $C$ .

Finally, doctors' choices are subject to three restrictions. First, the same number of patients must be admitted at each hospital. Second, Doctors can only admit as many patients as they would admit if there was no bonus. Third, all parameters are in  $\mathbb{R}^+$ , and  $0 < \lambda < 1$ .

#### Pre-Period: Neither Hospital Offers a Bonus

In order to know the capacity constraints that will constrain doctors in the full model, I first solve the model in the absence of the bonus (the "pre-period"). Doctors choose admission,  $A \in \{0, 1\}$ , and the quantity of care to provide,  $q$ . Since both hospitals are identical in the absence of the bonus, and doctors have to admit the same number of patients at each hospital, the hospital choice drops out.

Doctor's choose  $q$  and the hospital  $A \in \{0, 1\}$  to maximize the utility function:

$$\begin{aligned}
\max_{A,q} U(A, q; \beta) &= \underbrace{\lambda [aq]}_{\text{profit}} + (1 - \lambda) \underbrace{\left[ \beta q + (\gamma q - C) * \mathbb{1}\{A = 1\} - \frac{b}{2} q^2 \right]}_{\text{patient's utility from treatment}} \\
&= \max \left\{ \underbrace{\lambda [aq^*(\beta)] + (1 - \lambda) \left[ (\beta + \gamma) q^*(\beta) - C - \frac{b}{2} q^*(\beta)^2 \right]}_{\mathbf{V}_1(\beta) = U(q^*(\beta); \beta, A=1)}, \right. \\
&\quad \left. \underbrace{\lambda [aq^*(\beta)] + (1 - \lambda) \left[ \beta q^*(\beta) - \frac{b}{2} q^*(\beta)^2 \right]}_{\mathbf{V}_0(\beta) = U(q^*(\beta); \beta, A=0)} \right\}
\end{aligned}$$

**Proposition 1:** Under some parameter conditions, there exists a  $\beta^A$  such that all patients with  $\beta < \beta^A$  are not admitted, and all patients with  $\beta \geq \beta^A$  are admitted.

**Proof:** Need to know the doctor's utility as a function of  $\beta$ .

The value function is:

$$V(\beta) = \max \left\{ \underbrace{\lambda [aq_{(1)}] + (1 - \lambda) \left[ (\beta + \gamma) q_{(1)} - C - \frac{b}{2} q_{(1)}^2 \right]}_{V_1(\beta)}, \underbrace{\lambda [aq_{(0)}] + (1 - \lambda) \left[ \beta q_{(0)} - \frac{b}{2} q_{(0)}^2 \right]}_{V_0(\beta)} \right\} \quad (10)$$

Where from the first order conditions:

$$q_{(1)} = \frac{1}{b} \left[ \left( \frac{\lambda}{1 - \lambda} \right) (a_1) + \beta + \gamma \right] \quad (11)$$

and

$$q_{(0)} = \frac{1}{b} \left[ \left( \frac{\lambda}{1 - \lambda} \right) (a_1) + \beta \right] \quad (12)$$

The doctor's utility as a function of  $\beta$  is the upper envelope of  $V_1(\beta)$  and  $V_0(\beta)$ : the utility if all patients are admitted and if all patients are not admitted (see Figure 4). Assume the doctor admits all patients with  $\beta \in [\beta^A, \bar{\beta}]$ , and does not admit patients with  $\beta \in [0, \beta^A]$ . Now suppose a doctor were to admit a patient with  $\beta_1 < \beta^A$ . Since  $V_1(\beta_1) < V_0(\beta_1)$ , a



doctor would never choose to admit this patient. Likewise, suppose a doctor were to not admit a patient with  $\beta_2 > \beta^A$ . Now  $V_1(\beta_2) > V_0(\beta_2)$ , and again the doctor would be worse off. (See Figure A.2). Thus, patients with  $\beta \in [\beta^A, \bar{\beta}]$  are all admitted, and the rest are not admitted.

In order to solve the model in the post-period, it is necessary to know  $\beta^A$ . Define  $\beta^A$  such that  $U(q_{(0)}, \beta^A) = U(q_{(1)}, \beta^A)$ . Therefore,  $\beta^A$  solves:

$$\lambda [aq_{(0)}(\beta^A)] + (1 - \lambda) \left[ \beta^A q_{(0)}(\beta^A) - \frac{b}{2} q_{(0)}(\beta^A)^2 \right] \quad (13)$$

$$= \lambda [aq_{(1)}(\beta^A)] + (1 - \lambda) \left[ (\beta + \gamma) q_{(1)}(\beta^A) - C - \frac{b}{2} q_{(1)}(\beta^A)^2 \right] \quad (14)$$

$$\Rightarrow \beta^A = \frac{2a\gamma\lambda + 2bC\lambda - 2bC - \gamma^2\lambda + \gamma^2}{2\gamma(\lambda - 1)} \quad (15)$$

### Post-Period: Hospital 1 Offers a Bonus

Doctors again choose the quantity of care,  $q$ , the hospital,  $H \in \{0, 1\}$ , and admission,  $A \in \{0, 1\}$ . Now, however, hospital 1 introduces a cost reduction bonus, which is only available for doctors treating admitted patients. The bonus generates a difference between hospitals, and so the hospital choice becomes relevant. In addition, doctors are constrained by the pre-period capacity—they can only admit  $\bar{\beta} - \beta^A$  patients, and they must distribute the admitted patients evenly across hospitals.

Doctors choose  $q$ ,  $H \in \{0, 1\}$ , and  $A \in \{0, 1\}$  to maximize the utility function

$$\begin{aligned}
\max_{A,H,q} U(A, H, q; \beta) &= \underbrace{\lambda [aq + \max \{\alpha_0 - \alpha_1 q, 0\} * \mathbb{1} \{H = 1, A = 1\}]}_{\text{profit}} \\
&\quad + \underbrace{(1 - \lambda) \left[ \beta q + (\gamma q - C) * \mathbb{1} \{A = 1\} - \frac{b}{2} q^2 \right]}_{\text{patient's utility from treatment}} \\
&= \max \left\{ \underbrace{\lambda [aq^*(\beta) + \alpha_0 - \alpha_1 q^*(\beta)] + (1 - \lambda) \left[ (\beta + \gamma) q^*(\beta) - C - \frac{b}{2} q^*(\beta)^2 \right]}_{\mathbf{V}_2(\beta) = U(q^*(\beta); \beta, H=1, A=1)}, \right. \\
&\quad \underbrace{\lambda [aq^*(\beta)] + (1 - \lambda) \left[ (\beta + \gamma) q^*(\beta) - C - \frac{b}{2} q^*(\beta)^2 \right]}_{\mathbf{V}_1(\beta) = U(q^*(\beta); \beta, H=0, A=1)}, \\
&\quad \left. \underbrace{\lambda [aq^*(\beta)] + (1 - \lambda) \left[ \beta q^*(\beta) - \frac{b}{2} q^*(\beta)^2 \right]}_{\mathbf{V}_0(\beta) = U(q^*(\beta); \beta, A=0)} \right\}
\end{aligned}$$

Subject to the capacity constraint: a maximum of  $\beta'$  patients can be admitted at each hospital, where  $\beta' = \frac{\bar{\beta} - \beta^A}{2}$

**Proposition 2:** Under some parameter restrictions, there exists a  $\tilde{\beta}$  such that patients with  $\beta \in [0, \tilde{\beta}]$  are admitted at the bonus hospital, patients with  $\beta \in [\tilde{\beta}, \tilde{\beta} + \beta^A]$  are not admitted, and the remaining patients with  $\beta \in [\tilde{\beta} + \beta^A, \bar{\beta}]$  are admitted at either the bonus or non-bonus hospital.

**Proof:** Need to know the doctor's utility as a function of  $\beta$ .

The value function is

$$\begin{aligned}
V(\beta) &= \max \left\{ \underbrace{\lambda [aq_{(1)} + \alpha_0 - \alpha_1 q_{(1)}] + (1 - \lambda) \left[ (\beta + \gamma) q_{(1)} - C - \frac{b}{2} q_{(1)}^2 \right]}_{V_2(\beta)}, \right. \\
&\quad \left. \underbrace{\lambda [aq_{(2)}] + (1 - \lambda) \left[ (\beta + \gamma) q_{(2)} - C - \frac{b}{2} q_{(2)}^2 \right]}_{V_1(\beta)}, \underbrace{\lambda [aq_{(0)}] + (1 - \lambda) \left[ \beta q_{(0)} - \frac{b}{2} q_{(0)}^2 \right]}_{V_0(\beta)} \right\}
\end{aligned}$$

subject to the capacity constraint; only  $\beta'$  patients can be admitted to each hospital.

i. If the doctor chooses  $q$  under the first term, it must satisfy the following FOC:

$$q_{(2)} = \frac{1}{b} \left[ \left( \frac{\lambda}{1-\lambda} \right) (a - \alpha_1) + \beta + \gamma \right] \quad (16)$$

ii. If the doctor chooses  $q$  under the second term, it must satisfy the following FOC:

$$q_{(1)} = \frac{1}{b} \left[ \left( \frac{\lambda}{1-\lambda} \right) (a) + \beta + \gamma \right] \quad (17)$$

iii. If the doctor chooses  $q$  under the third term, it must satisfy the following FOC:

$$q_{(0)} = \frac{1}{b} \left[ \left( \frac{\lambda}{1-\lambda} \right) (a) + \beta \right] \quad (18)$$

Under certain conditions, the value function of the doctor is composed of three segments of the three parts of the value function, which maximize total utility (see Figure 5). Assume doctors decide which patients to admit and where to admit them by dividing their patients into three segments of  $\beta$ . They then admit the low  $\beta$  patients to the bonus hospital, do not admit the middle  $\beta$ s, and admit the highest  $\beta$ s to either hospital (randomizing over hospital such that they admit  $\beta'$  patients at both hospitals). Define the cut points as  $\tilde{\beta}$  and  $\tilde{\beta} + \beta^A$ . There is no patient  $\beta_2$  with  $\tilde{\beta} \leq \beta_2 \leq \tilde{\beta} + \beta^A$  where the doctor would prefer to admit  $\beta_2$  if it meant giving up admission for any patient  $\beta_1 < \tilde{\beta}$  or  $\beta_3 > \tilde{\beta} + \beta^A$ ; the doctor would be strictly worse off. This situation is depicted in Figure A.3.

The  $\tilde{\beta}'$  that partitions the range of  $\beta$  into these three groups solves  $U(q_{(2)}, \tilde{\beta}') - U(q_{(0)}, \tilde{\beta}') = U(q_{(1)}, \tilde{\beta}' + \beta^A) - U(q_{(0)}, \tilde{\beta}' + \beta^A)$ :

$$\begin{aligned} \tilde{\beta}' \text{ solves: } & \left( \lambda [aq_{(2)} + \alpha_0 - \alpha_1 q_{(2)}] + (1-\lambda) \left[ (\tilde{\beta} + \gamma) q_{(2)} - C - \frac{b}{2} q_{(2)}^2 \right] \right) \\ & - \left( \lambda [aq_{(0)'}] + (1-\lambda) \left[ \tilde{\beta} q_{(0)'} - \frac{b}{2} q_{(0)'}^2 \right] \right) \\ & = \left( \lambda [aq_{(1)}] + (1-\lambda) \left[ (\tilde{\beta} + \beta^A + \gamma) q_{(1)} - C - \frac{b}{2} q_{(1)}^2 \right] \right) \\ & - \left( \lambda [a_1 q_{(0)''}] + (1-\lambda) \left[ (\tilde{\beta} + \beta^A) q_{(0)''} - \frac{b_2}{2} q_{(0)''}^2 \right] \right) \end{aligned} \quad (19)$$

Where:

$$\begin{aligned} q_{(2)} &= \frac{1}{b} \left[ \left( \frac{\lambda}{1-\lambda} \right) (a - \alpha_1) + \tilde{\beta} + \gamma \right] \\ q_{(0)'} &= \frac{1}{b_2} \left[ \left( \frac{\lambda}{1-\lambda} \right) (a) + \tilde{\beta} \right] \\ q_{(1)} &= \frac{1}{b} \left[ \left( \frac{\lambda}{1-\lambda} \right) (a) + \tilde{\beta} + \beta^A + \gamma \right] \\ q_{(0)''} &= \frac{1}{b} \left[ \left( \frac{\lambda}{1-\lambda} \right) (a) + \tilde{\beta} + \beta^A \right] \\ \beta^A &= \frac{2a\gamma\lambda + 2bC\lambda - 2bC - \gamma^2\lambda + \gamma^2}{2\gamma(\lambda-1)} \end{aligned}$$

$$\tilde{\beta}' = \frac{\lambda(\lambda\alpha_1(2a-2\gamma-\alpha_1)+2b\alpha_0(\lambda-1)+2\gamma\alpha_1)+(\lambda-1)(2a\gamma\lambda+2bC\lambda-2bC-\gamma^2\lambda+\gamma^2)}{2\lambda\alpha_1(\lambda-1)} \quad (20)$$

However, because of the capacity constraint,  $\tilde{\beta} = \min \left\{ \tilde{\beta}', \frac{\tilde{\beta}-\beta^A}{2} \right\}$ .

**Proposition 3:** The direction of the change in  $q$  conditional on  $\beta$  from the pre- to the post-period for bonus-generating patients ( $\beta \in [0, \tilde{\beta}]$ ) is ambiguous.

**Proof:** If a patient would be admitted even without the bonus (in the pre-period), the introduction of the bonuses is associated with a lower  $q$ . For patients who are not admitted in the absence of the bonuses, however, the relevant comparison is between the  $q$  chosen under the bonus scheme ( $q_{(2)}$ ), and the  $q$  chosen when a patient is not admitted ( $q_{(0)}$ ). From the first order conditions of the doctor's value function, the optimal  $q$  when a patient is admitted at the bonus hospital is  $q_{(2)} = \frac{1}{b} \left[ \left( \frac{\lambda}{1-\lambda} \right) (a - \alpha_1) + \beta + \gamma \right]$ , and the optimal  $q$  when a patient is not admitted is  $q_{(0)} = \frac{1}{b} \left[ \left( \frac{\lambda}{1-\lambda} \right) (a) + \beta \right]$ . Whether the quantity of care provided for the bonus generating patients is higher or lower than the counterfactual of neither hospital offering a bonus is determined by the relative size of  $\gamma$  and  $\alpha_1$ . If  $\frac{1}{b} \left[ \gamma - \left( \frac{\lambda}{1-\lambda} \right) (\alpha_1) \right] < 0$ , the quantity of care provided for patients with  $\beta \in [0, \tilde{\beta}]$  is less when hospital 1 implements the bonus scheme than when neither hospital implements the bonus. On the other hand, if  $\frac{1}{b} \left[ \gamma - \left( \frac{\lambda}{1-\lambda} \right) (\alpha_1) \right] > 0$ , the quantity of care provided for patients with  $\beta \in [0, \tilde{\beta}]$  is greater when hospital 1 implements the bonus scheme than when neither hospital implements the bonus.

Figure A.4 shows both cases: 4.A demonstrates the case where the quantity of care provided for the bonus generating patients is less under the bonus program than the counterfactual of no bonuses; 4.B shows the opposite.

### Parameter conditions

The above interior solution exists as long as three sets of parameter restrictions hold. First,  $V_0(\beta)$  and  $V_1(\beta)$  cross; in the absence of the bonus, some patients are admitted and some patients are not admitted. Second, the bonuses are large enough to matter; the bonuses induce the doctor to admit the healthiest patient over the “healthiest” of the sick patients they formerly admitted. The second condition holds as long as  $U(q_{(2)}, 0) - U(q_{(0)}, 0) > U(q_{(1)}, \beta^A) - U(q_{(0)}, \beta^A)$ . Finally, the doctors always want to admit the sickest patients:  $U(q_{(2)}, \tilde{\beta}) - U(q_{(0)}, \tilde{\beta}) < U(q_{(1)}, \tilde{\beta}) - U(q_{(0)}, \tilde{\beta})$ .

## Working Paper Series

A series of research studies on regional economic issues relating to the Seventh Federal Reserve District, and on financial and economic topics.

The Effects of the Massachusetts Health Reform on Financial Distress <i>Bhashkar Mazumder and Sarah Miller</i>	<b>WP-14-01</b>
Can Intangible Capital Explain Cyclical Movements in the Labor Wedge? <i>François Gourio and Leena Rudanko</i>	<b>WP-14-02</b>
Early Public Banks <i>William Roberds and François R. Velde</i>	<b>WP-14-03</b>
Mandatory Disclosure and Financial Contagion <i>Fernando Alvarez and Gadi Barlevy</i>	<b>WP-14-04</b>
The Stock of External Sovereign Debt: Can We Take the Data at ‘Face Value’? <i>Daniel A. Dias, Christine Richmond, and Mark L. J. Wright</i>	<b>WP-14-05</b>
Interpreting the <i>Pari Passu</i> Clause in Sovereign Bond Contracts: It’s All Hebrew (and Aramaic) to Me <i>Mark L. J. Wright</i>	<b>WP-14-06</b>
AIG in Hindsight <i>Robert McDonald and Anna Paulson</i>	<b>WP-14-07</b>
On the Structural Interpretation of the Smets-Wouters “Risk Premium” Shock <i>Jonas D.M. Fisher</i>	<b>WP-14-08</b>
Human Capital Risk, Contract Enforcement, and the Macroeconomy <i>Tom Krebs, Moritz Kuhn, and Mark L. J. Wright</i>	<b>WP-14-09</b>
Adverse Selection, Risk Sharing and Business Cycles <i>Marcelo Veracierto</i>	<b>WP-14-10</b>
Core and ‘Crust’: Consumer Prices and the Term Structure of Interest Rates <i>Andrea Ajello, Luca Benzoni, and Olena Chyruk</i>	<b>WP-14-11</b>
The Evolution of Comparative Advantage: Measurement and Implications <i>Andrei A. Levchenko and Jing Zhang</i>	<b>WP-14-12</b>
Saving Europe?: The Unpleasant Arithmetic of Fiscal Austerity in Integrated Economies <i>Enrique G. Mendoza, Linda L. Tesar, and Jing Zhang</i>	<b>WP-14-13</b>
Liquidity Traps and Monetary Policy: Managing a Credit Crunch <i>Francisco Buera and Juan Pablo Nicolini</i>	<b>WP-14-14</b>

## **Working Paper Series** *(continued)*

Quantitative Easing in Joseph's Egypt with Keynesian Producers <i>Jeffrey R. Campbell</i>	<b>WP-14-15</b>
Constrained Discretion and Central Bank Transparency <i>Francesco Bianchi and Leonardo Melosi</i>	<b>WP-14-16</b>
Escaping the Great Recession <i>Francesco Bianchi and Leonardo Melosi</i>	<b>WP-14-17</b>
More on Middlemen: Equilibrium Entry and Efficiency in Intermediated Markets <i>Ed Nosal, Yuet-Yee Wong, and Randall Wright</i>	<b>WP-14-18</b>
Preventing Bank Runs <i>David Andolfatto, Ed Nosal, and Bruno Sultanum</i>	<b>WP-14-19</b>
The Impact of Chicago's Small High School Initiative <i>Lisa Barrow, Diane Whitmore Schanzenbach, and Amy Claessens</i>	<b>WP-14-20</b>
Credit Supply and the Housing Boom <i>Alejandro Justiniano, Giorgio E. Primiceri, and Andrea Tambalotti</i>	<b>WP-14-21</b>
The Effect of Vehicle Fuel Economy Standards on Technology Adoption <i>Thomas Klier and Joshua Linn</i>	<b>WP-14-22</b>
What Drives Bank Funding Spreads? <i>Thomas B. King and Kurt F. Lewis</i>	<b>WP-14-23</b>
Inflation Uncertainty and Disagreement in Bond Risk Premia <i>Stefania D'Amico and Athanasios Orphanides</i>	<b>WP-14-24</b>
Access to Refinancing and Mortgage Interest Rates: HARPing on the Importance of Competition <i>Gene Amromin and Caitlin Kearns</i>	<b>WP-14-25</b>
Private Takings <i>Alessandro Marchesiani and Ed Nosal</i>	<b>WP-14-26</b>
Momentum Trading, Return Chasing, and Predictable Crashes <i>Benjamin Chabot, Eric Ghysels, and Ravi Jagannathan</i>	<b>WP-14-27</b>
Early Life Environment and Racial Inequality in Education and Earnings in the United States <i>Kenneth Y. Chay, Jonathan Guryan, and Bhashkar Mazumder</i>	<b>WP-14-28</b>
Poor (Wo)man's Bootstrap <i>Bo E. Honoré and Luojia Hu</i>	<b>WP-15-01</b>
Revisiting the Role of Home Production in Life-Cycle Labor Supply <i>R. Jason Faberman</i>	<b>WP-15-02</b>

## Working Paper Series *(continued)*

Risk Management for Monetary Policy Near the Zero Lower Bound <i>Charles Evans, Jonas Fisher, François Gourio, and Spencer Krane</i>	<b>WP-15-03</b>
Estimating the Intergenerational Elasticity and Rank Association in the US: Overcoming the Current Limitations of Tax Data <i>Bhashkar Mazumder</i>	<b>WP-15-04</b>
External and Public Debt Crises <i>Cristina Arellano, Andrew Atkeson, and Mark Wright</i>	<b>WP-15-05</b>
The Value and Risk of Human Capital <i>Luca Benzoni and Olena Chyruk</i>	<b>WP-15-06</b>
Simpler Bootstrap Estimation of the Asymptotic Variance of U-statistic Based Estimators <i>Bo E. Honoré and Lujia Hu</i>	<b>WP-15-07</b>
Bad Investments and Missed Opportunities? Postwar Capital Flows to Asia and Latin America <i>Lee E. Ohanian, Paulina Restrepo-Echavarria, and Mark L. J. Wright</i>	<b>WP-15-08</b>
Backtesting Systemic Risk Measures During Historical Bank Runs <i>Christian Brownlees, Ben Chabot, Eric Ghysels, and Christopher Kurz</i>	<b>WP-15-09</b>
What Does Anticipated Monetary Policy Do? <i>Stefania D'Amico and Thomas B. King</i>	<b>WP-15-10</b>
Firm Entry and Macroeconomic Dynamics: A State-level Analysis <i>François Gourio, Todd Messer, and Michael Siemer</i>	<b>WP-16-01</b>
Measuring Interest Rate Risk in the Life Insurance Sector: the U.S. and the U.K. <i>Daniel Hartley, Anna Paulson, and Richard J. Rosen</i>	<b>WP-16-02</b>
Allocating Effort and Talent in Professional Labor Markets <i>Gadi Barlevy and Derek Neal</i>	<b>WP-16-03</b>
The Life Insurance Industry and Systemic Risk: A Bond Market Perspective <i>Anna Paulson and Richard Rosen</i>	<b>WP-16-04</b>
Forecasting Economic Activity with Mixed Frequency Bayesian VARs <i>Scott A. Brave, R. Andrew Butters, and Alejandro Justiniano</i>	<b>WP-16-05</b>
Optimal Monetary Policy in an Open Emerging Market Economy <i>Tara Iyer</i>	<b>WP-16-06</b>
Forward Guidance and Macroeconomic Outcomes Since the Financial Crisis <i>Jeffrey R. Campbell, Jonas D. M. Fisher, Alejandro Justiniano, and Leonardo Melosi</i>	<b>WP-16-07</b>

## **Working Paper Series** *(continued)*

Insurance in Human Capital Models with Limited Enforcement <i>Tom Krebs, Moritz Kuhn, and Mark Wright</i>	<b>WP-16-08</b>
Accounting for Central Neighborhood Change, 1980-2010 <i>Nathaniel Baum-Snow and Daniel Hartley</i>	<b>WP-16-09</b>
The Effect of the Patient Protection and Affordable Care Act Medicaid Expansions on Financial Wellbeing <i>Luojia Hu, Robert Kaestner, Bhashkar Mazumder, Sarah Miller, and Ashley Wong</i>	<b>WP-16-10</b>
The Interplay Between Financial Conditions and Monetary Policy Shock <i>Marco Bassetto, Luca Benzoni, and Trevor Serrao</i>	<b>WP-16-11</b>
Tax Credits and the Debt Position of US Households <i>Leslie McGranahan</i>	<b>WP-16-12</b>
The Global Diffusion of Ideas <i>Francisco J. Buera and Ezra Oberfield</i>	<b>WP-16-13</b>
Signaling Effects of Monetary Policy <i>Leonardo Melosi</i>	<b>WP-16-14</b>
Constrained Discretion and Central Bank Transparency <i>Francesco Bianchi and Leonardo Melosi</i>	<b>WP-16-15</b>
Escaping the Great Recession <i>Francesco Bianchi and Leonardo Melosi</i>	<b>WP-16-16</b>
The Role of Selective High Schools in Equalizing Educational Outcomes: Heterogeneous Effects by Neighborhood Socioeconomic Status <i>Lisa Barrow, Lauren Sartain, and Marisa de la Torre</i>	<b>WP-16-17</b>
Monetary Policy and Durable Goods <i>Robert B. Barsky, Christoph E. Boehm, Christopher L. House, and Miles S. Kimball</i>	<b>WP-16-18</b>
Interest Rates or Haircuts? Prices Versus Quantities in the Market for Collateralized Risky Loans <i>Robert Barsky, Theodore Bogusz, and Matthew Easton</i>	<b>WP-16-19</b>
Evidence on the within-industry agglomeration of R&D, production, and administrative occupations <i>Benjamin Goldman, Thomas Klier, and Thomas Walstrum</i>	<b>WP-16-20</b>
Expectation and Duration at the Effective Lower Bound <i>Thomas B. King</i>	<b>WP-16-21</b>



## **Working Paper Series** *(continued)*

The Term Structure and Inflation Uncertainty <i>Tomas Breach, Stefania D'Amico, and Athanasios Orphanides</i>	<b>WP-16-22</b>
The Federal Reserve's Evolving Monetary Policy Implementation Framework: 1914-1923 <i>Benjamin Chabot</i>	<b>WP-17-01</b>
Neighborhood Choices, Neighborhood Effects and Housing Vouchers <i>Morris A. Davis, Jesse Gregory, Daniel A. Hartley, and Kegan T. K. Tan</i>	<b>WP-17-02</b>
Wage Shocks and the Technological Substitution of Low-Wage Jobs <i>Daniel Aaronson and Brian J. Phelan</i>	<b>WP-17-03</b>
Worker Betas: Five Facts about Systematic Earnings Risk <i>Fatih Guvenen, Sam Schulhofer-Wohl, Jae Song, and Motohiro Yogo</i>	<b>WP-17-04</b>
The Decline in Intergenerational Mobility After 1980 <i>Jonathan Davis and Bhashkar Mazumder</i>	<b>WP-17-05</b>
Is Inflation Default? The Role of Information in Debt Crises <i>Marco Bassetto and Carlo Galli</i>	<b>WP-17-06</b>
Does Physician Pay Affect Procedure Choice and Patient Health? Evidence from Medicaid C-section Use <i>Diane Alexander</i>	<b>WP-17-07</b>
Just What the Nurse Practitioner Ordered: Independent Prescriptive Authority and Population Mental Health <i>Diane Alexander and Molly Schnell</i>	<b>WP-17-08</b>
How do Doctors Respond to Incentives? Unintended Consequences of Paying Doctors to Reduce Costs <i>Diane Alexander</i>	<b>WP-17-09</b>