Federal Reserve Bank of Chicago

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

Diane Alexander

REVISED
February 7, 2018

WP 2017-09

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

Diane Alexander†

February 7, 2018

Abstract

Billions of dollars have been spent on pilot programs searching for ways to reduce healthcare costs. I study one such program, where hospitals pay doctors bonuses for reducing the total hospital costs of admitted Medicare patients. 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.

†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 through the Center for Medicare and Medicare Innovation (CMMI).[1] 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, recent controversy has surrounded experimentation within Medicare, and in particular the mandatory participation of doctors in Medicare pilot programs. Tom Price, the first secretary of health and human services under President Trump, accused the Obama administration of trying to "commandeer clinical decision-making" by forcing doctors to participate. New pilot programs were scaled back and delayed in the first year of the Trump administration, and plans were released to "lead the Innovation Center in a new direction" (Verma, 2017). While the Obama administration favored large mandatory demonstrations, the Trump administration appears to favor smaller, voluntary demonstrations. How important are scale and mandatory participation for Medicare pilot programs?

I shed new light on the importance of experimental design in health care using the New Jersey Gainsharing Demonstration, a pilot program in which hospitals paid doctors bonuses for reducing the total treatment costs for Medicare admissions. The bonuses were designed to increase when total treatment costs decreased, and thus discourage the use of treatments with low marginal benefits. 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.

[1] 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).

Under the Gainsharing Demonstration, only the treatment of an admitted Medicare patient in a participating hospital could generate a bonus for the physician. Not all patients are admitted, not all hospitals participated in the demonstration, and most doctors in New Jersey admit patients in more than one hospital. Thus, doctors could increase their expected bonuses in three ways under the program: change which patients to admit, change where to admit them, and change how patients were treated.

Doctors responded to the bonuses by reallocating admission across patients—both by changing admission thresholds and diverting healthier patients into participating hospitals. Non-surgical patients admitted to participating hospitals had 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 exploit this knowledge 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 lowered costs or changed their procedure use in response to the bonuses.

My empirical strategy leverages the fact that most doctors in New Jersey treat patients in more than one hospital. I measure the effect of the bonuses by comparing changes in a doctor's behavior around implementation at a participating hospital to the same doctor in a non-participating hospital, using hospital discharge records. One critique of this doctor-level difference-in-difference specification is that 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 within-doctor identification strategy, changing practice styles would look like a null effect. Using an alternative strategy based on doctor-level program exposure, I rule out this alternative interpretation. Consistent with the main results, there is no evidence that the bonuses are associated with lower costs; if anything, costs appear to rise with program exposure.

Changing the composition of admitted patients has the potential to negatively affect both patients and Medicare itself. For patients, admission 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. Furthermore, 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 comparatively 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. The response of physicians to the New Jersey Gainsharing Demonstration highlights the importance of program design to both the effectiveness of the payment model, and the ability to generate internally valid estimates of the demonstration's efficacy.

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).[2] 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).[3] 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

[2]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.

[3]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.

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, and that hospitals respond to readmission penalties with selective readmission of returning patients (Brown et al., 2011; Duggan, 2004; Duggan and Hayford, 2013; Gupta, 2016; Leibowitz et al., 1992). 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. An exhaustive literature review is beyond the scope of this paper.[4] These lessons, however, 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 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 concludes.

[4] See for example Duflo (2004); Cullen et al. (2013); Allcott (2015)

2 Institutional Background

The employment relationship between doctors and the hospitals is complicated, and varies from place to place. For the most part, 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

Patients treated in hospitals are either admitted to the hospital (an "inpatient"), or treated on an outpatient basis. Patients treated only in the emergency department (ED) before being sent home are designated outpatient, as well as those sent to the hospital for diagnostic tests or same-day surgery (surgery that does not require an overnight hospital stay). On the other hand, admitted patients are under the care of a doctor with admitting privileges, who as written an order to admit the patient, and gives instructions for their care while in the hospital.

There are two major types of hospital admissions: elective and emergency. Emergency admissions originate in the hospital's emergency department, whereas elective admissions originate outside the hospital, such as a personal doctor seeing a patient in an office or clinic. Elective admissions involve a known medical complaint that requires further workup, treatment or surgery. Elective surgical admissions are those which are scheduled in advance, such as an elective knee surgery. For elective admissions, a personal doctor will generally request or arrange for you to be taken to a particular hospital, and has often reserved a bed.

Doctors treating non-emergent patients in hospitals have three main decisions to make: where to send the patient, whether they should be admitted, and the course of treatment. When deciding where to send a patient, doctors are limited to hospitals where they have pre-arranged relationships—so-called admitting or surgical privileges. 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 treats 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 decide whether to admit a patient and treat them, or treat the patient in the hospital on an outpatient basis. 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. Doctors considering admission weigh the benefits against the costs. While admitted patients are intensely monitored, and receive more care, admission is also 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. In nearly all diagnosis groups there are both patients treated with and without being admitted. Among Medicare patients with cardiac arrhythmia, for example, 51 percent of patients in my data are emergent admissions, 18 percent are elective admissions, 20 percent are treated outpatient in the ED, and 12 percent are outpatients with no ED revenue.

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 a fee-for-service system, whereas hospitals 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).[5] Traditional Medicare is no exception. Medicare Part A pays hospitals a fixed sum based on the patient's diagnosis (called diagnosis related groups, or "DRGs") for treating admitted patients. Conversely, physician services are paid for by Medicare Part B, a prospective payment system where doctors are paid separately each service provided to the patient.

Given these payment systems, 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 privileges, 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

[5]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).

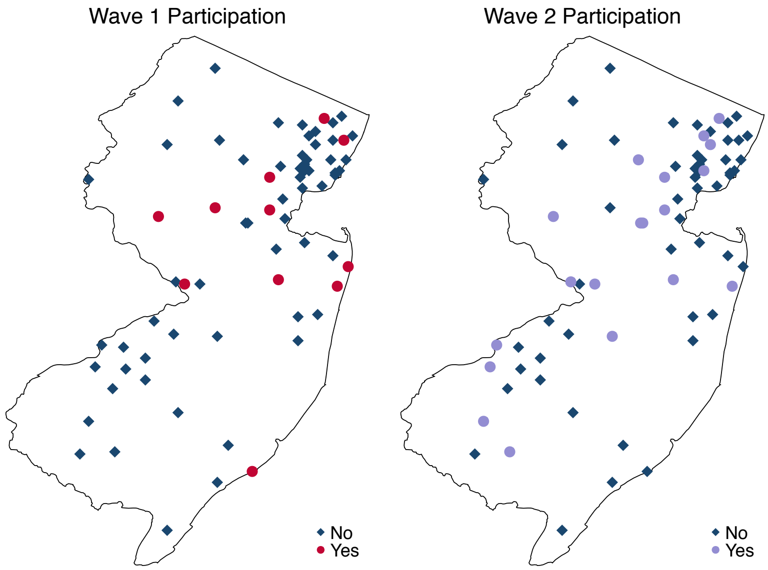
same privileges are what bring people into the hospital in the first place. Hospitals would like to use pay incentives to align the incentives of doctors with their own, but it is difficult in the current legal environment. Federal law constrains the ability of hospitals and doctors to participate in cost reduction programs, with the rationale that hospitals will pressure doctors into giving too little care, which would be bad for patient welfare.[6] 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. Under the program, doctors are still paid separately for each service provided by Medicare, 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.

[6] 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).

Figure 1: Hospital Locations



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.

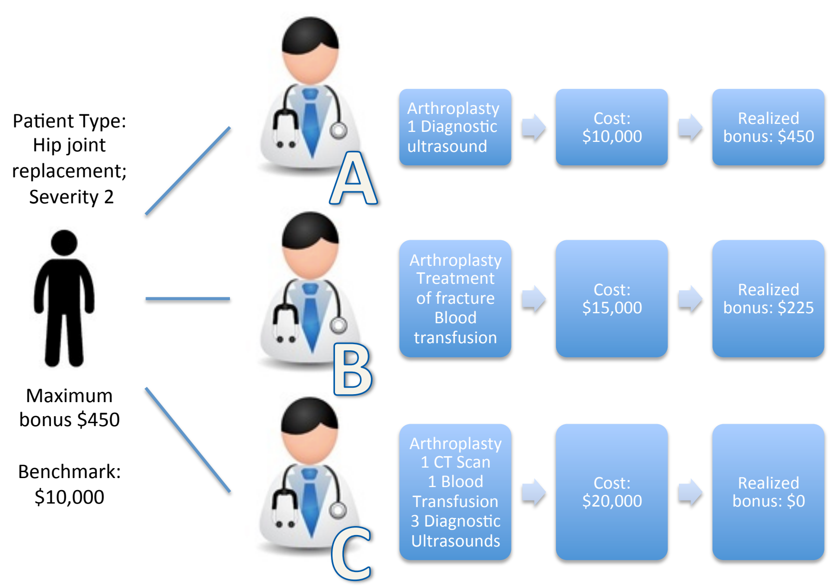
The Gainsharing Demonstration took place in two waves, which both applied only to doctors treating admitted Medicare patients. The initial phase took place in twelve New Jersey hospitals from July 1st, 2009 to July 1st, 2012. Eight of the original twelve hospitals opted to extend the program through March 31st, 2013. Based on the reported success of the Gainsharing Demonstration, the New Jersey Hospital Association applied for and secured 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 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.

While I do not have data on take up, anecdotal evidence suggests high physician participation in the demonstration. 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 involved in patient care, only the responsible doctor is eligible to receive a 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 from the hospital 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.

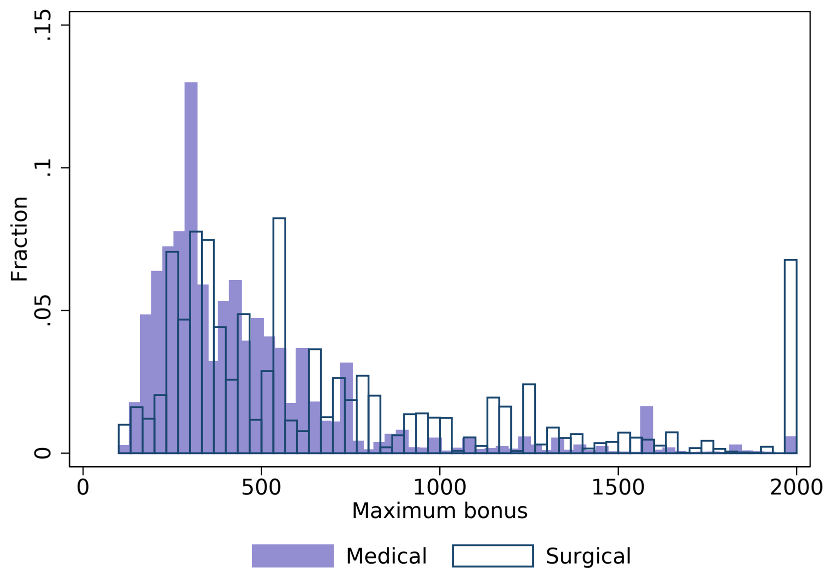
Figure 2: Hypothetical Bonus Calculation



Notes: Cost refers to the hospital treatment costs.

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 co-morbidities.[7] I recreate these maximum bonuses using list charges from hospital billing records deflated by Medicare's hospital level cost-to-charge ratio (more details 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.

While the formulas for calculating the maximum bonuses are opaque, doctors were given quarterly "dashboards", which gave them real time feedback on their 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 doctors to quickly become familiar with which types of patients generated large bonuses.

The rationale behind the formulas used to calculate maximum bonuses is that high cost variance within a diagnosis is a red flag, and indicates the existence of high cost patients who could be treated more cheaply. 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

[7]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.

cause of cost variation. Alternatively, high cost variance within a group of patients could be due to 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, and the patient treated by doctor C is receiving expensive but necessary care. 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

	Wave 1			Wave 2		
Participation	No	Yes	Diff	No	Yes	Diff
# of Hospitals (Gen. Medical/Surgical)	53	12		42	23	
Nongoverment 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 he Leapfrog Group (http://www.hospitalsafetyscore.org/). $*$ $p < 0.1$,$**$ $p < 0.05$,$***$ $p < 0.01$

As can be seen in Table 1, the selection process 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 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. 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 β, 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, where the weight placed on profit is λ. Doctors are paid a reimbursement rate, a, for each unit of care, q, provided to the patient. The payment, a, does not not depend on the hospital choice or whether the patient is admitted. Thus, the doctor's profit from treating a patient is aq. 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 patient's utility from medical treatment is concave in q, with sicker patients (those with a higher β) benefiting more from medical care. The utility a patient derives from medical care is:

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

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 γ). 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.[8],[9]

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:

[8] The randomization can interpreted as patients having a slight preference for the closest hospital, and patients being evenly distributed across space.

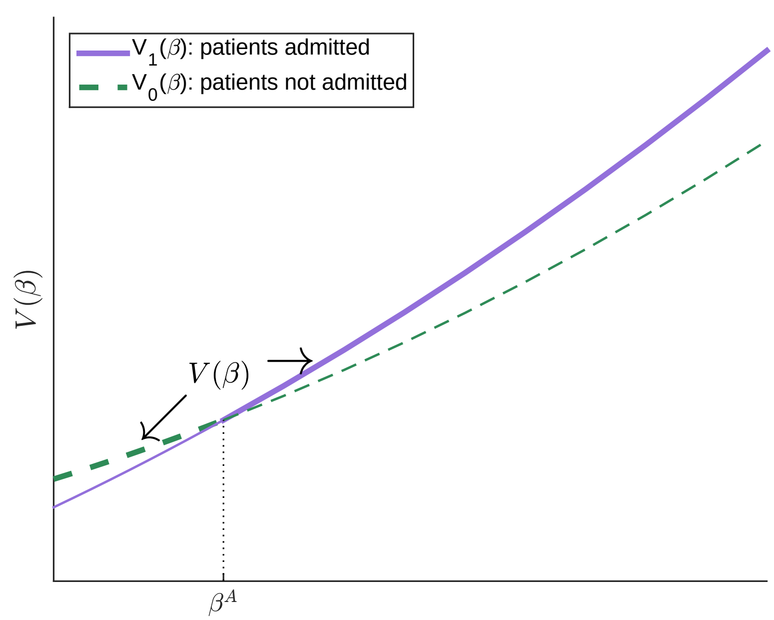
[9] Doctors could assign patients such that any proportion goes to each hospital; I use 50-50 to keep examples simple.

$$\max_{q,A} U(q,A;\beta) = \lambda \underbrace{[aq]}_{profit} + (1-\lambda) \underbrace{\left[\beta q + (\gamma q - C) * \mathbb{1}\{A=1\} - \frac{b}{2} q^2\right]}_{patient's\,utility\,from\,treatment}$$

$$= \max \left\{ \underbrace{\lambda [aq^*(\beta)] + (1-\lambda)\left[(\beta+\gamma) q^*(\beta) - C - \frac{b}{2} q^*(\beta)^2\right]}_{\mathbf{V_1(\beta)}=U(q^*(\beta);\beta,A=1)}, \underbrace{\lambda [aq^*(\beta)] + (1-\lambda)\left[\beta q^*(\beta) - \frac{b}{2} q^*(\beta)^2\right]}_{\mathbf{V_0(\beta)}=U(q^*(\beta);\beta,A=0)} \right\}$$

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 β) dislike admission, while for sicker patients (high β), admission is beneficial. Since doctors take into account patient's preferences, there is a sickness threshold β^A which defines the optimal admission rule.

Figure 4: Doctor's Utility as a Function of β: without Bonus
Without Bonus



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

Proposition 1: Under some parameter conditions, there exists a β^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 $V_0(\beta) > V_1(\beta)$. β^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 Gainsharing 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:

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

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 α_0, and α_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, $\beta' = \frac{\bar{\beta}-\beta^A}{2}$.[10] 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.

[10] The capacity constraint β' is just a number; doctors can admit any patients they want, and are not constrained to pick patients in an interval of β.

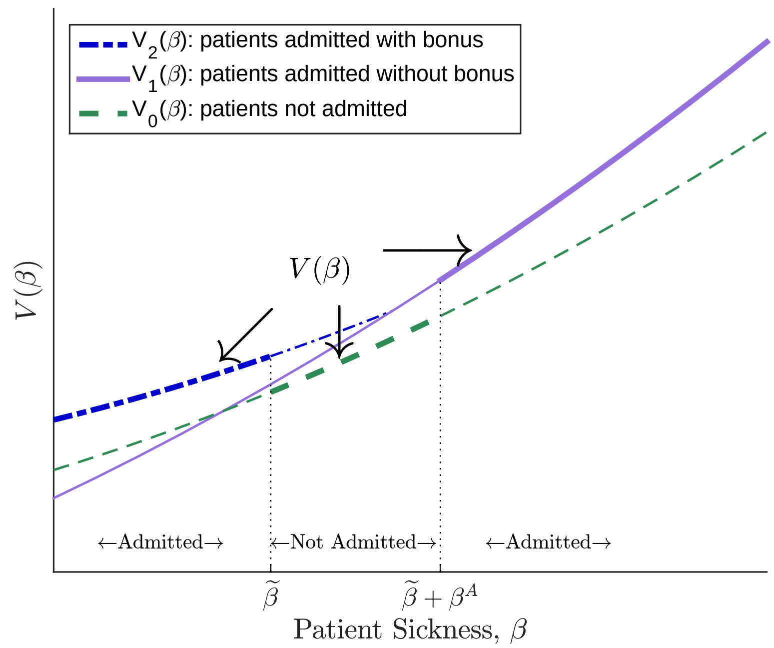
Doctors now choose $A \in \{0,1\}$, $H \in \{0,1\}$, and q to maximize the utility function

$$
\begin{aligned}
\max_{q,H,A} U(q,H,A;\beta) &= \lambda \underbrace{\left[aq + \max\{\alpha_0 - \alpha_1 q, 0\} * \mathbb{1}\{H=1, A=1\}\right]}_{profit} \\
&\quad + (1-\lambda) \underbrace{\left[\beta q + (\gamma q - C) * \mathbb{1}\{A=1\} - \frac{b}{2} q^2\right]}_{patient's\, utility\, from\, treatment} \\
&= \max \left\{ \underbrace{\lambda \left[aq^*(\beta) + \alpha_0 - \alpha_1 q^*(\beta)\right] + (1-\lambda)\left[(\beta+\gamma) q^*(\beta) - C - \frac{b}{2} q^*(\beta)^2\right]}_{\mathbf{V_2(\beta)} = U(q^*(\beta);\beta,H=1,A=1)}, \right. \\
&\quad \underbrace{\lambda \left[aq^*(\beta)\right] + (1-\lambda)\left[(\beta+\gamma) q^*(\beta) - C - \frac{b}{2} q^*(\beta)^2\right]}_{\mathbf{V_1(\beta)} = U(q^*(\beta);\beta,H=0,A=1)}, \\
&\quad \left. \underbrace{\lambda \left[aq^*(\beta)\right] + (1-\lambda)\left[\beta q^*(\beta) - \frac{b}{2} q^*(\beta)^2\right]}_{\mathbf{V_0(\beta)} = U(q^*(\beta);\beta,A=0)} \right\}
\end{aligned}
$$

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 increasing in β. 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 β) 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 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 β patients because of the bonus, and the highest β patients because these patients have the highest utility from treatment.

Figure 5: Doctor's Utility as a Function of β: with Bonus



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

Proposition 2: Under some parameter restrictions, there exists a $\widetilde{\beta}$ such that patients with $\beta \in \left[0, \widetilde{\beta}\right]$ are admitted at the bonus hospital, patients with $\beta \in \left[\tilde{\beta}, \tilde{\beta} + \beta^A\right]$ are not admitted, and the remaining patients with $\beta \in \left[\widetilde{\beta} + \beta^A, \bar{\beta}\right]$ 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 β patients at the bonus hospital up until $\widetilde{\beta}$. They will also admit the sickest $\bar{\beta} - \left(\widetilde{\beta} + \beta^A\right)$ patients, randomizing over hospital choice such that they admit β' total patients at each hospital. The patients with βs 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.

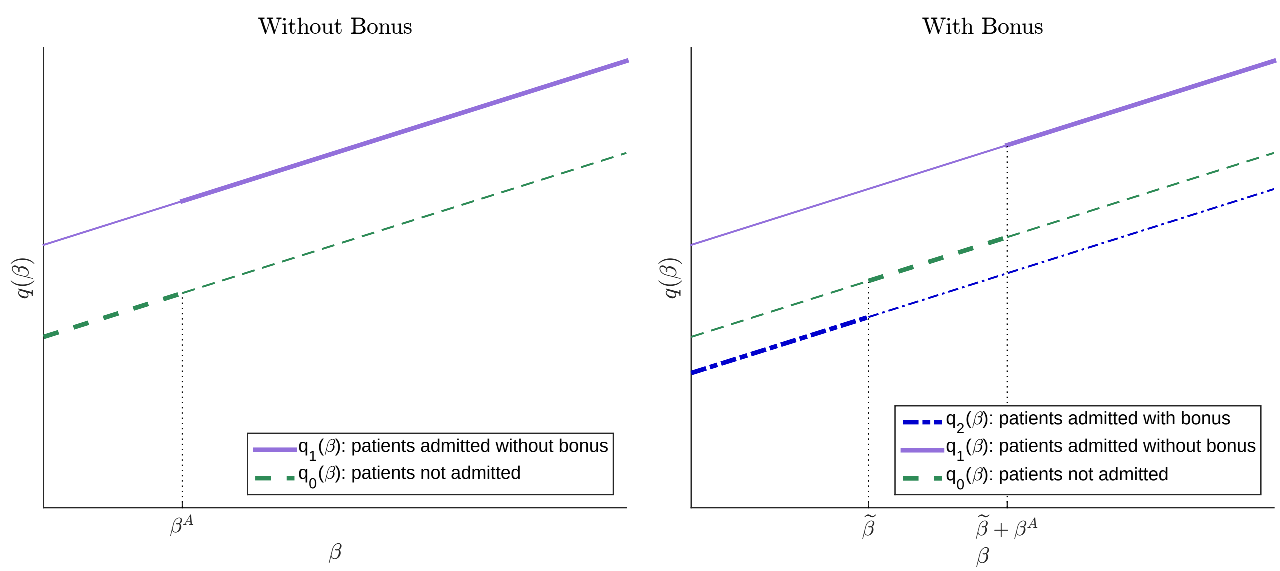
The cost reduction bonuses introduce two distortions. First, the bonuses increase the probability of admission for the healthiest patients and decrease the probability of admission for sicker patients. Many patients with $\beta < \tilde{\beta}$ are not admitted without the bonus (the "pre-period"), and all are admitted in the when the bonus is introduced (the "post-period"). On the other hand, many "medium sick" patients with $\beta \in \left[\tilde{\beta}, \tilde{\beta} + \beta^A\right]$ 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 α_1). The second is upward pressure on q from admission (represented by γ).

Proposition 3: The direction of the change in q conditional on β from the pre- to the post-period for bonus-generating patients ($\beta \in \left[0, \widetilde{\beta}\right]$) 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 γ and α_1. For more details, see the Mathematical Appendix.

Figure 6: Optimal Quantity of Care as a Function of β



Notes: The bold line sections show the optimal quantity of care provided as a function of β, 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 (β). 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.[11] I also create doctor identifiers using the recorded license numbers of doctors and surgeons. The final file includes unique identifiers for both 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.

The main analysis sample consists of all visits where a Medicare beneficiary was admitted to a New Jersey hospital. 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 over the sample period were also dropped,

[11]The Levenshtein edit distance is used to match names, because of problems with typos and misspellings (stata command strgroup).

as these doctors likely did not have enough patients to qualify for the bonus program.

The full sample consists of approximately 1.3 million medical visits and 580,000 surgical hospital visits. Of these, 69 percent of medical and 27 percent of surgical visits were admitted through the emergency room. The route of admission is important, as doctors have no scope to send a patient already in an emergency room to a different hospital based on their latent health. They could, however, become more likely to admit healthier ER patients in participating hospitals. In the analysis that follows, I will use route of admission to help disentangle sorting and admission decision responses.

Summary statistics are shown in Table 2. The admitted Medicare beneficiaries are predominately white, with an average age of 75, and are slightly over half female. The doctors treating these patients worked at 2.3 hospitals on average, with 38 percent ever working in both a participating and non-participating hospital. The average maximum bonus a doctor could earn for treating a surgical patient was \$697, versus \$513 for a medical patient. 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)).[12]

[12] Medicare facility charge for repair of knee ligament (CPT 27405), 2012.

Table 2: Main Sample Characteristics

Medicare Patients	Medical	Surgical	Doctors	
Characteristics				
Age	75.327	74.465	Avg. # of patients	366.793
White	0.784	0.826	Avg. # of hospitals	2.340
Black	0.138	0.100	Med. # of hospitals	2.000
Woman	0.569	0.534	Ever in policy hosp	0.695
ER revenue	0.694	0.267	Ever in other hosp	0.682
In policy hospital	0.462	0.535	Ever in both types	0.377
APR-DRGs	163	118	Avg. MPI (medical)	\$466
N	1,271,574	580,140	Avg. MPI (surgical)	\$706
Outcomes			N	3,639
Latent health:				
Charlson index	2.670	1.803		
Charlson index = 0	0.294	0.440		
Surgical risk index	2.160	1.633		
Surgical risk index = 0	0.143	0.226		
Costs and Quantity:				
Length of stay	5.753	6.893		
Total costs	9,333	18,226		
CT scan	0.057	0.037		
MRI	0.024	0.013		
Diagnostic ultrasound	0.035	0.058		
Any imaging	0.120	0.191		

Notes: admitted Medicare patients in general medical/surgical hospitals (2006-2013). APR-DRG stands for All Patient Refined-Diagnosis Related Group. In policy hosp is fraction treated at hospitals that ever participate in the program.

My primary measure of latent patient health is the Charlson co-morbidity index. This index is designed to predict the one-year mortality for hospital inpatients based on the presence of comorbid conditions, with a higher score on the index denoting a sicker patient. The index is computed based on the presence of 17 conditions, each weighted by the associated risk of death, and has been widely validated.[13] Furthermore, the Charlson index has been shown to be strongly predictive of hospital resource utilization, which makes it uniquely well

[13]The Charlson Co-morbidity Index is a weighted sum over the following conditions (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)

suited for measuring latent health as it relates to cost of treatment (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, which exploits the time-series dimension of the data. The leave-out index uses data on all hospital visits made by each patient (regardless of admission), excluding any diagnoses recorded during the current visit.[14] Thus, the measure is an index of latent health of patients that are known to the doctor (or at least correlated with information known to the doctor), but is not used in the bonus formula.

While the Charlson co-morbidity index is a useful summary measure of patient health, it was developed to measure the mortality risks of medical inpatients, and may be less sensitive as a measure of preoperative physical status. Therefore, I supplement the Charlson index with an index based on general surgical risk factors. Specifically, I use all conditions in the universal surgical risk calculator that can be identified via ICD-9-CM diagnosis code (Best et al., 2002; Bilimoria et al., 2013).[15] I construct an surgical risk factor index that is the sum of these indicators, again leaving out diagnoses recorded during the index visit. A higher score on the surgical risk factor index represents a sicker patient.

Patients admitted under a medical diagnosis have a higher disease burden on average compared to surgical visits, as measured by both the average Charlson index and surgical risk factor index, as well as the proportion scoring a zero on each index. Patients admitted under a medical diagnosis are about a year older, more likely to be black, and much more likely to be admitted through the emergency room. These patients are also more varied. There are 163 medical diagnosis groups, compared to 118 surgical diagnosis groups, and just 10 diagnosis groups make up half of all patients admitted under surgical codes (Table A.2 lists the most common diagnosis groups for medical and surgical patients in the sample).

To examine whether the bonuses changed procedure use or lowered costs, I also look across several measures. The first two are summary measures of resource use: length of stay and total costs. 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 total hospital costs incurred during a visit are estimated using the total list charges reported in

[14]If a patient appears just once in the data, they are assigned a zero. This strategy will introduce some measurement error, as a patient could have a serious disease, but not a previous hospital visit. However, given the seriousness of the conditions used in the Charlson co-morbidity index and the high disease burden of Medicare patients, I expect the vast majority of patients with one of these conditions will appear in the data multiple times. Of admitted visits covered by Medicare, just 3.4% of admitted medical visits and 5.6% of admitted surgical visits are to patients seen only once in the data.

[15]These conditions are disseminated cancer, diabetes, hypertension, dyspnea, COPD, acute renal failure, and ascites or congestive heart failure within 30 days preoperatively. The exact ICD-9-CM codes used to define these indicators are listed in Table A.3.

the discharge data, deflating them by Medicare's hospital-year level cost-to-charge ratio, and then converting them to real 2010 dollars. The Medicare cost-to-charge ratio is explicitly designed to translate list charges into an estimate of the resource cost of inpatient care. Surgical cases are more resource intensive on both measures, spending on average an extra day in the hospital and incurring nearly twice the costs of medical cases.

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).[16] 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. [17]

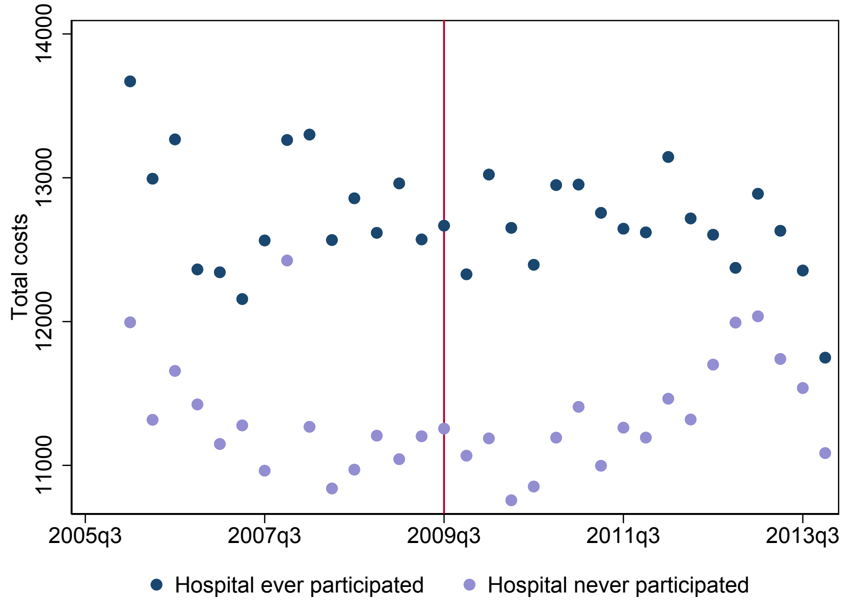
4.2 Empirical Strategy

The main challenge in identifying the effect of the cost-reduction bonuses on doctor decision-making is that participating hospitals are different from hospitals that did not take up the program. In particular, hospitals which seek out a program designed to reduce wasteful spending may be on different treatment cost trajectories from those that do not. Figure 7 plots average treatment costs for admitted Medicare patients across hospitals by future program participation, and shows that hospitals that took up the program had higher average costs than those that did not. The level difference in costs reflects the fact that participation was not randomly assigned; hospitals with higher average costs initially had more to gain from a program incentivizing doctors to reduce costs. However, there is no evidence of differential pre-trends between the two groups with respect to costs.

[16]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)

[17]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).

Figure 7: Average Costs by Hospital Participation in Demonstration



Notes: Average quarterly costs for admitted Medicare patients, averaged across hospitals that ever participated in the Gainsharing Demonstration, versus hospitals that never participated. Costs are deflated by Medicare cost-to-charge ratios.

Despite no evidence of differential pre-trends in costs across the hospital groups, I use a within-doctor rather than a within-hospital identification strategy for two reasons. First, studying the responses of individual doctors to the introduction of the bonus program allows for a clean connection between the conceptual framework in Section 3 and the empirical analysis. Second, including doctor fixed effects allows me to control for time-invariant physician-level characteristics. In the within-doctor specification, I do not have to worry about different compositions of surgeons and physicians across hospitals driving the results, since 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 (in quarters). $Policy_{dht}$ 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 β_1. The patient characteristics included in X_{it} vary by specification. They are omitted when looking at the effect of the bonuses on latent health, as the goal is measure the effect of the bonuses on the composition of admitted patients. When considering the effect of the bonuses on costs and procedure use, however, I want to control for changes in patient composition. Thus, these regressions thus include age, sex, race, and measures of latent health. Hospital, quarter-by-year, and doctor fixed effects are also included in all regressions (λ_h, λ_t and λ_d). In some specifications, type (APR-DRG by SOI pairs) fixed effects are also included. 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 is that it will not be able to detect any responses to the bonuses that occur in both hospitals. For now, I rule out this type of behavior. I address this issue with an alternative identification strategy in section 6, which examines total resource use as a function of program exposure.

All analyses are done 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 Results

5.1 Effects of Bonuses on Latent Health of Admitted Patients

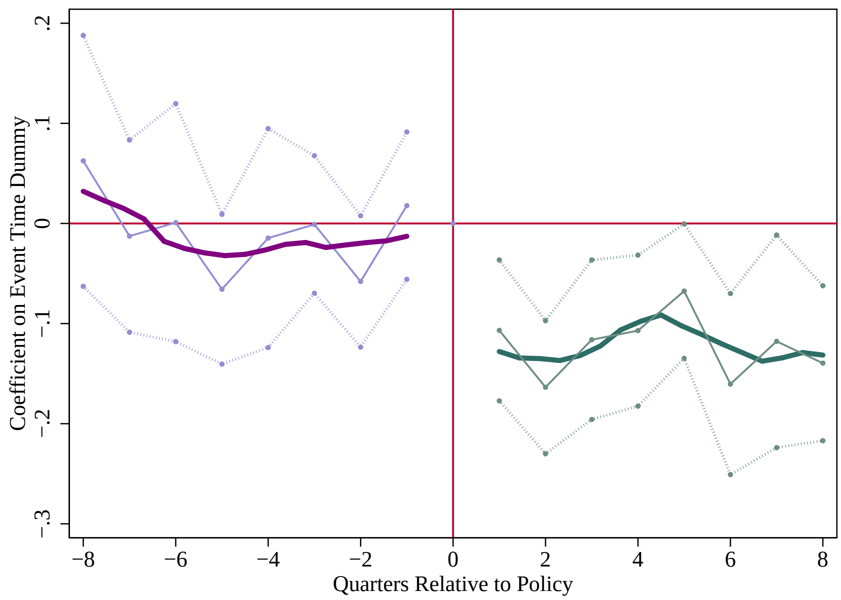
Doctors admit healthier bonus-eligible patients in participating hospitals in response to the program. Figure 8 displays the effect of the bonus policy on the average Charlson co-morbidity index of medical patients in event time, where the implementation of the policy is normalized to $t = 1$. The event 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.[18] After the policy is introduced, there is a clear drop in the average co-morbidity burden of Medicare patients with non-

[18] The model is fully saturated; hospitals which never participated are assigned an event time of -8.

surgical diagnoses.[19] These medical patients admitted at participating hospitals under the Gainsharing Demonstration became healthier, relative to the patients admitted by the same doctor at a nonparticipating hospital.

Figure 8: Healthier Patients Sent to Participating Hospitals

Charlson Comorbidity Index: Medical Patients



Notes: These are event study plots created by regressing the Charlson Comorbidity Index on a full set of event time indicators, as well as hospital, quarter, type (APR-DRG by SOI), and doctor fixed effects. Reported are the coefficients for event time, which plot the time path of the Charlson Comorbidity Index of patients admitted at participating hospitals, relative to non-participating hospitals, before and after the program went into effect. The dashed lines represent 95% confidence intervals, where standard errors are clustered at the hospital level. Time is normalized relative to the quarter that the hospital took up the bonus program.

The event time result in Figure 8 is presented in regression form in Table 3. The bonuses are associated with a decrease in the average Charlson co-morbidity index of medical patients of 0.09. 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 to 6 percent across seven OECD countries (Quan et al., 2011).[20] The bonuses are also associated with an increase in the probability that admitted medical patients have a Charlson score of zero; these patients can be thought of as the ones in the best overall health. The results are similar with and without type fixed effects (APR-DRG by SOI), implying that the change in latent health occurs within the types over which the bonuses are defined.

[19]As a placebo check, Figure A.2 shows that there is no analogous improvement in latent health for "near Medicare" patients (aged 50 to 64) treated at under the program. These patients have many of the same health problems as the Medicare population, but are too young to quality for Medicare coverage.

[20]Calculation assumes mortality decreases linearly between a Charlson score of 3 and 2.

The effect of the bonuses on latent health is much larger and more precisely measured for patients admitted under medical diagnoses, compared to surgical cases. Columns 5 and 6 show essentially no effect of the bonuses on the average latent health of surgical patients treated under the program—the coefficients are nearly an order of magnitude smaller than those for medical patients in columns 1 and 2. While the point estimates in columns 7 and 8 suggest that surgical patients treated in participating hospitals are more likely to have a Charlson score of zero, these estimates are not statistically significant.

Table 3: Effect of Bonuses on Latent Health: All Admitted Medicare Beneficiaries

	Medical Patients				Surgical Patients			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Charlson	Charlson	CCI=0	CCI=0	Charlson	Charlson	CCI=0	CCI=0
Policy	-0.091***	-0.092***	0.006*	0.007**	-0.010	-0.014	0.005	0.006
	(0.028)	(0.028)	(0.003)	(0.003)	(0.035)	(0.029)	(0.004)	(0.004)
Type FEs	-	x	-	x	-	x	-	x
Mean dep. var.	2.670	2.670	0.294	0.294	1.803	1.803	0.440	0.440
Clusters	74	74	74	74	74	74	74	74
Observations								
N	1,271,574	1,271,574	1,271,574	1,271,574	580,140	580,140	580,140	580,140

Notes: Quarter-by-year, doctor, and hospital fixed effects included in all regressions. Standard errors clustered at the hospital level. $*p < 0.1$, $** \ p < 0.05$, $*** \ p < 0.01$

The results in Table 3 suggest that doctors were able to game the bonuses much more effectively for medical patients than for surgical patients. An alternative interpretation, however, is that the the diagnoses included in the Charlson index are more closely tied to the costs of medical conditions than surgical ones. While it is difficult to completely rule out this alternative interpretation, Table A.4 uses a surgical risk index as an alternative measure of latent health, and again the bonuses have no effect. Finally, Table A.5 looks at the effect of the bonuses on a wide range of individual co-morbidities associated with more complex and expensive patients. Medical patients admitted at participating hospitals are less likely to have a wide range of individual chronic conditions, but again this is not the case for surgical patients.

There are two conceptual reasons why doctors treating surgical patients may be less responsive to the bonuses. First, there is likely to be less discretion over admission for surgical cases. Many surgical diagnoses have nearly 100 percent admission rates, which blunts the ability of doctors to manipulate this margin. Surgical cases tend to follow strict protocols, which may also prevent doctors from manipulating admission. Second, given that surgical patients are younger, healthier, and more homogenous, it may be harder to know which patients will be lower cost, and thus sorting surgical cases may be less lucrative than sorting medical cases.

In the conceptual framework, the increase in average latent health of patients treated under the bonuses is driven both by doctors newly admitting healthier patients, as well as doctors sorting healthier patients into participating hospitals. In the next two sections, I investigate each channel individually by alternately shutting down the admission margin and the sorting margin.

5.1.1 Admission

In order to isolate changes in which patients are admitted from changes in where patients are admitted, I look at two subsamples of the data where sorting is not possible: patients who were admitted through the emergency room, and patients whose doctors work in just one hospital. 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. Similarly, doctors cannot sort patients between hospitals if they only have admitting privileges in one. When looking at single hospital doctors, I use a difference-in-difference specification at the hospital level (equation 3 without doctor fixed effects), as there is no within-doctor policy variation in the post period.

Table 4: Effect of Bonuses on Latent Health: No Sorting

Panel A: Admitted through Emergency Room

	Medical Patients				Surgical Patients			
	(1) Charlson	(2) Charlson	(3) CCI=0	(4) CCI=0	(5) Charlson	(6) Charlson	(7) CCI=0	(8) CCI=0
Policy	-0.069**	-0.080***	0.004	0.006	-0.052	-0.039	0.012	0.010
	(0.032)	(0.024)	(0.003)	(0.003)	(0.047)	(0.037)	(0.008)	(0.006)
Type FEs	-	x	-	x	-	x	-	x
Mean dep. var.	2.620	2.620	0.302	0.302	1.881	1.881	0.446	0.446
Clusters	72	72	72	72	71	71	71	71
Observations								
N	881,970	881,970	881,970	881,970	154,730	154,730	154,730	154,730

Panel B: Single hospital doctors

	Medical Patients				Surgical Patients			
	(1) Charlson	(2) Charlson	(3) CCI=0	(4) CCI=0	(5) Charlson	(6) Charlson	(7) CCI=0	(8) CCI=0
Policy	-0.084**	-0.076**	0.009**	0.007	-0.034	-0.027	0.013	0.009
	(0.035)	(0.037)	(0.004)	(0.005)	(0.046)	(0.037)	(0.009)	(0.008)
Type FEs	-	x	-	x	-	x	-	x
Mean dep. var.	2.464	2.464	0.314	0.314	1.662	1.662	0.464	0.464
Clusters	58	58	58	58	58	58	58	58
Observations								
N	263,659	263,659	263,659	263,659	99,966	99,966	99,966	99,966

Notes: Quarter-by-year, doctor, and hospital fixed effects included in all regressions. Standard errors clustered at the hospital level. ∗p $<$ 0.1, ∗∗ p $<$ 0.05, ∗∗∗ p $<$ 0.01

Table 4 shows that even when doctors are unable to sort patients between hospitals, admitted Medicare patients with medical diagnoses are healthier when admitted under the program. Thus, doctors must be admitting healthier patients in response to the cost reduction bonuses. For medical patients, the magnitudes are similar across the two samples, though slightly smaller when compared to the main results in Table A.4. For surgical patients, the magnitudes are larger when the sorting margin is eliminated, though still the coefficients are still imprecisely estimated.

5.1.2 Sorting

In addition to changing which patients are admitted, do doctors also sort healthier patients into participating hospitals, conditional on patient type? I again attempt create a subsample that eliminates one margin of adjustment—in this case the admission margin—leaving sorting as the only way doctors could respond to the bonuses other than changing their practice style. In Table 5, I first exclude patients admitted through the emergency room, as there is no scope for moving these patients between hospitals. Panel A of Table 5 shows similar patterns to the main results in Table 3; bonuses are associated with better latent health for medical patients, but not surgical patients.

Table 5: Effect of Bonuses on Latent Health: Excluding Emergency Room

Panel A: Excluding Emergency Room

	Medical Patients				Surgical Patients			
	(1) Charlson	(2) Charlson	(3) CCI=0	(4) CCI=0	(5) Charlson	(6) Charlson	(7) CCI=0	(8) CCI=0
Policy	-0.128***	-0.094**	0.010*	0.006	0.014	0.001	0.001	0.003
	(0.047)	(0.043)	(0.005)	(0.004)	(0.035)	(0.030)	(0.004)	(0.004)
Type FEs	-	x	-	x	-	x	-	x
Mean dep. var.	2.785	2.785	0.275	0.275	1.775	1.775	0.439	0.439
Clusters	73	73	73	73	73	73	73	73
Observations								
N	389,604	389,604	389,604	389,604	425,410	425,410	425,410	425,410

Panel B: Excluding Emergency Room, High Admission DRGs

	Medical Patients				Surgical Patients			
	(1) Charlson	(2) Charlson	(3) CCI=0	(4) CCI=0	(5) Charlson	(6) Charlson	(7) CCI=0	(8) CCI=0
Policy	-0.129**	-0.101*	0.002	-0.000	0.002	0.002	0.002	0.002
	(0.063)	(0.058)	(0.006)	(0.006)	(0.030)	(0.026)	(0.008)	(0.007)
Type FEs	-	x	-	x	-	x	-	x
Mean dep. var.	3.323	3.323	0.212	0.212	1.365	1.365	0.517	0.517
Clusters	72	72	72	72	73	73	73	73
Observations								
N	143,832	143,832	143,832	143,832	161,002	161,002	161,002	161,002

Notes: Quarter-by-year, doctor, and hospital fixed effects included in all regressions. Standard errors clustered at the hospital level. $*p < 0.1$, $** p < 0.05$, $*** p < 0.01$

It is still possible, however, that some of the change in patient health in this sample is due to changes in which patients are admitted, rather than sorting. To focus more tightly on the sorting channel, I further narrow the sample to patients in diagnoses that are nearly always admitted.[21] In these diagnoses, any changes in the latent health must be due to doctors sorting patients between hospitals. While the results are less precisely estimated, the point estimates in Panel B of Table 5 suggest that doctors do respond to the bonuses by sorting healthier patients into participating hospitals. As in the previous section, however, the results are concentrated among doctors treating patients with medical diagnoses; there is no evidence that healthier surgical patients are more likely to end up in participating hospitals after the bonuses go into effect.

While I cannot calculate the relative importance of each channel, it appears that doctors treating non-surgical cases respond to the bonuses both by newly admitting healthier patients, and sorting patients across hospitals. Although the costs of doctors sorting patients between hospitals may be minimal (at least to patients and doctors), changing the pool of patients seen at each hospital may have important implications for hospital profitability. Furthermore, the fact that a healthier mix of patients is admitted when doctors are given bonuses for low cost admissions is worrying for two reasons. First, the healthier patients who are admitted under the program but not otherwise face substantial costs (both in time and money) from admission. Second, if capacity constraints bind, some sicker patients will not be admitted who would have been otherwise. For the sicker patients, the welfare effect of not being admitted depends on whether the admissions decisions were optimal in the pre-period, which is beyond the scope of this paper. However, if the admissions decisions were either optimal or too low in the absence of the bonuses, these patients could also be worse off on average from receiving less intensive treatment.

5.1.3 Up-Coding

Both the model and the empirical results assume that the assignment of a patient into a particular type (characterized by a diagnosis and a severity of illness pair) is 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.

[21] These are the top third of the sample of medical and surgical patients, based on the average admission rate in each APR-DRG in the Medicare population. The average admission rate in the medical high admission rate sample is 93.8. The average admission rate in the surgical high admission rate sample is 99.7.

The diagnosis groups used by Medicare (MS-DRGs) 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 MS-DRGs for diabetes with major complications (637), diabetes with complications (638), 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)).

While doctors may not be able to change the diagnosis group, it is possible that doctors could respond to the Gainsharing Demonstration by trying to move their patients into higher severity of illness bins.[22] Influencing the severity of illness (SOI) designation should theoretically be difficult (or at least indirect), 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 record more co-morbidities 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 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)	(5)	(6)
	Full Sample	ER Admission	High Admission	Full Sample	ER Admission	High Admission
Policy	0.006	0.012	-0.021	0.013	0.002	0.009
	(0.021)	(0.030)	(0.029)	(0.016)	(0.026)	(0.020)
Mean dep. var.	2.288	2.255	2.689	2.050	2.198	2.070
Clusters	74	72	72	74	71	73
Observations	1,271,574	881,970	143,832	580,140	154,730	161,002

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. $*p < 0.1$, $** \ p < 0.05$, $*** \ p < 0.01$

[22]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.

There is no association in the data between the bonuses and the average severity of illness within APR-DRGs, 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 fixed effects, rather than APR-DRG by severity of illness fixed effects. If up-coding occurred in response to the cost reduction bonuses, then within each APR-DRG, the patients admitted under the bonus policy should have a higher severity, on average. The introduction of the bonus program appears to have no effect on the average severity of illness, either in the full sample, or the subgroups used in the previous sections. 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 successfully 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 observed changes in latent health.

5.1.4 Placebo Tests

To confirm that the association between the bonuses and latent health is not spurious, I conduct two placebo tests. First, I hold fixed the true hospital participation in the Gainsharing Demonstration, but randomly assign start dates for the program, and repeat the main within-doctor regressions using randomly assigned dates. Second, I randomly assign New Jersey hospitals to participate in the program, holding constant both the number of participating hospitals and the timing of the program. I repeat the main within-doctor regressions using randomly assigned participation. The CDFs of the coefficients from both simulations (each based on 100 repetitions) are presented in Figure A.3.

The coefficients from the true regressions are represented by a red vertical line, and the 90th percentile by a red horizontal line. In both cases, the true coefficient is well above the 90th percentile. When hospitals are randomly assigned to participate, the true coefficient is much larger than any coefficient generated under the simulation. The results of these simulations suggest that it is extremely unlikely that the findings in section 5 are due to chance.

5.2 Effects of Bonuses on Costs and Quantity of Services Provided

Despite the fact that the program was explicitly designed to reduce costs, the bonuses have no effect on costs or resource use, conditional on patient health. Table 7 looks at the effect of the bonuses on measures of procedure use and total treatment costs. All regressions in Table 7 control for an array of measures of latent health: the Charlson index, the surgical risk factor index, and the individual chronic conditions included in Table A.5. Conditional on these latent health measures, the bonuses are not associated with significant decreases in length of stay, the use of 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).

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

Panel A: Medical Patients

	(1) Length of Stay	(2) CT Scan	(3) MRI	(4) Diag. Ultra	(5) Any Imaging	(6) Total Costs
Policy	-0.119	-0.001	-0.003	-0.001	-0.002	268.367
	(0.108)	(0.003)	(0.002)	(0.004)	(0.006)	(362.159)
Mean dep. var.	5.753	0.057	0.024	0.035	0.120	9,332.836
Clusters	74	74	74	74	74	58
Observations	1,271,574	1,271,574	1,271,574	1,271,574	1,271,574	1,047,516

Panel B: Surgical Patients

	(1) Length of Stay	(2) CT Scan	(3) MRI	(4) Diag. Ultra	(5) Any Imaging	(6) Total Costs
Policy	-0.083	-0.004	-0.001	0.002	0.007	796.911
	(0.150)	(0.002)	(0.001)	(0.006)	(0.011)	(575.311)
Mean dep. var.	6.893	0.037	0.013	0.058	0.191	18,226.473
Clusters	74	74	74	74	74	58
Observations	580,140	580,140	580,140	580,140	580,140	497,869

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. $*p < 0.1$, $** \ p < 0.05$, $*** \ p < 0.01$

While regressions reported in Table 7 control for a number of latent health measures, it is likely that I cannot completely control for differences in underlying health status. Given the fact that healthier patients were admitted into participating hospitals, 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 overall effect.

Despite not reducing costs overall, one might expect the bonuses to be more successful if doctors are not able to manipulate where patients are treated or whether they were admitted. However, there is no evidence that the bonuses reduced treatment costs for surgical patients, who displayed little evidence of manipulation on either the sorting or admissions margin. Table A.6 narrows the sample to single hospital doctors for whom the sorting margin is eliminated, and again finds no evidence that the bonuses are associated with lower costs. Surgeons and doctors working in a single hospital provide an interesting glimpse into what to expect if the program was scaled up. Given the null results in both groups it is unlikely that the bonuses would be more successful if all hospital participated, or even if both the sorting and admission margins were eliminated.

Table 8: Replicating the Initial Evaluation

	(1) Total Costs	(2) Total Costs	(3) Total Costs	(4) Total Costs	(5) Total Costs
policy	-415***	-415	-348	514	540
	(134)	(762)	(783)	(531)	(518)
Health controls	-	-	x	x	x
Comparison hospitals	-	-	-	x	x
Doctor fixed effects	-	-	-	-	x
Mean dep. var.	9,412	9,412	9,412	9,310	9,310
Clusters	.	11	11	58	58
Observations	204,812	204,812	204,812	862,748	862,748

Notes: Quarter-by-year, hospital, and diagnosis by severity of illness fixed effects included in all regressions. Health controls are the Charlson index, the surgical risk factor index, indicators for scoring zero on each index, indicators for the chronic conditions in Table A.5, and for age categories, sex, and race. Comparison hospitals are those where the Gainsharing policy was not implemented. $*p < 0.1$, $** p < 0.05$, $*** p < 0.01$

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 for medical patients. As in the initial study, I only include hospitals that eventually take up the initial demonstration, 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. Table 8 also demonstrates the extent to which low hanging fruit is not being utilized from a data analysis perspective. When evaluating the first wave of the Gainsharing Demonstration, simply including comparison hospitals would have given essentially the correct result. While it is possible to conclude that the bonus program lowered costs, this conclusion does not hold up to a more thorough investigation.

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 ex ante program exposure. 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. In particular, the exposure variable measures 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).

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 (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 5.2:

$$outcome_{dt} = \beta_0 + \beta_1 exposure_{dt} + \lambda_d + \lambda_t + \epsilon_{dt} \quad (4)$$

where λ_d and λ_t are doctor and quarter fixed effects. The coefficient of interest is β_1, which I interpret as the effect of program exposure on total costs and procedure use, net of sorting. Regressions are weighted by the number of admitted patients, though the results are nearly identical if no weights are used.

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 or admission rates. However, the exposure to the Gainsharing Demonstration is associated with higher average 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) Admitted Patients	(2) Adm. Rate	(3) Total Costs	(4) Avg. Costs	(5) CT Scans	(6) MRIs	(7) Diag. Ultra.	(8) Diag. Imaging	(9) Hosp. Days
Simulated share	-0.438	-0.005	3,690.032	249.402**	0.271	0.073	-0.082	0.283	-2.854
	(1.958)	(0.005)	(19,681.525)	(116.680)	(0.269)	(0.114)	(0.174)	(0.395)	(11.649)
Mean dep. var.	51.037	0.849	494,287.462	10,567.110	3.858	1.404	2.015	8.229	301.810
Clusters	3,466	3,466	3,323	3,323	3,466	3,466	3,466	3,466	3,466
Observations	87,568	87,568	80,276	80,276	87,568	87,568	87,568	87,568	87,568

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, diagnosic 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. Regressions and means are weighted by the number of admitted Medicare patients. $*p < 0.1$, $** p < 0.05$, $*** p < 0.01$

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 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 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. While external validity is always a concern when deciding whether to expand a pilot program, this paper calls into question the internal validity of pilot program evaluations in US health care.

In early 2018, the Trump administration announced plans for a new, voluntary bundled payment model, "Bundled Payments for Care Improvement Advanced". Under the new model, providers will receive a bonus if the spending for admitted Medicare patients with certain diagnoses is below a target. The new demonstration shares two key features with the New Jersey Gainsharing Demonstration: eligibility for the bonus is conditional on admission, and participation is voluntary. Medicare spends billions of dollars on pilot programs to learn how to reduce health care costs. The results of the Gainsharing Demonstration call into question the ability of pilot programs, which generally only cover small parts of tightly connected health care markets, to answer this question in the first place.

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.

Best, William R, Shukri F Khuri, Maureen Phelan, Kwan Hur, William G Henderson, John G Demakis, and Jennifer Daley, "Identifying patient preoperative risk factors and postoperative adverse events in administrative databases: results from the department of veterans affairs national surgical quality improvement program," 2002, *194* (3), 257–266.

Bilimoria, Karl Y, Yaoming Liu, Jennifer L Paruch, Lynn Zhou, Thomas E Kmiecik, Clifford Y Ko, and Mark E Cohen, "Development and Evaluation of the Universal ACS NSQIP Surgical Risk Calculator: A Decision Aid and Informed Consent Tool for Patients and Surgeons," 2013, *217* (5), 833–842.

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.

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.

Gupta, Atul, "IImpact of perfperform pay for hospitals: the readmissions reduction program," Technical Report 2016.

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," *Medicare 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.

Verma, Seema, "Medicare and Medicaid Need Innovation," 2017.

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.

8 Appendix: For Online Publication

8.1 Figures

Figure A.1: Physician Dashboard

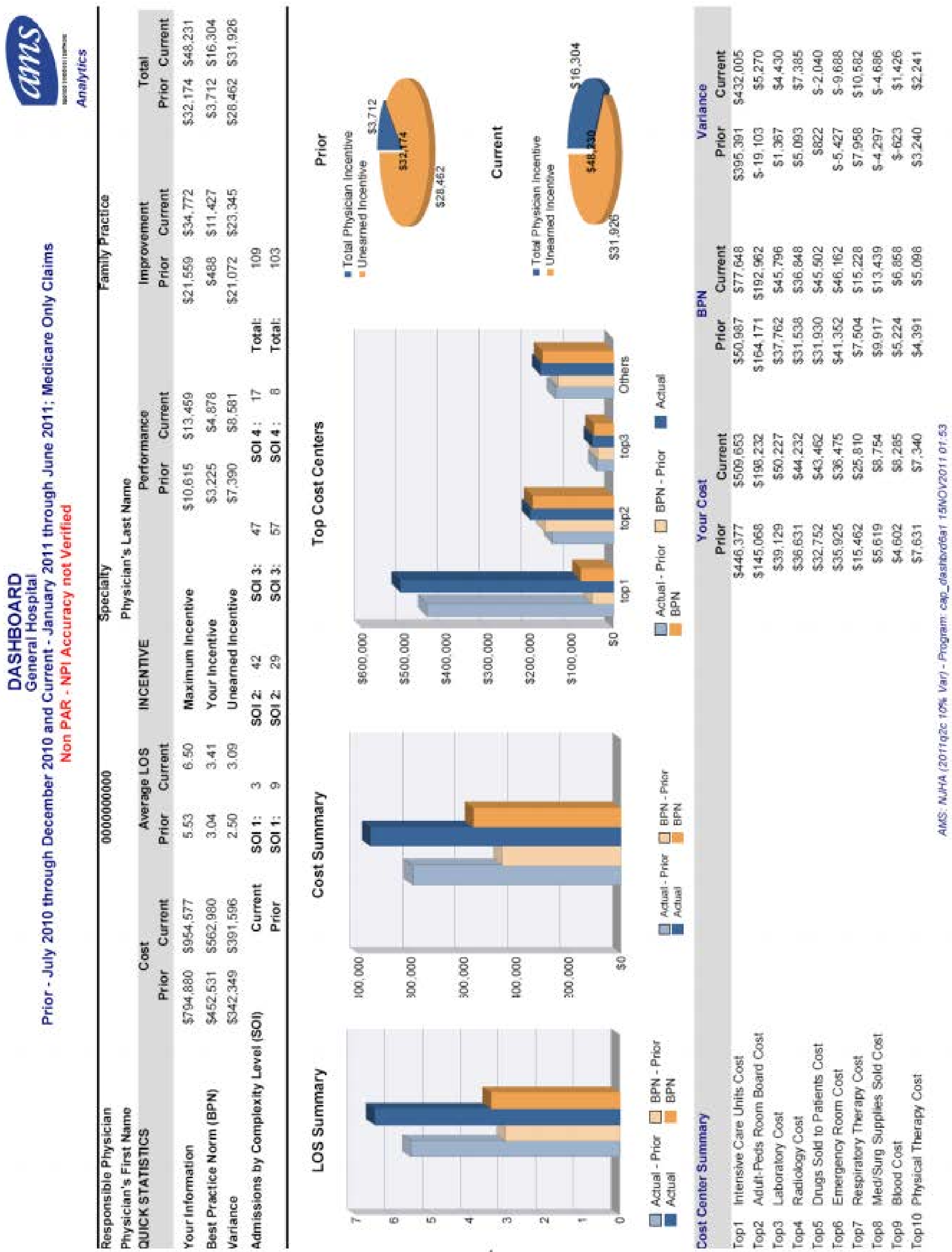
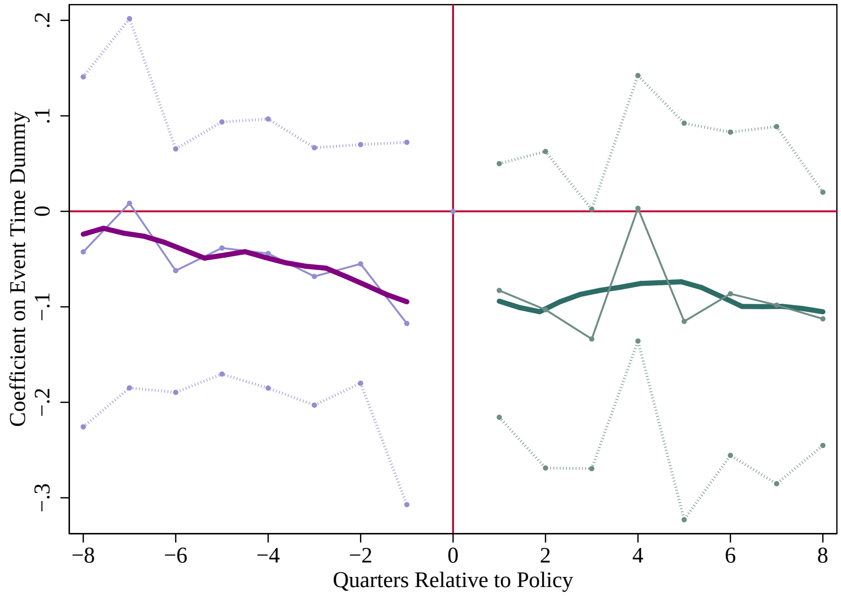


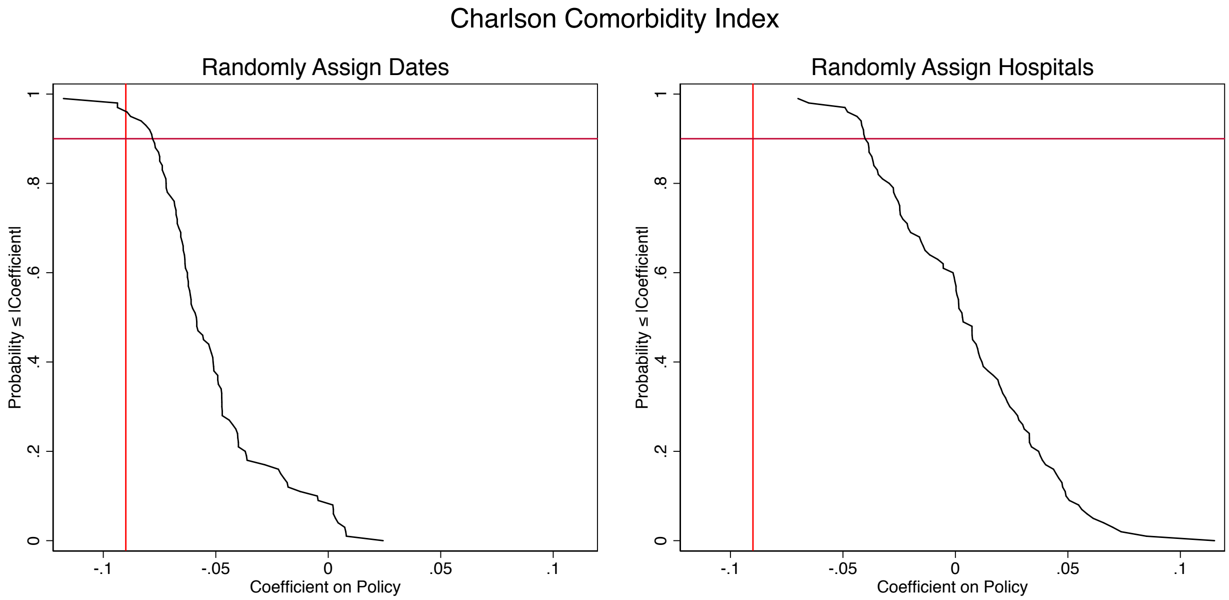
Figure A.2: Placebo Test: Near Medicare Patients (50-64)

Charlson Comorbidity Index: Medical Patients



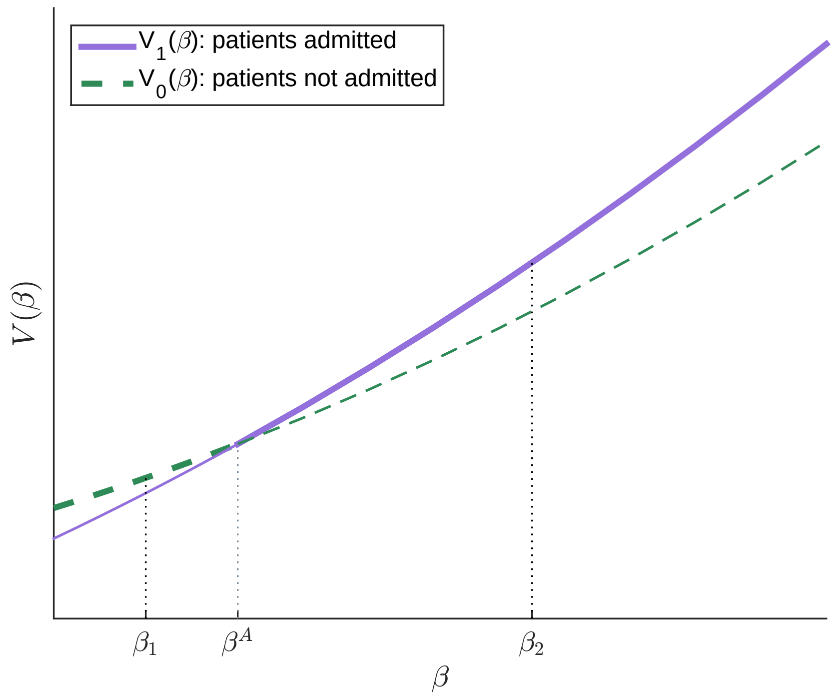
Notes: These are event study plots created by regressing the Charlson Comorbidity Index on a full set of event time indicators, as well as hospital, quarter, type (APR-DRG by SOI), and doctor fixed effects. Reported are the coefficients for event time, which plot the time path of the Charlson Comorbidity Index of patients admitted at participating hospitals, relative to non-participating hospitals, before and after the program went into effect. The dashed lines represent 95% confidence intervals, where standard errors are clustered at the hospital level. Time is normalized relative to the quarter that the hospital took up the bonus program.

Figure A.3: Placebo Test: Randomize Dates and Program Participation



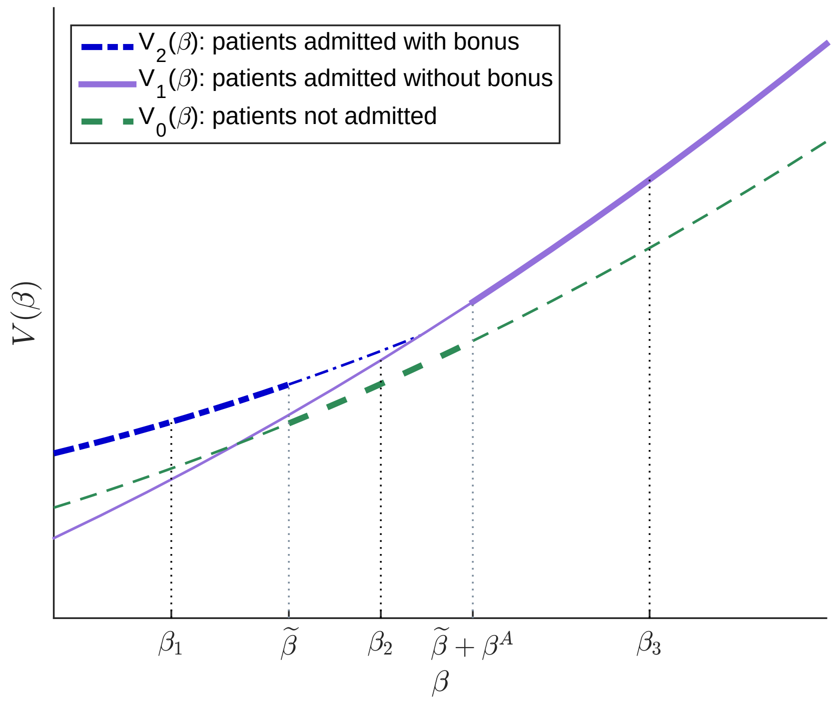
Each plot based on 100 repetitions. For randomizing dates, the starting dates of the Gainsharing Demonstration and the BPCI Model 1 program are randomly assigned (two random dates are chosen; the first is assigned to the Gainsharing Demonstration, and the second to the BPCI Model 1 program), with the hospitals participating held fixed. For randomizing participation, the participation into the the Gainsharing Demonstration and the BPCI Model 1 program was 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^{th} percentile.

Figure A.4: Doctor's Utility as a Function of β: without Bonuses



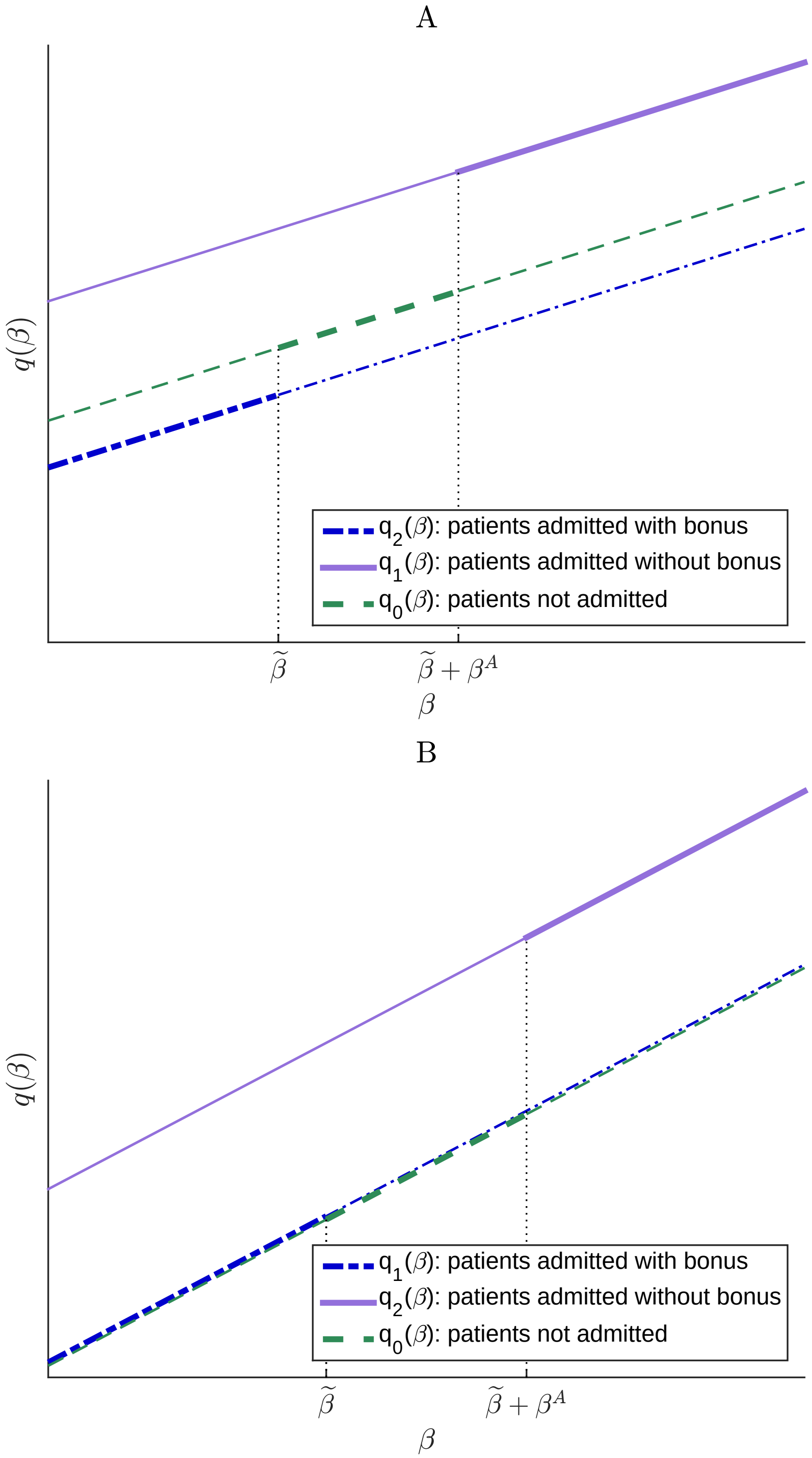
The bold line sections show the optimal decision rule as a function of β.

Figure A.5: Doctor's Utility as a Function of β: with Bonuses



The bold line sections show the optimal decision rule as a function of β.

Figure A.6: Optimal Quantity of Care as a Function of β: with Bonuses



The bold line sections show the quantity of care provided along the optimal decision rule. Figure A.6A shows the optimal quantity of care under one set of parameters; Figure A.6B shows the optimal quantity of care under another set of parameters.

8.2 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.

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

Medical Patients	Surgical Patients
194: heart failure	175: percutaneous cardiovascular procedures w/o ami
140: chronic obstructive pulmonary disease	302: knee joint replacement
720: septicemia & disseminated infections	301: hip joint replacement
139: other pneumonia	221: major small & large bowel procedures
201: cardiac arrhythmia & conduction disorders	173: other vascular procedures
463: kidney and urinary tract infections	308: hip & femur procedures for trauma except joint replacement
460: renal failure	171: perm cardiac pacemaker implant w/o AMI, heart failure or shock
383: cellulitis & other bacterial skin infections	174: percutaneous cardiovascular procedures w/ AMI
204: syncope & collapse	263: laparoscopic cholecystectomy
45: CVA & precerebral occulusion w/ infarct	24: extracranial vascular procedures
190: acute myocardial infarction	
198: angina pectoris & coronary atherosclerosis	
137: major respiratory infections & inflammations	
197: peripheral & other vascular disorders	
192: cardiac catheterization for ischemic heart disease	
253: other & unspecified gastrointestinal hemorrhage	
347: other back & neck disorders, fractures & injuries	

Notes: APR-DRGs from 3M's grouping software. These APR-DRGs make up 50 percent of admitted Medicare patients for medical and surgical visits, respectively.

Table A.3: Components of Surgical Risk Factor Index

Medicare Patients	Medical	Surgical	ICD-9-CM Codes
Ascites (all)	0.034	0.020	789.5
Ascites (30d prior)	0.003	0.001	
Congestive heart failure (all)	0.395	0.258	428.0
Cong. heart failure (30d prior)	0.049	0.24	
Dyspnea	0.001	0.000	786.0
Acute renal failure	0.312	0.199	584
Disseminated cancer	0.064	0.052	196-199
COPD	0.316	0.183	490-496
Ventilator dependent	0.223	0.121	518.8
Cardiac event	0.289	0.252	410-414
Diabetes	0.400	0.331	249-250
Hypertension	0.792	0.722	401-405

Notes: Admitted Medicare patients in general medical/surgical hospitals (2006-2013).

Table A.4: Effect of Bonuses on Surgical Risk Factor Index

	Medical Patients				Surgical Patients			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Surgical	Surgical	SRF=0	SRF=0	Surgical	Surgical	SRF=0	SRF=0
Policy	-0.005	-0.005	-0.004	-0.004	-0.002	-0.003	-0.002	-0.002
	(0.014)	(0.013)	(0.003)	(0.003)	(0.011)	(0.010)	(0.003)	(0.003)
Type FEs	-	x	-	x	-	x	-	x
Mean dep. var.	2.160	2.160	0.143	0.143	1.633	1.633	0.226	0.226
Clusters	74	74	74	74	74	74	74	74
Observations								
N	1,271,574	1,271,574	1,271,574	1,271,574	580,140	580,140	580,140	580,140

Notes: Quarter-by-year, doctor, and hospital fixed effects included in all regressions. Standard errors clustered at the hospital level. $*p < 0.1$, $** \; p < 0.05$, $*** \; p < 0.01$

Table A.5: Effect of Bonuses on Individual Chronic Conditions

Panel A: Surgical Patients

	(1) Chron. heart dis.	(2) Atherosclerosis	(3) Asthma	(4) Diabetes	(5) Arthritis	(6) Viral infection	(7) HIV/AIDS
Policy	-0.0061	-0.0064	-0.0028	-0.0077*	-0.0036	-0.0024	-0.0004*
	(0.0079)	(0.0053)	(0.0019)	(0.0041)	(0.0026)	(0.0015)	(0.0002)
Type FEs	x	x	x	x	x	x	x
Mean dep. var.	0.0909	0.2278	0.1125	0.2133	0.2072	0.0521	0.0019
Clusters	74	74	74	74	74	74	74
Observations							
N	1,271,574	1,271,574	1,271,574	1,271,574	1,271,574	1,271,574	1,271,574

Panel B: Surgical Patients

	(1) Chron. heart dis.	(2) Atherosclerosis	(3) Asthma	(4) Diabetes	(5) Arthritis	(6) Viral infection	(7) HIV/AIDS
Policy	-0.0026	0.0008	-0.0005	-0.0052	-0.0001	-0.0014	0.0001
	(0.0048)	(0.0033)	(0.0017)	(0.0038)	(0.0020)	(0.0012)	(0.0002)
Type FEs	x	x	x	x	x	x	x
Mean dep. var.	0.0692	0.1667	0.0656	0.1630	0.1450	0.0305	0.0008
Clusters	74	74	74	74	74	74	74
Observations							
N	580,140	580,140	580,140	580,140	580,140	580,140	580,140

Notes: Quarter-by-year, doctor, and hospital fixed effects included in all regressions. Standard errors clustered at the hospital level. $*p < 0.1$, $** \ p < 0.05$, $*** \ p < 0.01$

Table A.6: Effect of Program Costs and Procedure Use: Single Hospital Doctors

Panel A: Medical Patients

	(1) Length of Stay	(2) CT Scan	(3) MRI	(4) Diag. Ultra	(5) Any Imaging	(6) Total Costs
Policy	-0.031	0.003	0.001	0.000	0.002	876.505
	(0.172)	(0.005)	(0.003)	(0.003)	(0.008)	(664.989)
Mean dep. var.	6.007	0.027	0.018	0.033	0.088	9,537.632
Clusters	58	58	58	58	58	52
Observations	263,659	263,659	263,659	263,659	263,659	250,664

Panel B: Surgical Patients

	(1) Length of Stay	(2) CT Scan	(3) MRI	(4) Diag. Ultra	(5) Any Imaging	(6) Total Costs
Policy	-0.119	-0.004	-0.001	0.018	0.029*	1,491.843*
	(0.178)	(0.004)	(0.004)	(0.013)	(0.016)	(856.217)
Mean dep. var.	7.469	0.028	0.012	0.077	0.189	20,398.844
Clusters	58	58	58	58	58	53
Observations	99,966	99,966	99,966	99,966	99,966	95,065

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. $*p < 0.1$, $** p < 0.05$, $*** p < 0.01$

8.3 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^{th} percentile of costs:

$$0.1 * \left(\frac{1}{n} \sum_{i=1}^{n} \left(c_i - c_{25^{th}pctile} \right) \right) \tag{5}$$

where c_i is the cost of care for a patient in 2007 (before the program), and $c_{25^{th}pctile}$ is the 25^{th} 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.

$$\underbrace{\frac{1}{3} * MaxBonus * \frac{90th\ pctile - rate\ yr\ cost}{90th\ pctile - 25th\ pctile}}_{Performance\ Incentive} + \underbrace{\frac{2}{3} * MaxBonus * \frac{base\ yr\ cost - rate\ yr\ cost}{75th\ pctile - 25th\ pctile}}_{Improvement\ Incentive} \tag{6}$$

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.

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

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.

8.4 New Jersey Hospital and ED Discharge Data

The primary data are the New Jersey Uniform Billing Records, which cover all hospital discharges in New Jersey from 2006 to 2013. These data are compiled by the state from information that all general medical and surgical hospitals are required to submit about every individual encounter with a patient. Whether or not the visit resulted in admission, it appears in the raw data. These records include the hospital of treatment, admission and discharge dates, the patient's diagnosis, procedures performed, the type of insurance coverage, limited patient demographics, and total list charges. In addition, the raw file includes both identifying information on patients and medical license numbers of the attending doctors and surgeons, which I use to construct patient and doctor identifiers.

From this raw data, I construct a panel by matching patient records across visits by sex, date of birth, and first and last names.[23] This process creates a unique patient identifier, so that I can follow each patient over the sample period and across hospitals in the Uniform Billing Records. The matching algorithm does a good job catching slight misspellings without lumping together names that look different in a manual inspection of these data. This matching process was conducted in Trenton at the Department of Health and Senior Services, and these data were then de-identified on site. The ability to observe all visits to all hospitals is a particular strength of these data.

[23]The Levenshtein edit distance is used to match names, because of problems with typos and misspellings (stata command strgroup).

8.5 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 $\beta \sim U\left(\left[0,\bar{\beta}\right]\right)$. Doctors choose A, H, and q to maximize a weighted average of their profits and the patient's utility from receiving treatment,

$$\begin{aligned} \max_{A,H,q} U\left(A,H,q;\beta\right) &= \lambda \underbrace{\left[aq + \max\left\{\alpha_0 - \alpha_1 q, 0\right\} * \mathbb{1}\left\{H=1, A=1\right\}\right]}_{Doctor's\, profits} \\ &+ \left(1-\lambda\right) \underbrace{\left[\beta q + \left(\gamma q - C\right) * \mathbb{1}\left\{A=1\right\} - \frac{b}{2}q^2\right]}_{Patient's\, utility\, of\, treatment} \end{aligned} \tag{8}$$

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: $\max\left\{\alpha_0 - \alpha_1 q, 0\right\}$. 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 (γ), 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:

$$\max_{A,q} U(A,q;\beta) = \lambda \underbrace{[aq]}_{profit} + (1-\lambda)\underbrace{\left[\beta q + (\gamma q - C) * \mathbb{1}\{A=1\} - \frac{b}{2}q^2\right]}_{patient's\ utility\ from\ treatment}$$

$$= \max\left\{\underbrace{\lambda [aq^*(\beta)] + (1-\lambda)\left[(\beta+\gamma) q^*(\beta) - C - \frac{b}{2}q^*(\beta)^2\right]}_{\mathbf{V_1(\beta)}=U(q^*(\beta);\beta,A=1)},\right.$$

$$\left.\underbrace{\lambda [aq^*(\beta)] + (1-\lambda)\left[\beta q^*(\beta) - \frac{b}{2}q^*(\beta)^2\right]}_{\mathbf{V_0(\beta)}=U(q^*(\beta);\beta,A=0)}\right\}$$

Proposition 1: Under some parameter conditions, there exists a β^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 β.

The value function is:

$$V(\beta) = \max\left\{\underbrace{\lambda\left[aq_{(1)}\right] + (1-\lambda)\left[(\beta+\gamma) q_{(1)} - C - \frac{b}{2}q_{(1)}^2\right]}_{V_1(\beta)}, \underbrace{\lambda\left[aq_{(0)}\right] + (1-\lambda)\left[\beta q_{(0)} - \frac{b}{2}q_{(0)}^2\right]}_{V_0(\beta)}\right\} \tag{9}$$

Where from the first order conditions:

$$q_{(1)} = \frac{1}{b}\left[\left(\frac{\lambda}{1-\lambda}\right)(a_1) + \beta + \gamma\right] \tag{10}$$

and

$$q_{(0)} = \frac{1}{b}\left[\left(\frac{\lambda}{1-\lambda}\right)(a_1) + \beta\right] \tag{11}$$

The doctor's utility as a function of β 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 $\beta \in \left[\beta^A, \bar{\beta}\right]$, and does not admit patients with $\beta \in \left[0, \beta^A\right]$. 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.4). Thus, patients with $\beta \in \left[\beta^A, \bar{\beta}\right]$ are all admitted, and the rest are not admitted.

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

$$\lambda\left[aq_{(0)}\left(\beta^A\right)\right] + (1-\lambda)\left[\beta^A q_{(0)}\left(\beta^A\right) - \frac{b}{2}q_{(0)}\left(\beta^A\right)^2\right] \tag{12}$$

$$= \lambda\left[aq_{(1)}\left(\beta^A\right)\right] + (1-\lambda)\left[(\beta+\gamma)\,q_{(1)}\left(\beta^A\right) - C - \frac{b}{2}q_{(1)}\left(\beta^A\right)^2\right] \tag{13}$$

$$\Rightarrow \beta^A = \frac{2a\gamma\lambda + 2bC\lambda - 2bC - \gamma^2\lambda + \gamma^2}{2\gamma(\lambda-1)} \tag{14}$$

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

$$\max_{A,H,q} U(A,H,q;\beta) = \lambda \underbrace{\left[aq + \max\{\alpha_0 - \alpha_1 q, 0\} * \mathbb{1}\{H=1, A=1\}\right]}_{profit} + (1-\lambda) \underbrace{\left[\beta q + (\gamma q - C) * \mathbb{1}\{A=1\} - \frac{b}{2}q^2\right]}_{patient's\, utility\, from\, treatment}$$

$$= \max\left\{ \underbrace{\lambda\left[aq^*(\beta) + \alpha_0 - \alpha_1 q^*(\beta)\right] + (1-\lambda)\left[(\beta+\gamma)q^*(\beta) - C - \frac{b}{2}q^*(\beta)^2\right]}_{\mathbf{V_2(\beta)}=U(q^*(\beta);\beta,H=1,A=1)}, \underbrace{\lambda\left[aq^*(\beta)\right] + (1-\lambda)\left[(\beta+\gamma)q^*(\beta) - C - \frac{b}{2}q^*(\beta)^2\right]}_{\mathbf{V_1(\beta)}=U(q^*(\beta);\beta,H=0,A=1)}, \underbrace{\lambda\left[aq^*(\beta)\right] + (1-\lambda)\left[\beta q^*(\beta) - \frac{b}{2}q^*(\beta)^2\right]}_{\mathbf{V_0(\beta)}=U(q^*(\beta);\beta,A=0)} \right\}$$

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

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

Proof: Need to know the doctor's utility as a function of β.

The value function is

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

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:

$$q_{(2)} = \frac{1}{b}\left[\left(\frac{\lambda}{1-\lambda}\right)(a-\alpha_1)+\beta+\gamma\right] \tag{15}$$

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

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

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

$$q_{(0)} = \frac{1}{b}\left[\left(\frac{\lambda}{1-\lambda}\right)(a)+\beta\right] \tag{17}$$

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 β. They then admit the low β patients to the bonus hospital, do not admit the middle βs, and admit the highest βs to either hospital (randomizing over hospital such that they admit $\beta^{'}$ patients at both hospitals). Define the cut points as $\tilde{\beta}$ and $\widetilde{\beta}+\beta^A$. There is no patient β_2 with $\tilde{\beta} \leq \beta_2 \leq \widetilde{\beta}+\beta^A$ where the doctor would prefer to admit β_2 if it meant giving up admission for any patient $\beta_1 < \widetilde{\beta}$ or $\beta_3 > \widetilde{\beta}+\beta^A$; the doctor would be strictly worse off. This situation is depicted in Figure A.5.

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

$$\begin{aligned}\tilde{\beta}^{'} \text{ solves: } & \left(\lambda\left[aq_{(2)}+\alpha_0-\alpha_1 q_{(2)}\right]+(1-\lambda)\left[\left(\tilde{\beta}+\gamma\right)q_{(2)}-C-\tfrac{b}{2}q_{(2)}^2\right]\right) \\ & -\left(\lambda\left[aq_{(0)'}\right]+(1-\lambda)\left[\tilde{\beta}q_{(0)'}-\tfrac{b}{2}q_{(0)'}^2\right]\right) \\ & = \left(\lambda\left[aq_{(1)}\right]+(1-\lambda)\left[\left(\widetilde{\beta}+\beta^A+\gamma\right)q_{(1)}-C-\tfrac{b}{2}q_{(1)}^2\right]\right) \\ & -\left(\lambda\left[a_1 q_{(0)''}\right]+(1-\lambda)\left[\left(\widetilde{\beta}+\beta^A\right)q_{(0)''}-\tfrac{b_2}{2}q_{(0)''}^2\right]\right)\end{aligned} \tag{18}$$

Where:

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

$$\widetilde{\beta}' = \frac{\lambda(\lambda\alpha_1(2a-2\gamma-\alpha_1)+2b\alpha_0(\lambda-1)+2\gamma\alpha_1)+(\lambda-1)\left(2a\gamma\lambda+2bC\lambda-2bC-\gamma^2\lambda+\gamma^2\right)}{2\lambda\alpha_1(\lambda-1)} \tag{19}$$

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

Proposition 3: The direction of the change in q conditional on β from the pre- to the post-period for bonus-generating patients ($\beta \in \left[0, \widetilde{\beta}\right]$) 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 γ and α_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 $\beta \in \left[0, \widetilde{\beta}\right]$ 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 $\beta \in \left[0, \widetilde{\beta}\right]$ is greater when hospital 1 implements the bonus scheme than when neither hospital implements the bonus.

Figure A.6 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 patient over the "healthiest" of the sick patients they formerly admitted. The second condition holds as long as $U\left(q_{(2)}, 0\right) - U\left(q_{(0)}, 0\right) > U\left(q_{(1)}, \beta^A\right) - U\left(q_{(0)}, \beta^A\right)$. Finally, the doctors always want to admit the sickest patients: $U\left(q_{(2)}, \widetilde{\beta}\right) - U\left(q_{(0)}, \tilde{\beta}\right) < U\left(q_{(1)}, \bar{\beta}\right) - U\left(q_{(0)}, \bar{\beta}\right)$.