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

Job Market Paper, Latest Version:

http://scholar.princeton.edu/sites/default/files/dalexand/files/alexander_jmp.pdf

Diane Alexander*

November 12, 2015

Abstract

In an effort to find effective strategies for reducing healthcare costs, the Affordable Care Act has spent billions of dollars on pilot programs. In this paper, I study a Medicare pilot program in New Jersey where hospitals paid doctors bonuses for reducing the total costs of a given admission (a "bundled payment"). I identify the effects of the bonus by comparing the behavior of a given doctor who works at multiple hospitals, some of which participate in the program and others that do not. I find that doctors respond to the bonuses by changing the composition of admitted patients, and sorting healthier patients to participating hospitals–even conditional on the program's risk-adjustment criteria. Conditional on admission and patient health, however, doctors do not reduce costs or change procedure use. That doctors can identify low-cost patients in response to payment incentives is important for policy design going forward. In addition, the gaming behavior of doctors suggests that it is problematic to extrapolate the results of this and similar pilot programs to a nationwide reform.

*PhD Candidate, Princeton University. I thank Janet Currie, Ilyana Kuziemko, Hank Farber, Tom Vogl, Anne Case, Thomas Winberry, and participants in the Center for Health and Wellbeing and Public Finance workshops at Princeton University for 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.

1 Introduction

Lowering the growth in health care costs has long been a key 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) has earmarked billions of dollars for pilot programs.[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 fact that doctors in the US are paid separately for each service provided ("fee-for-service"), potentially encouraging them to perform unnecessary procedures. These programs offer the opportunity to study how much and on which margins doctors respond to altered payment schemes, an important open question in the literature. However, the small-scale nature of pilot programs leaves them susceptible to gaming and selection bias, making it unclear whether the information they generate is actually informative for a nationwide reform.

In this paper, I study a pilot program that paid doctors for reducing costs. I ask both how doctors responded to these changing incentives, and whether the information learned is useful for informing a national reform. In particular, I analyze the effects of the New Jersey Gainsharing Demonstration, under which hospitals paid doctors bonuses for reducing the total costs of treatment for each Medicare admission. The bonuses were tied to the total costs incurred during a hospital stay, and were designed to experiment with *bundled payments*—an incentive scheme where doctors are paid one fee for treating a patient, rather than separately for each service provided.

[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).

I find three main results of this program. First, doctors change which patients are admitted, as opposed to treated and then sent home. In particular, doctors are more likely to admit patients whose treatment generates high expected bonuses, and less likely to admit patients who generate low expected bonuses. Second, doctors often work in more than one hospital, and thus can change *where* patients are admitted. Doctors sort healthier patients into participating hospitals (even conditional on the program's risk-adjustment criteria), as these patients are cheaper to treat and therefore generate bigger bonuses. Third, conditional on admission and patient health, doctors do not change their procedure choice or otherwise lower treatment costs. Thus, the bonuses caused doctors to change their admission decisions—where and whether a patient was admitted—rather than reduce costs.

My empirical strategy leverages the fact that many doctors treat patients in more than one hospital. I measure the effect of the bonuses by comparing changes in a doctor's behavior at a participating hospital to the same doctor in a non-participating hospital. I worked at the New Jersey Department of Health to construct a unique dataset which allows me use a within-doctor specification. These data allow me to follow both both patients and doctors over time, and across all hospitals in New Jersey. The data include admission and discharge dates, all diagnoses and procedure codes, payer and patient demographic information, codes for doctors and patients, and list charges.

The bonuses given to doctors under the pilot program were designed to lower costs by reducing the incentive to provide treatments with low marginal benefits. In practice, patients were divided into types by diagnosis and severity of illness, and a maximum bonus was assigned to each group. Doctors were then paid a fraction of this maximum bonus after treating an eligible patient, depending on how close they got to pre-program cost benchmarks. Because of the limited scope of the pilot program, doctors could only receive a bonus if they treated an admitted Medicare patient at a participating hospital.

My first main result is that the cost-reduction bonuses change the patterns of admission across patient types. Admission is an important outcome, both in terms of costs and patient health; it is the difference between intense and prolonged monitoring by health care professionals, and being sent home after treatment.[2] I find that the cost reduction bonuses are associated with an increase the admission of patients in high-bonus types, relative to baseline. Conversely, patients in low-bonus types are less likely to be admitted. As capacity constraints and program rules limit the ability of doctors to increase overall admission rates, doctors instead reallocate admission across patients.

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

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

[2] Generally, admitted patients are assigned a hospital bed and spend at least one night at the hospital.

While sorting patients between hospitals does not necessarily change treatment costs, it can cause "naive" evaluations of the policy to be biased. In an evaluation of the first wave of the program, the Agency for Healthcare Research and Policy published an article reporting that the bonuses reduced costs per admission by eight percent (AHRQ, 2014). However, the evaluation only compared the costs of admitted patients at participating hospitals, before and after the program was implemented. My results suggest that a simple pre- versus post-comparison of admitted patients is misleading; it does not take into account changes in admission, nor does it not capture the fact that the composition of admitted patients at these hospitals changed in response to the program.

Related Literature My paper contributes to three main strands of literature. First, it is 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 Folk, 1996, Yip, 1998].[3] 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].[4] Unfortunately, studying the effect of payment structure

[3]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.

[4]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

on doctor decision-making is hampered both by data availability, and the fact that doctors practicing under different payment schemes may differ on unobservable characteristics. Therefore, how much and on what margins doctors will respond to payment reform policies remains an open question.

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

Third, problems and limitations of pilot programs have been widely studied in economics, particularly in development, education, and environmental economics [Duflo, 2004, Cullen et al., 2013, Allcott, 2015]. However, these lessons have generally not been applied to U.S. health care reform. The Centers for Medicare and Medicaid Services (CMS) has spent millions of dollars on pilot programs (or "demonstrations") since the 1960s, without considering whether partial equilibrium effects generated by such programs would hold in general equilibrium. Furthermore, the results of these pilot programs help direct the annual spending of Medicare, a 600 billion dollar 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.

changing treatment intensity and coding practices in response to DRG specific price changes.

Roadmap The rest of the paper is organized as follows. Section 2 describes the bonus program, and the specific incentives it created for doctors. To formalize the economics behind my empirical results, in Section 3 I develop a model of doctor decision-making. I show that the bonuses in the Gainsharing Demonstration clearly incentivize doctors to change who is admitted, and to sort patients between hospitals. The effects of the bonuses on resource use is ambiguous, and so remains a purely empirical question. 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.

2 Institutional Background

Before describing the details of the pilot program, I first lay out the institutional arrangements of doctors treating patients within a hospital setting. I describe how doctors and hospitals are paid, focusing on Medicare payment rules. Finally, I discuss the New Jersey Gainsharing Demonstration, the details of the bonus calculation formula, and its implementation.

2.1 How Doctors Treat Patients within Hospitals

When treating a patient in a hospital, doctors must decide both whether a patient should be admitted, and where to send the patient. Doctors can either decide to admit a patient and treat them, or they can treat the patient in the hospital and then discharge them. The technical definition of admission is simply that a doctor has written an order of admission. In practice, admitted patients generally stay at least over night and occupy a bed. Within all patient types in my data, there are both patients treated with and without being admitted. When considering whether to admit a patient, a doctor must weigh the benefits against the costs; admitted patients are intensely monitored, and receive more care. On the other hand, admission is costly for the patient, both in terms of time and money. In addition, admitted patients spend more time in the hospital, and thus face a higher risk of contracting a hospital acquired infections, which are often resistant to treatment.

When considering where to treat a patient, doctors are limited to choosing between hospitals where they have pre-arranged relationships. The exact employment relationship between doctors and the hospitals they work within is complicated, and varies from place to place. For the most part, however, doctors treating patients in hospitals are independent contractors, rather than hospital employees. These doctors have arrangements with hospitals which allow them to see patients there—so-called admitting or surgical privileges.[5] Doctors often have such privileges at more than one hospital (in my data, the average doctor is seen to treat patients at two different hospitals). That doctors tend to work at more than one hospital is key to my identification strategy, as I will compare the behavior of the same doctor that works both in a hospital that offers the bonuses and one that does not.

2.2 How Doctors and Hospitals are Paid

2.2.1 Status Quo

For the most part, doctors in the US are paid under the fee-for-service system, and traditional Medicare is no exception. Many argue that this fee-for-service system incentivizes additional care on the margin, and is thought to cause doctors to provide treatments with low or zero marginal benefits. Conversely, hospitals in the US are not paid according to each individual service performed. Instead, 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].[6] Medicare, which makes up approximately a third of the average hospital's net revenue, pays hospitals a fixed sum based on the patient's diagnosis (called diagnosis related groups, or

[5] Even emergency room doctors are usually not employed by the hospital, but are provided by separate business.

[6] 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].

"DRGs")—no matter how expensive the patient is to treat. Thus for Medicare patients, hospitals are incentivized to use fewer services.

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 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.[7] Medicare demonstration projects, however, are typically granted waivers to these statutes.

2.2.2 Under the Cost-Reduction Bonus Program

The New Jersey Gainsharing Demonstration was created by the New Jersey Hospital Association to reduce costs by aligning the incentives of doctors with those of hospitals. Under the program, doctors are still paid separately for each service provided, but are now also paid bonuses for lowering costs per visit. These bonuses are paid by the hospital to the doctor, and are supposed to reduce costs by lowering the use of unnecessary procedures. Doctors treating admitted patients at participating hospitals are eligible to receive one bonus per visit, where the maximum bonus they can receive depends on the patient's diagnosis and severity of illness.

[7] 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) [OIG, 1999].

When a hospital joins the Demonstration, doctors working in the hospital have the option to sign up for the program. While I do not have on which or how many doctors signed up, anecdotal evidence suggests take-up was high. There is no reason for a 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 cost reduction bonus under the Gainsharing Demonstration. For medical cases, this is the attending doctor, and for surgical cases, it is the surgeon.

The bonus a doctor receives through the Gainsharing Demonstration for treating an eligible (admitted and covered by Medicare) patient is calculated in three steps. First, patients are divided into types based on their diagnosis and how sick they are (for example, one type would be "hip joint replacement, severity of illness level two"), using 3M's All Patient Refined Diagnosis Related Group (APR-DRG) system.[8] 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 costs for that patient type relative to pre-program costs. A hypothetical bonus calculation example is presented in Figure 1. 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.

The maximum bonuses are calculated using 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 cost distribution 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

[8] 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 will discuss this more later in the paper, however, I believe the scope for up-coding is minimal in the Gainsharing Demonstration.

capture the fact that the same diagnosis (e.g. "peptic ulcer and gastritis") may be more or less serious depending on a patient's age and comorbidities. 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 1, and the distribution of maximum bonuses is shown in Figure 2.

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

The formulas used to calculate maximum bonuses are based on the idea that high cost variance within a diagnosis is bad, as it suggests that there are high cost patients who could be getting the same treatment as low cost patients. The bonuses are designed to make patients in diagnoses with high cost variance especially profitable. However, the association between cost variance and waste is just a theory. Alternatively, the high cost variance could be due to medical reasons, rather than doctor behavior. If 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 Implementation of Gainsharing Demonstration

The Gainsharing Demonstration took place in two waves, which both applied only to admitted Medicare patients.[9] The initial phase took place in twelve New Jersey hospitals from

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

July 1st, 2009 to July 1st, 2012. Eight of the original twelve hospitals opted to extend the program through March 31st, 2013. On April 1st, the program was renamed the BPCI Model 1 program, and was expanded to 23 hospitals (including six of the original twelve). My data go from 2006 through the end of 2013, so while I use the variation from the start of the BPCI Model 1 program, the bulk of my variation comes from the first wave of the program, and the 2012 extension.

The hospitals that formed the demonstration and its expansion appear to be similar to other New Jersey hospitals, on average.[10] Figure 3 shows that the participating hospitals are scattered around the state, and are thoroughly interspersed with non-participating hospitals (a complete list of participating hospitals can be found in the appendix). As can been seen in Table 2, 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. That hospitals with more Medicare patients are more like to participate is to be expected, as hospitals with the most Medicare patients also have the most to gain from a program designed to reduce the costs of treating this population. In addition, hospitals that participated in the first wave were more likely to receive a grade of A on a hospital quality report card. By the second wave, however, these differences disappear. 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 trajectories than non-participating hospitals. However, the identifying variation used in this paper is within doctor, which sidesteps many of the difficulties posed by differential trends at the hospital level.

[10] 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. As can be seen in Table 2, this appears to have been successful.

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. The basic setup of the model is first introduced; 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. I first 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

I consider one doctor treating a population of patients with mass one, where all patients are within a single diagnosis-severity of illness cell. I assume for now that the cell is exogenously defined, though I will later 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 weight placed on profit is given by λ. Doctors are paid a reimbursement rate, a, for each unit of care, q, provided to the patient. The reimbursement rate, 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.

Doctors do not only maximize profit—they also care about the patient's utility from treatment. A doctor's concern for their patient's welfare can be understood as altruism on behalf of their patients, or as the doctor acting to preserve their reputation. The utility a patient derives from medical care is:

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

where C is a fixed cost of admission. 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 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 not just on the amount of care provided, but also 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.[11][12]

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

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

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:

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

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

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

Now, I consider what happens when cost reduction bonuses 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 or patients admitted to the bonus hospital, $\beta' = \frac{\bar{\beta}-\beta^A}{2}$.[13] Doctors are constrained by the number or 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. Research has shown that patients essentially do whatever doctors tell them [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.

[13] 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

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

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

Whether or not there are bonuses, the admitted patients are always those with the largest (positive) difference between the utility a doctor receives from admitting them and not admitting them. Before the bonuses are introduced, this difference is largest for the sickest patient ($\beta = \bar{\beta}$), and is decreasing in 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.

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.

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 will 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. In the this section, I describe the data and strategy used to investigate each of these questions.

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.[14] I also create doctor identifiers using the recorded license numbers of doctors and surgeons. The final de-identified file includes codes for patients and doctors, allowing me to track them over time and across all hospitals in New Jersey. The ability to follow patients and doctors over time and across hospitals 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 annual information on hospitals from the American Hospital Association (AHA) annual survey, and Medicare's cost-to-charge ratio series.

In addition, I supplement the billing records with the bonus amounts that doctors could have received for treating each patient during the program. While I do not have access to the actual maximum bonuses used, I recreate them based on the formula provided by the New Jersey Hospital Association. The first step is to define the patient types, which is done by passing the billing records through 3M's All Patient Refined - Diagnosis Related Group (APR-DRG) software. For each record, the software creates a diagnosis group, a severity of illness (SOI) category, and designates the visit as medical or surgical. I then combine this information with cost data from the billing records and the bonus formula used in the Gainsharing Demonstration to reconstruct the maximum bonuses that a doctor could earn by treating each patient.

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

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

4.2 Estimation Sample

The main analysis sample consists of visits where an admitted patient was covered by Medicare, as these are the cases which can generate bonuses for doctors. I restrict the sample to patients seen in general medical and surgical hospitals that were open throughout the sample period. This restriction mainly excludes psychiatric and rehabilitation facilities, which were not targeted by the program. Visits to doctors with very few admitted patients were also dropped, as these doctors likely did not have enough patients to qualify for the bonus program. Finally, my main analysis omits visits where the patient is admitted through the emergency room, as emergency room doctors are not making a hospital choice decision. However, I will come back to patients admitted through the emergency room in a separate analysis at the end.

The main sample includes approximately $400,000$ each of medical and surgical visits, which were conducted by $3,474$ doctors in 73 hospitals. The patients are predominately white, with an average age of 75, and a high disease burden. The doctors worked at 2.2 hospitals on average, with 35% working in both a participating and non-participating hospital (see Table 3). The average maximum bonus a doctor could earn for a surgical patient was $697, and for a medical patient was $513. While few doctors take home the whole amount, even receiving half would be a windfall (for comparison, in 2012 Medicare paid doctors

$675.99 to repair a knee ligament [Smith, 2012]).[15]

4.3 Empirical Strategy

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

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

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

where $Policy_{dht}$ 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_{idht} vary slightly by specification, but in general contain age, race, and sex, and dummies for patient type. Hospital, quarter, and doctor fixed effects are also included in all regressions (λ_t, λ_h, and λ_d). When I look at the effect of the program on which patients are admitted to the hospital, I further want to know whether patients are differentially affected, depending on the size of the maximum bonus attached to their type. In this case, I interact the policy variable with the bonus size:

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

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

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

5 Results

All results are presented separately for medical and surgical patients, as there are important differences between these groups. For one, surgical patients have higher admission rates (with a few APR-DRGs at 100% 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 I look at length of stay and diagnostics. In addition, the consequences for the patient of changing admission and the quantity of services may be different for medical and surgical cases, which could lead to distinct program effects across the two groups.

5.1 The Admission Margin

5.1.1 Outcomes

When considering the effect of the bonuses on admission, I would ideally know whether each admitted patient who was treated under the bonus program would have been admitted if the bonus program did not exist. While I can never know this counterfactual, I estimate it from my data. First, I take data from before the program started, which includes both patients that were admitted and those that were not admitted. Using this pre-program data, I regress admission on a large set of observable characteristics. Next, I use the results of this regression to predict whether or not each patient in my main sample (all of whom are actually admitted) would have been admitted in the pre-period. I call this variable the baseline admission probability, as it answers the question: would a patient with these characteristics have been admitted in the pre-period? If the introduction of the bonus program is associated with a decrease in the baseline admission probability of admitted patients, it is consistent with the bonuses inducing doctors to admit some patients who would not have been otherwise.

5.1.2 Results

Table 4 shows the effect of the bonuses on the baseline admission probability. On average, the introduction of the bonuses had no effect, which is consistent both with program rules and binding capacity constraints. While the bonuses are associated with a small decrease in the average baseline admission probability of both medical and surgical patients, it is not statistically significant (seen in columns 1 and 3 of Table 4, for medical and surgical patients respectively). This null result, however, could conceal important heterogeneity with respect to the size of the bonus. While doctors were barred from admitting all of their patients in response to the bonuses, they could change which patients were admitted. As columns 2 and 4 of 4 show, it is important to allow the effect of the policy to differ by the size of the maximum bonus.

For patients in high bonus types, the introduction of the bonuses is associated with a significant decrease in the baseline admission probability—implying that some patients in high bonus types would not have been admitted in the absence of the program. For these patients, the effect of the policy is the sum of the coefficients on the policy and its interaction with high bonus. The policy is associated with a statistically significant decrease of 0.048 percentage points for medical patients (Table 4, column 2) and 0.017 for surgical patients (Table 4, column 4), implying that doctors are more likely to admit patients in high bonus types when the policy is in effect.

Conversely, the bonuses are associated with a small increase in the baseline admission probability of low bonus patients, suggesting that some low bonus patients were not admitted, that would have been admitted in the absence of the program. The effect of the policy on the baseline admission of low bonus patients is simply the coefficient on the policy variable in columns 2 and 4 of Table 4, which is statistically significantly positive in both cases. Low bonus medical patients saw an 0.028 percentage point increase (3% of mean) in the baseline admission rates, and low bonus surgical patients saw an increase of 0.014 percentage points (1.5% of mean). The increase in the baseline admission probability for low bonus patients is consistent with doctors making room for more high bonus patients by not admitting some patients with low bonus types.

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

5.2 The Hospital Sorting Margin

5.2.1 Outcomes

Do doctors sort healthier patients into participating hospitals, conditional on patient type? To answer this question, I need measures of the latent health of patients that are known to the doctor (or at least correlated with information known to the doctor), but not to the bonus formula. My strategy will be to exploit the time-series dimension of the data, and use information from past patient visits.

My preferred measure of latent patient health is the Charlson Co-morbidity Index, which is computed based on 17 conditions weighted by the associated risk of death.[16] This index

[16] The Charlson Co-morbidity Index is a weighted sum over the following 17 conditions, where weights are in parentheses: acute myocardial infarction (1), congestive heart failure (1), peripheral vascular disease (1), cerebrovascular disease (1), dementia (1), chronic pulmonary disease (1), rheumatologic disease (connective

has been widely validated, and has shown to be strongly predictive of hospital resource utilization [Charlson et al., 2008]. In order to measure a patient's latent health (rather than the acute event that brought them to the hospital), I construct a "leave-out" version of the Charlson Co-morbidity Index. The leave-out index is constructed using data on each patient's previous visits to the hospital, excluding the current visit. Excluding the current visit helps me identify sorting, as diagnoses from the current visit are taken into account in when determining the severity of illness level, but diagnoses from previous visits are not.

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

5.2.2 Results

Taking the Charlson Co-morbidity Index as my preferred measure of patient health, I find that doctors admit healthier patients to bonus eligible hospitals in response to the program. Figure 7 displays the effect of the bonus policy on the average Charlson Co-morbidity Index of patients in event time (for medical patients), where the implementation of the policy is normalized to $t = 1$. After the policy comes into effect, there is a clear drop in the average co-morbidity burden of patients.

The sorting result depicted in Figure 7 is presented in regression form in column 1 of

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)

Table 5. The bonuses are associated with a decrease in the average Charlson Co-morbidity Index of medical patients of 0.11. Medical patients admitted in hospitals offering bonuses are significantly healthier. The patient sorting channel is quantitively substantial; using estimates from Quan et al. [2011], a decrease of 0.11 is associated with decreasing in-hospital mortality by $5 - 18\%$. In addition, doctors have shown that the Charlson Co-morbidity Index is strongly predictive of resource utilization [Charlson et al., 2008], suggesting that these healthier patients are indeed cheaper to treat. The Charlson Index of surgical patients does not respond to the program, probably because the diagnoses included in the index are more closely tied to medical conditions than surgical ones (Table A.1 lists the top 15 primary diagnoses for medical and surgical patients in my sample).

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

Exactly as suggested by the model, Table 5 shows that conditional on type, patients admitted by doctors at participating hospitals are healthier than patients admitted by the same doctors at non-participating hospitals. The mechanism is straightforward—conditional on diagnosis, healthier patients are cheaper to treat, and cheaper patients earn higher bonuses. By sorting patients across hospitals, doctors can earn a bonus without changing changing treatment conditional on admission and patient health.

One might wonder whether the results in Table 5 are driven by changes in the admission margin—patients that are only admitted because of the program are likely healthier on average—rather than pure sorting. Table 6 addresses this concern by repeating the analysis on the subsample of patients in diagnoses that are nearly always admitted, as it is unlikely that the composition of this sample would be affected by changing admission thresholds. While the results are less precisely estimated, the magnitudes are similar. Thus, it appears that doctors are responding to the bonuses by both changing admission thresholds and sorting patients across hospitals.

Not only do doctors sort healthier patients into participating hospitals, the hospitals that participated in the first wave are also higher quality—or at least score better on one particular hospital quality report. In general, if the participants in pilot programs are higher quality, another unintended consequence of programs which incentivize doctors and hospitals to cream-skim patients could be a worse match between patient health and treatment quality.

5.2.3 Up-Coding

Both the model and the sorting results assume that the diagnosis and severity of illness margins are unaffected by the Gainsharing Demonstration. One might be worried about this assumption, as during the 1990s many hospitals were accused of up-coding—exaggerating a patient's diagnosis to extract a higher reimbursement from Medicare. Silverman and Skinner [2004] found, for example, that between 1989 and 1996, the percentage point share of the most generous diagnosis groups (DRGs) for pneumonia and respiratory infections rose precipitously. One reason the diagnosis groups used by Medicare are particularly susceptible to up-coding is that there are often multiple DRGs for each diagnosis, where the most severe version pays a much higher amount. For example, there are separate Medicare diagnosis groups (MS-DRGs) for diabetes with complications (638), diabetes with major complications (637), and diabetes without complications (639), where the more severe codes are reimbursed at higher rates. In the diagnosis groups used for the bonus calculations, however, this feature is lacking. In order to upcode at the diagnosis level doctors would have to change the diagnosis conceptually, which seems unlikely (e.g., changing a diagnosis from "diabetes" (APR-DRG 420) to "malnutrition, failure to thrive, and other nutritional disorders" (APR-DRG 421)).

It is also possible that doctors could respond to the Gainsharing Demonstration by trying to move their patients into higher severity of illness bins.[17] Influencing the severity of illness (SOI) designation should be difficult, as it is imputed by software and not recorded by the doctor. The only way doctors can affect the severity of illness is to change which secondary diagnoses are recorded on a patient's chart. While the link between any one co-morbidity and the designation generated by the software is not clear, adding additional diagnoses to all patients could lead to higher average SOI designations. If doctors managed to inflate the severity of illness of all patients in response to the program, the average "true sickness level" of the patients in each cell would decrease—the sickest patients in the first severity bin would be shifted into the next bin, and so on up the chain. Up-coding, therefore, could generate similar patterns in the data as sorting.

There is no association between the bonuses and the average severity of illness, however, suggesting that up-coding is not a concern in this context (see Table 7). The regressions reported in Table 7 use the same empirical strategy outlined in equation 3, but with diagnosis rather than diagnosis by severity of illness fixed effects. One interpretation of this null result is that hospitals are able to closely monitor the coding practices of their doctors. The proximity of the payor (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 upcode patients when working within the walls of the entity making the payment, in comparison to a distant third party such as Medicare.

[17] 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.

5.3 The Quantity of Services Margin

5.3.1 Outcomes

To examine whether the bonuses changed procedure use or costs, I again use a variety of measures. The first two are summary measures of resource use: how long a patient stays in the hospital, and the total costs incurred. 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 costs incurred during a visit are estimated using the total list charges reported in the discharge data, deflating them by Medicare's hospital-year level cost-to-charge ratio, and then converting them to real 2010 dollars. Deflating the list charges is an important step, as list charges are closer to bargaining tools than a measure of the costs to the hospital of providing a service. The cost-to-charge ratio files, however, are explicitly designed to translate list charges into an estimate of the resource cost of inpatient care.

In addition to summary measures of resource use, I look specifically at the use of diagnostic imaging to proxy for the use of unnecessary procedures. While it is difficult to pinpoint any specific test as unnecessary, there is widespread agreement that diagnostic imaging is overused [Hillman and Goldsmith, 2010, Abaluck et al., 2015].[18] Thus, 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.[19]

[18]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]

[19]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].

5.3.2 Results

As shown in Table 8, the bonus policy has no significant effect on costs or resource use, conditional on admission and patient health—despite the fact that the program was designed to reduce costs. The program is not associated with significant decreases in length of stay, diagnostic imaging tests, or costs. Even taking the point estimates at face value, the magnitudes are small, and the signs are not even consistently negative. I also find 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].

Furthermore, while the regressions reported in Table 8 include controls for all latent health measures examined in the previous section, it is likely that I cannot completely control for differences in underlying health status. Given the fact that healthier patients were sorted into participating hospitals sorting, these patients may have required fewer resources from the start. Thus, the small decreases reported in some measures in Table 8 should be considered an upper bound on the true effect.

Given these results, how did the naive evaluation conclude that the program succeeded in decreasing costs? In Table 9, I attempt to replicate the naive evaluation of the first wave of the program. In column 1, I only include hospitals that eventually take up the program, with no health controls 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 renders the decrease in costs insignificant. In columns 3-5, I add comparison hospitals, health controls, and doctor fixed effects, and show that the sign flips from negative to positive. An incomplete analysis can conclude the program lowered costs, but this conclusion does not hold in to a more thorough investigation.

6 Extensions and Robustness Exercises

6.1 Placebo tests

To confirm that my results are not spurious, I conduct two placebo tests. First, I randomly assign New Jersey hospitals to participate in the program, holding constant both the number of participating hospitals and the timing. I repeat my main regressions using randomly assigned participation, and plot the cdfs of the resulting coefficients (based on 100 repetitions) in Figure A.5. Second, I hold fixed the true hospital participation, but randomly assign start dates for the program, and again repeat all of my main regressions. The cdfs of the coefficients from this second simulation (again based on 100 repetitions) are presented in Figure A.5. The coefficients from the true regressions are represented by a red vertical line, and the 90th percentile by a red horizontal line. In nearly all cases, the true coefficients are well above the 90th percentile—in a few cases, the true coefficients are larger than any coefficient generated under the simulation. The results of the simulations suggest that it is extremely unlikely that my findings are due to chance.

6.2 Other patient groups

6.2.1 Emergency Room Medicare Patients

While patients admitted through the emergency room were excluded from the main analysis, they provide both a useful placebo test for sorting, and another group of patients among which costs might have been reduced. Patients admitted through the emergency room (ER) cannot be sorted in response to the bonuses, as emergency room doctors cannot send a healthier than average ER patient to a different hospital. In Table A.4, I show that the sorting results pass this placebo test—the introduction of the bonuses is not associated with a change in underlying patient health for patients admitted through the emergency room.

In addition, the doctor who decides to admit a patient from the emergency room is not necessarily the doctor who would receive the bonus. While it varies from hospital to hospital, an attending in the emergency department generally decides to admit a patient, and then a different doctor is responsible for the patient after admission. The way the Gainsharing Program is designed, a bonus from treating an ER patient would most likely go to the second doctor, and thus the bonuses should not influence admission. This intuition is consistent with the data, as there is no effect of the bonus program on baseline admission for patients admitted through the ER (see Table A.3).

Exactly because there is no sorting margin and a limited admission margin, however, the bonuses could have a large affects on costs and procedure use for patients admitted through the emergency room. In Table A.5, however, I show that the effects of the bonuses on costs and quantity measures for these patients are similar to those in the main sample. While the point estimates for the effect of the policy on length of stay is larger for emergency room patients than in the main sample, they are not statistically significant.

6.2.2 Patients Not Covered by Medicare

Despite the fact that only Medicare patients are included in the bonus program, it is possible that the program could spill over into the treatment of other patients—particularly "near Medicare" patients. "Near Medicare" patients, aged 50 to 64, have many of the same health problems as the Medicare population but are too young to quality for Medicare coverage.[20] The bonus program could spill over into the care of near Medicare by crowding out care for near Medicare patients, as their treatment cannot generate a bonus.

I find no evidence that the program caused doctors to crowd out younger patients (see Tables A.6 through A.8). If anything, the results for near Medicare patients point in the same direction as the main results, though are roughly 80% smaller in magnitude. While any treatment spillovers are very small, these results are consistent with it being difficult for doctors to perfectly target Medicare patients—that some patients who are not eligible for the program still end up being treated like Medicare patients.

[20] Medicare patients who are younger than 65 are excluded from this analysis. Patients younger than 65 can be enrolled in Medicare, for example those on disability or with end-stage renal disease.

6.3 Alternative Strategy: Simulated Share Treated

The doctor-level difference-in-difference specification paints a clear picture: 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. Policymakers, however, may additionally want to know what effect the bonuses had on total costs, rather than relative costs. I use an alternative identification strategy to examine the effect of the program on total costs incurred by doctors, which asks whether ex-ante program exposure affects doctor-level resource use.

In order to isolate the effect of the bonuses on total costs net of sorting, I create a doctor-level exposure variable: the simulated share of a doctor's case load that is treated. Specifically, the simulated share is the fraction of a doctor's patients that would be affected by the program if the distribution of patients across hospitals was fixed in the pre-period. The simulated share is zero in the pre-period, and then rises to the fraction of a doctor's caseload treated at participating hospitals before the program started (2006-2008). The simulated share treated captures the fact that some doctors only admit patients to participating hospitals (their simulated share goes from zero to one), others are not exposed at all, and many doctors are in between. By construction, the simulated share reflects only ex ante exposure, and will not be affected by doctors sorting patients in response to the bonus program.

I collapse the data to the doctor-quarter level, and regress the simulated share treated regressed on cost and quantity measures:

$$outcome_{dt} = \beta_0 + \beta_1 share\ treated_{dt} + \lambda_d + \lambda_t + \epsilon_{dt} \tag{5}$$

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, net of sorting.

There is no evidence of any cost-saving response to the Gainsharing Demonstration in response to program exposure, backing up the main results. As shown in column 1 of Table 10, if anything, exposure to the Gainsharing Demonstration leads to an increase in costs (statistically significant at the 10 percent level). This result concords with the main results of the paper, and could be a result of lowering the threshold for admission (the effect of program exposure on the number of admitted patients is positive, but insignificant). There is no affect of the bonuses on the use of CT scans or MRIs, and a small positive affect on the use of diagnostic ultrasounds—also suggestive of a change in the composition of admission towards healthier patients. Overall, 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. That doctors are able to identify low-cost patients even within risk-adjusted groups has ramifications beyond the success or failure of this particular program—it is crucial information for policy design. Furthermore, for pilot programs to be informative, they must contain the same incentives as if the program was expanded nationwide. If doctors are able to sort patients in and out of pilot programs in response to changing incentives, this assumption fails.

The results of this paper are applicable to many active programs experimenting with doctor incentives. For example, the Affordable Care Act has heavily promoted Accountable Care Organizations (ACO) as an improvement on the fee-for-service system. ACOs are networks of doctors and hospitals who share responsibility for providing care to patients in an attempt to limit duplication of services and other unnecessary spending. Like Gainsharing, ACOs create incentives for savings by offering bonuses when providers reduce costs. Also like Gainsharing, ACOs only cover some patients, and doctors can influence which patients are covered by the program. In future work, I will investigate whether doctors participating in ACOs exhibit similar sorting behavior to those in the New Jersey Gainsharing Demonstration.

The Department of Health and Human Services (HHS) recently committed to moving 50 percent of Medicare payments (and 90 percent of fee-for-service Medicare payments) into "value-based" payment models by 2018. Many alternative payment models will be used to meet the 2018 goal—such as "Accountable Care Organizations (ACOs), advanced primary care medical home models, new models of bundling payments for episodes of care, and integrated care demonstrations for beneficiaries that are Medicare-Medicaid enrollees" [HHS, 2015]. The early evidence on whether these alternative payment models seem to work comes largely from pilot programs and other partial equilibrium settings. However, my paper calls into question the ability of pilot programs to provide this type of evidence, due to the sorting and selection behavior of providers. Furthermore, incomplete expansion of these payment models continues to leave room for providers to act strategically in ways that undermine the reform's success.

8 Figures

Figure 1: Hypothetical Bonus Calculation

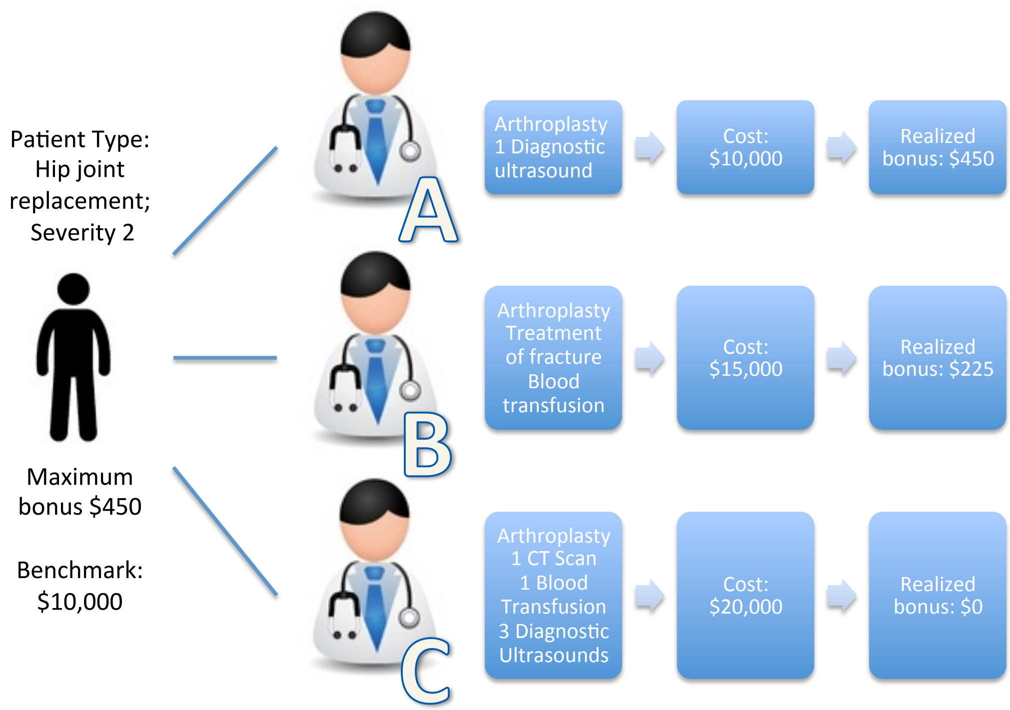
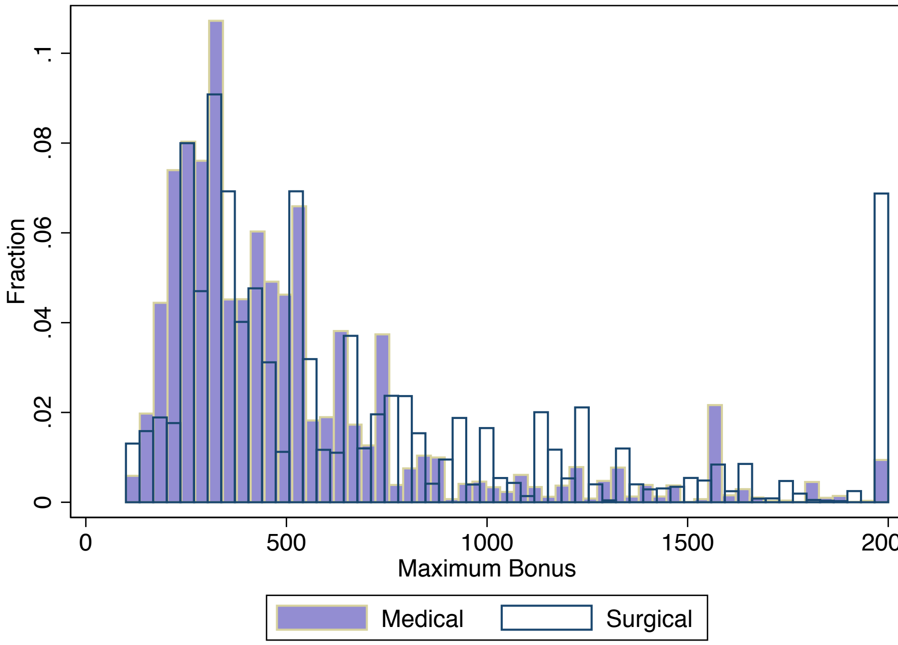
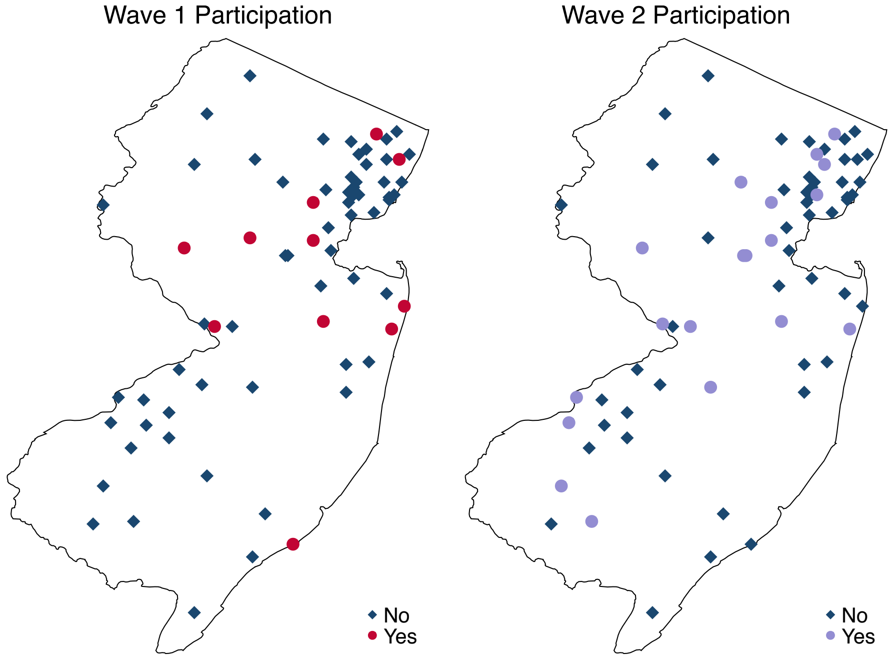


Figure 2: Distribution of Maximum Bonuses



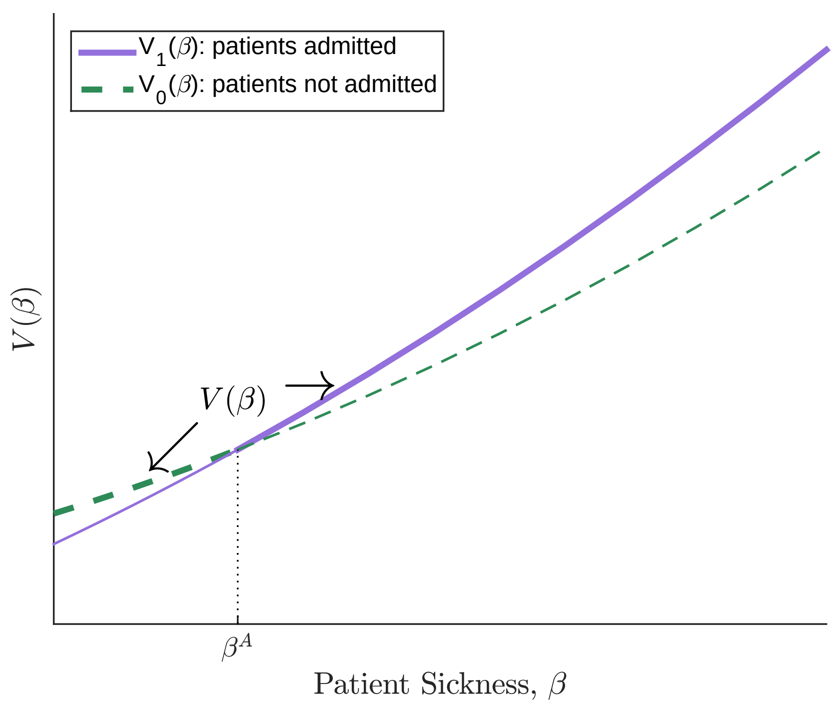
Notes: Each observation is a Medicare beneficiary's inpatient visit to a general medical/surgical hospital in New Jersey from 2006-2013, excluding visits that went through the emergency room.

Figure 3: Hospital Locations



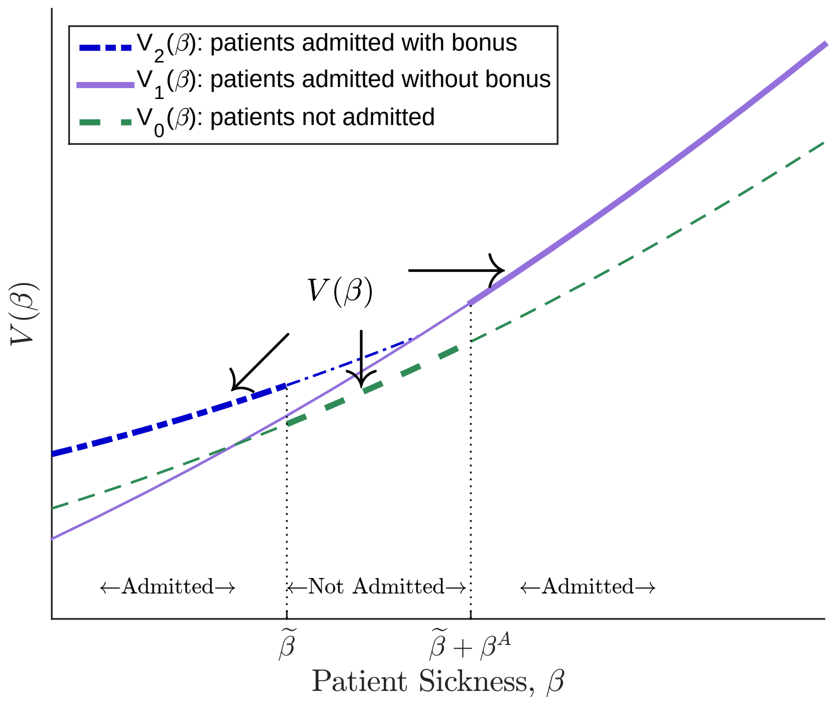
Notes: Blue diamonds are hospitals that never participated, red circles are hospitals that took up the bonuses in the first wave, and purple circles are hospitals that joined in the second wave.

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



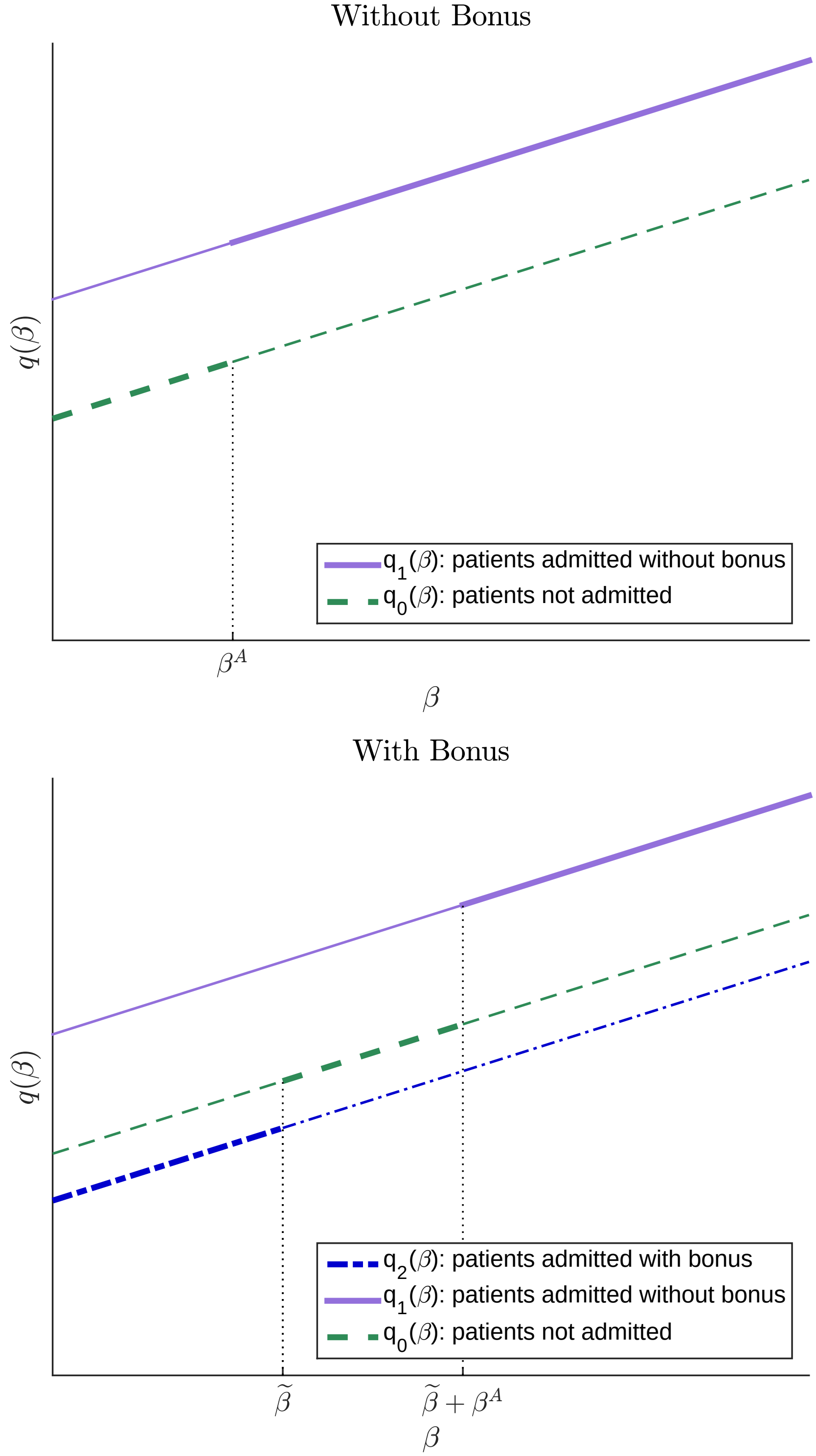
Notes: The bold line sections show the optimal decision rule as a function of β.

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



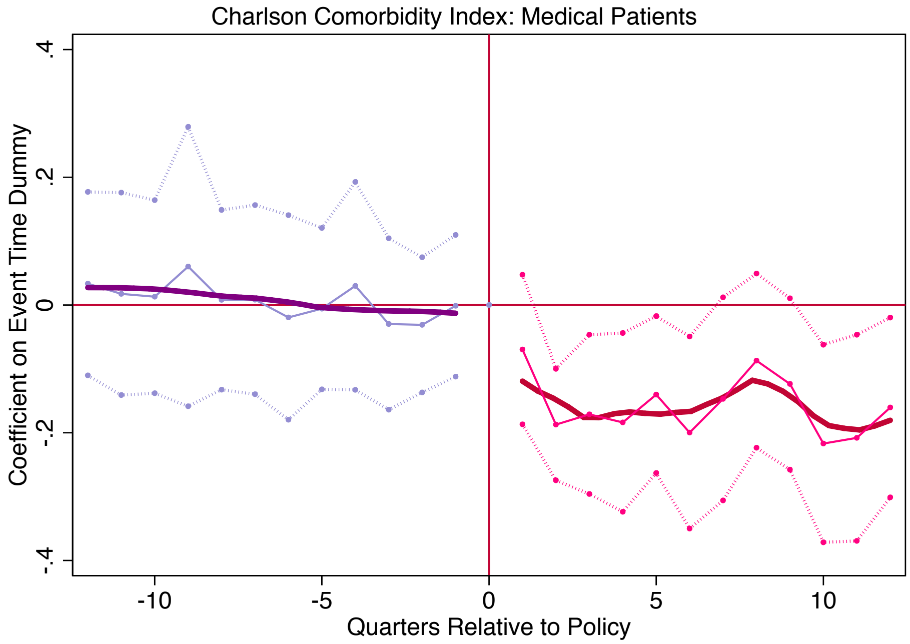
Notes: The bold line sections show the optimal decision rule as a function of β.

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.

Figure 7: Healthier Patients Sent to Participating Hospitals



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

9 Tables

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

All Patient Refined-Diagnosis Related Group (APR-DRG)	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	$1,4034	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 gainsharing formula. Number of patients are for admitted Medicare patients in main sample, which excludes those who went through the emergency room.

Table 2: 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: Data from the 2008 American Hospital Association Annual Survey—the year before the program was implemented. Medicaid/Medicare days are the total number of hospital days used by people with these insurers. Grade A refers to hospitals that scored an "A" on their hospital report card, as reported by he Leapfrog Group (http://www.hospitalsafetyscore.org/). $*p < 0.1, ** p < 0.05, *** p < 0.01$

Table 3: Main Sample Characteristics

Patients		Doctors		Outcomes	
N	815,014	N	3,474	Charlson index	2.26
N (medical)	389,604	Avg. # of patients	215	Total chronic	3.52
N (surgical)	425,410	Avg. # of hospitals	2.2	Viral Infections	0.02
Avg. age	74	Med. # of hospitals	2.0	Kidney disease	0.10
% white	81	Ever in policy hosp	74%	Asthma	0.03
% black	11	Ever in other hosp	39%	Length of stay	6.74
% woman	54	Ever in both types	35%	CT scans	0.03
% in policy hosp	19	Avg. max bonus (med.)	$513	MRIs	0.01
APR-DRGs (med.)	163	Avg. max bonus (surg.)	$697	Diag. imaging	0.15
APR-DRGs (surg.)	117				

Notes: Statistics for main sample of admitted Medicare patients in New Jersey general medical and surgical hospitals, from 2006-2013, which excludes visits that went through the emergency room. APR-DRG stands for All Patient Refined - Diagnosis Related Group. % under policy refers to the percent of patients treated at participating hospitals while the program was in effect.

Table 4: Bonuses Change Which Patients are Admitted

	Medical Patients		Surgical Patients	
	(1)	(2)	(3)	(4)
	Baseline Admission	Baseline Admission	Baseline Admission	Baseline Admission
policy	-0.004	0.028***	-0.008	0.014***
	(0.006)	(0.005)	(0.006)	(0.004)
high bonus		0.043***		-0.027***
		(0.012)		(0.005)
policy * high bonus		-0.076***		-0.031***
		(0.006)		(0.005)
Doctor FEs	✓	✓	✓	✓
Mean	0.90	0.90	0.95	0.95
Clusters	73	73	73	73
N	385845	385845	405400	405400

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

Table 5: Doctors Sort Healthier Patients into Participating Hospitals

Panel A: Medical Patients

	(1) Charlson	(2) Tot. Costs	(3) Tot. Chronic	(4) Asthma	(5) Viral Inf.	(6) Chron. Kidney
policy	-0.111**	-701.617	-0.076*	-0.004**	-0.002**	-0.008***
	(0.042)	(524.629)	(0.047)	(0.002)	(0.001)	(0.003)
Doctor FEs	✓	✓	✓	✓	✓	✓
Mean	2.785	17916.9	4.167	0.0401	0.0188	0.115
Clusters	73	73	73	73	73	73
N	389,604	389,604	389,604	389,604	389,604	389,604

Panel B: Surgical Patients

	(1) Charlson	(2) Tot. Costs	(3) Tot. Chronic	(4) Asthma	(5) Viral Inf.	(6) Chron. Kidney
policy	0.013	128.350	0.024	-0.002**	-0.001	0.001
	(0.033)	(219.007)	(0.058)	(0.001)	(0.001)	(0.003)
Doctor FEs	✓	✓	✓	✓	✓	✓
Mean	1.775	9048.8	2.917	0.0266	0.00948	0.0848
Clusters	73	73	73	73	73	73
N	425,410	425,410	425,410	425,410	425,410	425,410

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

Table 6: Doctors Sort Healthier Patients into Participating Hospitals: High Admission Diagnoses

Panel A: Medical Patients

	(1)	(2)	(3)	(4)	(5)	(6)
	Charlson	Tot. Costs	Tot. Chronic	Asthma	Chron. Kidney	Viral Inf.
policy	-0.127**	-528.816	-0.050	-0.003	-0.001	-0.008**
	(0.057)	(706.427)	(0.064)	(0.003)	(0.002)	(0.004)
Doctor FEs	✓	✓	✓	✓	✓	✓
Mean	3.103	21806.9	4.530	0.0461	0.0206	0.123
Clusters	73	73	73	73	73	73
N	162686	162686	162686	162686	162686	162686

Panel B: Surgical Patients

	(1)	(2)	(3)	(4)	(5)	(6)
	Charlson	Tot. Costs	Tot. Chronic	Asthma	Chron. Kidney	Viral Inf.
policy	0.005	244.348	0.016	-0.003*	-0.002*	-0.002
	(0.027)	(164.893)	(0.064)	(0.002)	(0.001)	(0.002)
Doctor FEs	✓	✓	✓	✓	✓	✓
Mean	1.242	5108.9	2.269	0.0240	0.00758	0.0494
Clusters	71	71	71	71	71	71
N	127163	127163	127163	127163	127163	127163

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

Table 7: Effect of Program on Population-Level Severity of Illness

	Medical		Surgical	
	(1)	(2)	(3)	(4)
	SOI	SOI	SOI	SOI
	All	High Adm.	All	High Adm.
policy	-0.026	-0.024	0.011	0.015
	(0.021)	(0.028)	(0.014)	(0.021)
Doctor FEs	✓	✓	✓	✓
Mean	2.363	2.560	1.996	1.945
Clusters	73	73	73	71
N	389,604	162,686	425,410	127,163

Notes: Quarter, doctor, hospital, and diagnosis 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

Table 8: 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	-0.093	-0.000	-0.001	-0.001	-0.002	391
	(0.196)	(0.003)	(0.002)	(0.003)	(0.006)	(516)
Doctor FEs	✓	✓	✓	✓	✓	✓
Mean	6.979	0.032	0.016	0.037	0.112	11459
Clusters	73	73	73	73	73	58
N	389,604	389,604	389,604	389,604	389,604	355,306

Panel B: Surgical Patients

	(1) Length of Stay	(2) CT Scan	(3) MRI	(4) Diag. Ultra	(5) Any Imaging	(6) Total Costs
policy	-0.042	-0.003	-0.000	0.001	0.008	889
	(0.100)	(0.003)	(0.001)	(0.006)	(0.010)	(560)
Doctor FEs	✓	✓	✓	✓	✓	✓
Mean	6.516	0.025	0.008	0.061	0.186	18557
Clusters	73	73	73	73	73	58
N	425,410	425,410	425,410	425,410	425,410	371,262

Notes: Quarter, 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 6. Standard errors clustered at the hospital level. $*p < 0.1$, $** \ p < 0.05$, $*** \ p < 0.01$

Table 9: Replicating the Naive Evaluation: Medical Patients

	(1) Total Costs	(2) Total Costs	(3) Total Costs	(4) Total Costs	(5) Total Costs
policy	-469**	-469	-364	414	606
	(219)	(869)	(849)	(636)	(612)
Health Controls			✓	✓	✓
Comparison Hospitals				✓	✓
Doctor FEs					✓
Mean	10,910	10,910	10,910	11,281	11,281
Clusters	-	11	11	58	58
N	85374	85,374	85,374	293,052	293,052

Notes: Quarter, hospital, and diagnosis by severity of illness fixed effects also included. $*$p $<$ 0.1, $**$ p $<$ 0.05, $***$ p $<$ 0.01

Table 10: Simulated Share Treated on Costs and Procedure Use

	(1) Total Costs	(2) Patients Admitted	(3) CT Scans	(4) MRIs MRIs	(5) Diagnostic Ultrasounds	(6) Any Diag. Imaging	(7) Total LOS
Simulated share	8432.72*	0.25	0.08	-0.01	0.08**	0.22**	1.48
	(4561.81)	(0.38)	(0.07)	(0.03)	(0.04)	(0.10)	(2.39)
Doctor FEs	✓	✓	✓	✓	✓	✓	✓
Mean	210940.9	19.99	1.078	0.437	0.835	2.862	122.8
Clusters	3394	3394	3394	3394	3394	3394	3394
N	84626	84626	84626	84626	84626	84626	84626

Notes: Doctor and quarter fixed effects included. The outcome variables are summed at the doctor quarter level. Total LOS is the sum of length of stay across a doctor's patients in each quarter. $*$p $<$ 0.1, $**$ p $<$ 0.05, $***$ p $<$ 0.01

References

Hospital gain-sharing program offers incentives to physicians based on their efficiency, producing significant cost savings without decline in quality. Agency for Healthcare Research and Quality, 2014.

Jason Abaluck, 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.

Diane Alexander. Does physician pay affect procedure choice and patient health? evidence from medicaid c-section use. Technical report, Princeton, 2015.

Hunt Allcott. Site selection bias in program evaluation. *The Quarterly Journal of Economics*, 130(3):1117–1165, 2015.

Jason Brown, Mark Duggan, Ilyana Kuziemko, and William Woolston. Will risk-adjustment decrease health care costs? new evidence from the medicare advantage program. Technical report, 2011.

Mary E. Charlson, Robert E. Charlson, Janey C. Peterson, Spyridon S. Marinopoulos, William M. Briggs, and James P. Hollenberg. The charlson comorbidity index is adapted to predict costs of chronic disease in primary care patients. *Journal of Clinical Epidemiology*, 61(12):1234–1240, 2008.

Jeffrey Clemens and Joshua Gottlieb. Do physicians' financial incentives affect medical treatment and patient health? *The American Economic Review*, 104(4):1320–1349, 2014.

Dominic Coey. Physician incentives and treatment choices in heart attack management. Technical report, SIEPR Discussion Paper No. 12-027, 2013.

Julie Berry Cullen, Steven D. Levitt, Erin Robertson, and Sally Sadoff. What can be done to improve struggling high schools? *The Journal of Economic Perspectives*, 27(2):133–152, 2013.

David Cutler. Empirical evidence on hospital delivery under prospective payment. Technical report, MIT, 1990.

David Cutler. The incidence of adverse medical outcomes under prospective payment. *Econometrica*, 63(1):29–50, 1995.

Leemore Dafny. How do hospitals respond to price changes? *American Economic Review*, 95(5):1525–1547, 2005.

Michael Dickstein. Physician vs. patient incentives in prescription drug choice. Technical report, 2014.

David Dranove and Paul Wehner. Physician-induced demand for childbirths. *The Journal of Health Economics*, 13(1):61–73, 1994.

David Dranove, Daniel Kessler, Mark McClellan, and Mark Satterthwaite. Is more information better? the effects of "report cards" on health care providers. *Journal of Political Economy*, 111(3):555–588, 2003.

Esther Duflo. Scaling up and evaluation. Technical report, Annual World Bank Conference on Development Economics, 2004.

Mark Duggan. Does contracting out increase the efficiency of government programs? evidence from medicaid hmos. *Journal of Public Economics*, 88(12):2549–2572, 2004.

Mark Duggan 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*, 32(3):505–535, 2013.

Randall Ellis and Thomas McGuire. Hospital response to prospective payment: Moral hazard, selection, and practice-style effects. *The Journal of Health Economics*, 15:257–277, 1996.

Darren Grant. Physician financial incentives and cesarean delivery: New conclusions from the healthcare cost and utilization project. *Journal of Health Economics*, 28(1):244–250, 2009.

Jonathan Gruber. Covering the uninsured in the united states. *Journal of Economic Literature*, 46(3):51–68, 2008.

Jonathan Gruber. The cost implications of health care reform. *The New England Journal of Medicine*, 362:2050–2051, 2010.

Jonathan Gruber and Maria Owings. Physician financial incentives and cesarean section delivery. *RAND Journal of Economics*, 27(1):99–123, 1996.

Jonathan Gruber, John Kim, and Dina Mayzlin. Physician fees and procedure intensity: the case of cesarean delivery. *Journal of Health Economics*, 18(4):473–490, 1999.

Jack Hadley, Jean M. Mitchell, and Jeanne Mandelblatt. Medicare fees and small area variations in breast-conserving surgery among elderly women. *Medical Care Research and Review*, 58(3):334–360, 2001.

Jack Hadley, James Reschovsky, Catherine Corey, and Stephen Zuckerman. Medicare fees and the volume of physicians' services. *Inquiry*, 46(4):372–390, 2009.

Bruce J. Hillman and Jeff C. Goldsmith. The uncritical use of high-tech medical imaging. *The New England Journal of Medicine*, 363:4–6, 2010.

Kate Ho and Ariel Pakes. Hospital choices, hospital prices and financial incentives to physicians. Technical report, NBER Working Paper No. 19333, 2014.

E. Keeler and T. Folk. Equalizing physician fees had little effect on cesarean rates. *Medicare Care Research and Review*, 53:465–471, 1996.

Jonathan Kolstad. Information and quality when motivation is intrinsic: Evidence from surgeon report cards. *American Economic Review*, 103(7):2875–2910, 2013.

Arleen Leibowitz, Joan L. Buchanan, and Joyce Mann. A randomized trial to evaluate the effectiveness of a medicaid hmo. *Journal of Health Economics*, 11(3):235–257, 1992.

Adam L. Lessler, Joshua A. Isserman, Rajan Agarwal, Harold I. Palevsky, and Jesse M. Pines. Testing low-risk patients for suspected pulmonary embolism: A decision analysis. *Annals of Emergency Medicine*, 55(4):316–326, 2010.

Willard Manning, 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*, 77(3):251–277, 1987.

Department of Health and Human Services. Better care. smarter spending. healthier people: Paying providers for value, not volume. Technical report, 2015.

OIG. Special advisory bulletin: Gainsharing arrangements and cmps for hospital payments to physicians to reduce or limit services to beneficiaries. Technical report, Office of Inspector General, 1999.

Peter Orszag and Ezekiel Emanuel. Health care reform and cost control. *New England Journal of Medicine*, 363(7):601–603, 2010.

Hude Quan, Bing Li, and Chantal M. 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*, 173(6):676–682, 2011.

Vijay M. Rao and David C. Levin. The overuse of diagnostic imaging and the choosing wisely initiative. *Annals of Internal Medicine*, 157(8):574–576, 2012.

doi: 10.7326/0003-4819-157-8-201210160-00535. URL +http://dx.doi.org/10.7326/0003-4819-157-8-201210160-00535.

Uwe E. Reinhardt. The pricing of u.s. hospital services: Chaos behind a veil of secrecy. *Health Affairs*, 25:57–69, 2006.

Elaine Silverman and Jonathan Skinner. Medicare upcoding and hospital ownership. *Journal of Health Economics*, 23:369–389, 2004.

Sherry L. Smith, editor. *Medicare RBRVS 2012: The Physician Guide*. American Medical Association, 2012.

Rebecca Smith-Bindman. Is computed tomography safe? *New England Journal of Medicine*, 363(1):1–4, 2010. doi: 10.1056/NEJMp1002530. URL http://www.nejm.org/doi/full/10.1056/NEJMp1002530. PMID: 20573919.

Justin W. Timbie, Peter S. Hussey, Lane F. Burgette, Neil S. Wenger, Afshin Rastegar, Ian Brantley, Dmitry Khodyakov, Kristin J. Leuschner, Beverly A. Weidmer, and Katherine L. Kahn.] medicare imaging demonstration final evaluation: Report to congress. Technical report, RAND Corporation, 2014.

Winnie Yip. 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*, 17:675–699, 1998.

10 Appendix

10.1 Bonus Calculation Details:

Maximum Bonus:

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

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

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

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

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

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

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

10.2 Mathematical Appendix: Model of doctor Decision-making

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

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

Doctors' profits in the normal hospital are the amount of services provided, q, multiplied by a reimbursement rate, a. If $A = 1$ and $H = 1$, doctors may also receive a bonus: $\max\left\{\alpha_0 - \alpha_1 q, 0\right\}$. 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:

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

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:

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

Where from the first order conditions:

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

and

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

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.1). 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:

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

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

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

Post-Period: Hospital 1 Offers a Bonus

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

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

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

Subject to the capacity constraint: 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

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

subject to the capacity constraint; only β' patients can be admitted to each hospital.

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

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

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

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

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

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

Under certain conditions, the value function of the doctor is composed of three segments of the three parts of the value function, which maximize total utility (see Figure 5). Assume doctors decide which patients to admit and where to admit them by dividing their patients into three segments of 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 β' 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.2.

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

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

Where:

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

$$q_{(0)'} = \tfrac{1}{b_2}\left[\left(\tfrac{\lambda}{1-\lambda}\right)(a) + \widetilde{\beta}\right]$$

$$q_{(1)} = \tfrac{1}{b}\left[\left(\tfrac{\lambda}{1-\lambda}\right)(a) + \widetilde{\beta} + \beta^A + \gamma\right]$$

$$q_{(0)''} = \tfrac{1}{b}\left[\left(\tfrac{\lambda}{1-\lambda}\right)(a) + \widetilde{\beta} + \beta^A\right]$$

$$\beta^A = \tfrac{2a\gamma\lambda + 2bC\lambda - 2bC - \gamma^2\lambda + \gamma^2}{2\gamma(\lambda-1)}$$

$$\widetilde{\beta}' = \frac{\lambda(\lambda\alpha_1(2a-2\gamma-\alpha_1)+2b\alpha_0(\lambda-1)+2\gamma\alpha_1)+(\lambda-1)\left(2a\gamma\lambda+2bC\lambda-2bC-\gamma^2\lambda+\gamma^2\right)}{2\lambda\alpha_1(\lambda-1)} \tag{20}$$

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

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.

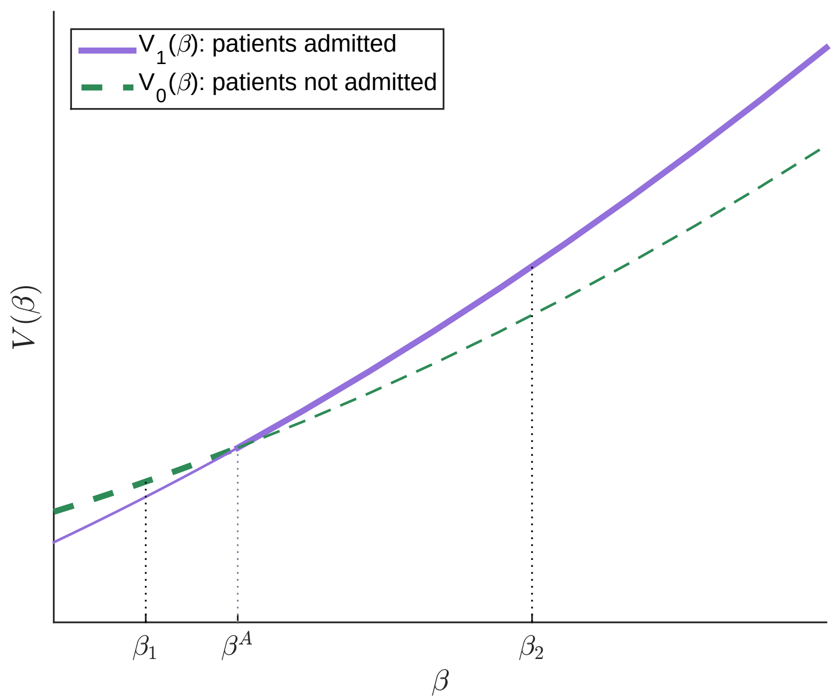
Figure A.3 shows both cases: 4.A demonstrates the case where the quantity of care provided for the bonus generating patients is less under the bonus program than the counterfactual of no bonuses; 4.B shows the opposite.

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: $U\left(q_{(2)}, \widetilde{\beta}\right) - U\left(q_{(0)}, \tilde{\beta}\right) < U\left(q_{(1)}, \bar{\beta}\right) - U\left(q_{(0)}, \bar{\beta}\right)$.

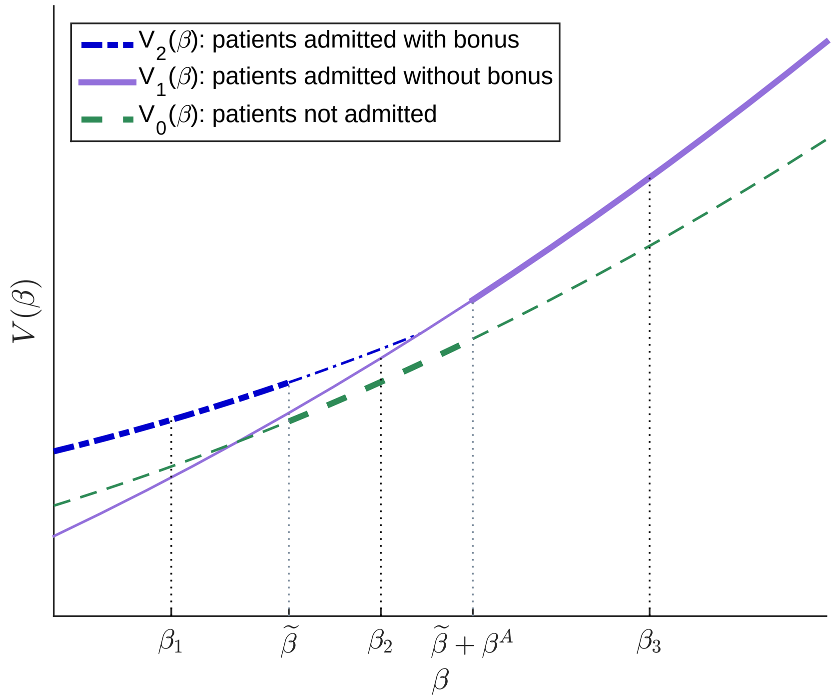
10.3 Appendix Figures

Figure A.1: Doctor's Utility as a Function of β: without Bonuses



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

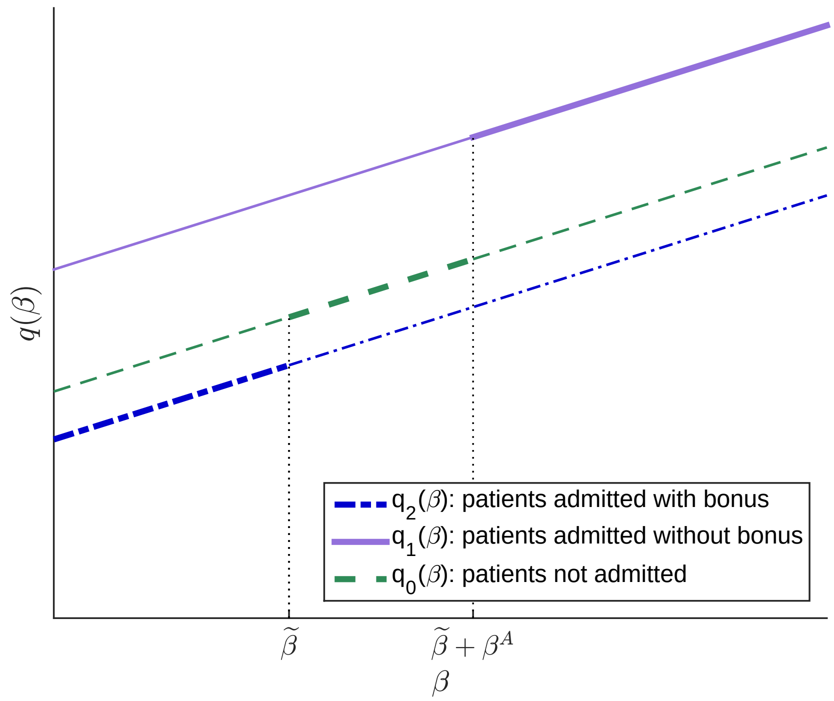
Figure A.2: Doctor's Utility as a Function of β: with Bonuses



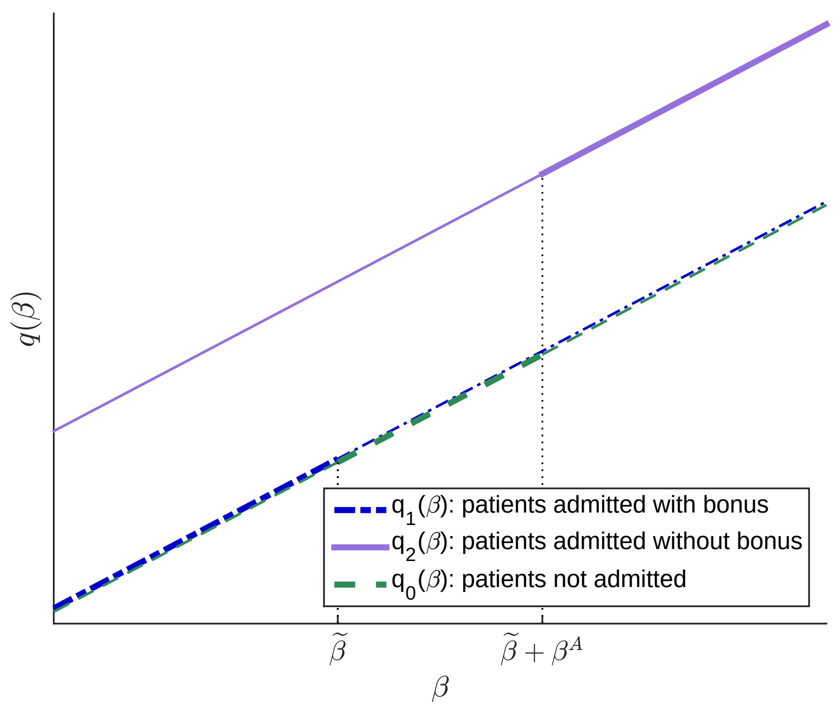
The bold line sections show the optimal decision rule as a function of β.

Figure A.3: Optimal Quantity of Care as a Function of β: with Bonuses

A

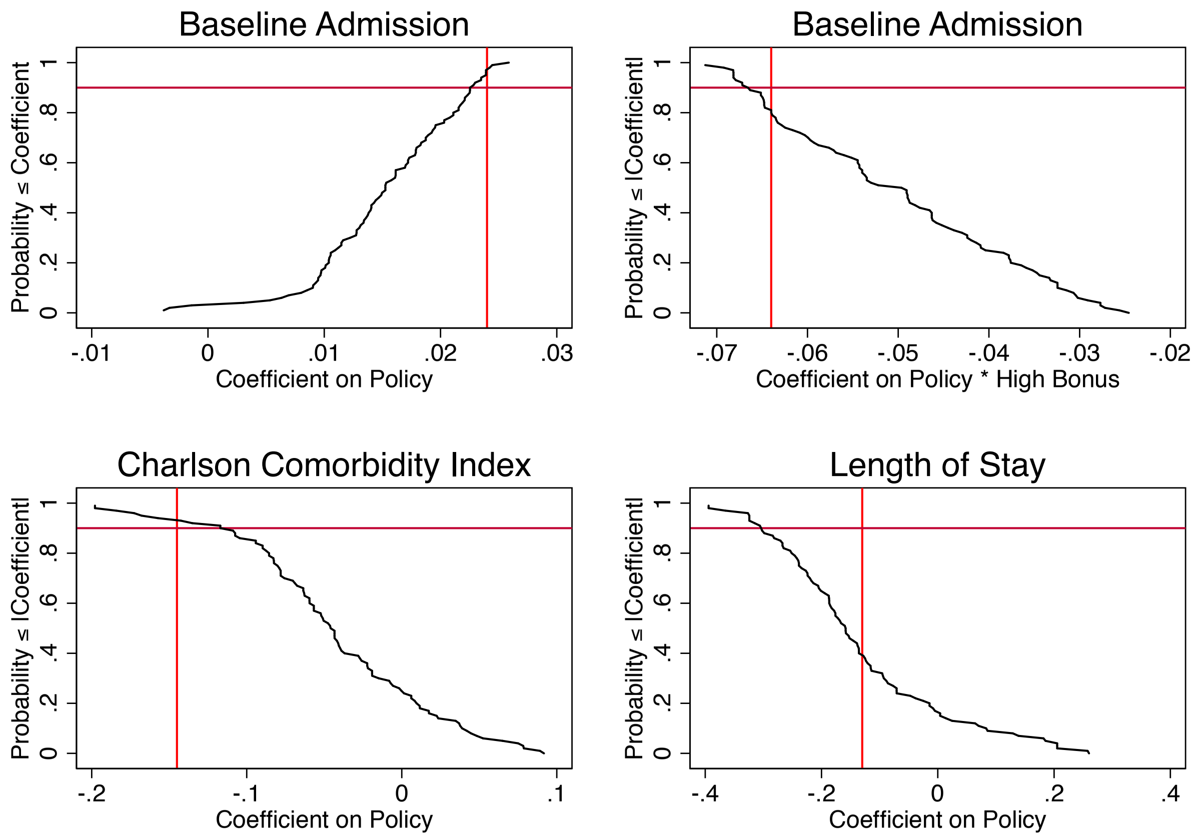


B



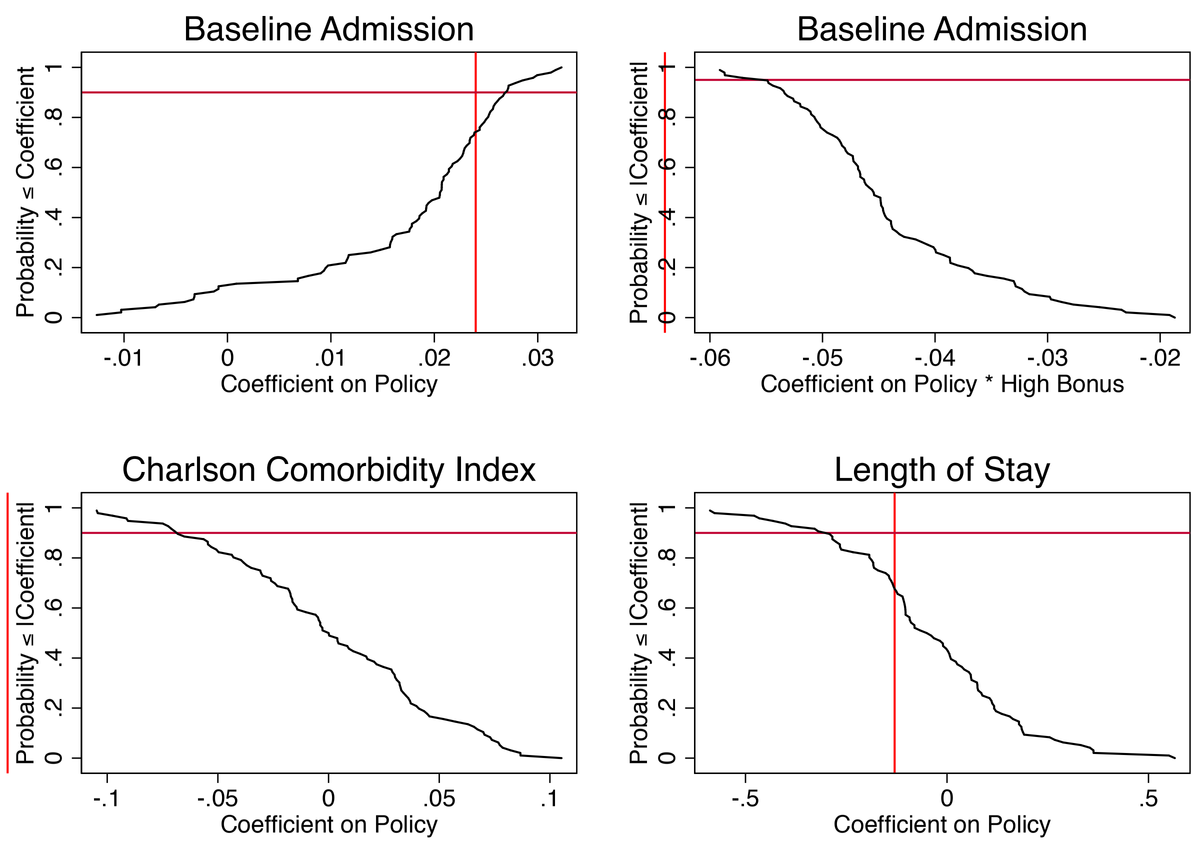
The bold line sections show the quantity of care provided along the optimal decision rule. Figure A.3A shows the optimal quantity of care under one set of parameters; Figure A.3B shows the optimal quantity of care under another set of parameters.

Figure A.4: Placebo Test: Randomize Dates



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

Figure A.5: Placebo Test: Randomize Program Participation



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

10.4 Appendix Tables

Table A.1: Top 15 Diagnoses for Medical and Surgical Patients
(CCS Diagnosis Codes)

Medical Patients			
Primary Diagnosis	Frequency	Percent	Cumulative
Diseases of the heart	128,993	16.06	16.06
Diseases of the urinary system	54,048	6.73	22.7
Upper gastrointestinal disorders	37,511	4.67	27.4
Complications	30,919	3.85	31.3
Symptoms; signs; and ill-defined condit	28,173	3.51	34.8
Benign neoplasms	28,106	3.5	38.3
Lower gastrointestinal disorders	27,110	3.38	41.6
Spondylosis; intervertebral disc disord	26,493	3.3	44.9
Cerebrovascular disease	23,615	2.94	47.9
Respiratory infections	20,821	2.59	50.5
Chronic obstructive pulmonary disease a	16,985	2.11	52.6
Anemia	16,504	2.05	54.6
Eye disorders	15,052	1.87	56.5
Diseases of arteries; arterioles; and c	14,117	1.76	58.3
Other nervous system disorders	13,446	1.67	60

Surgical Patients			
Primary Diagnosis	Frequency	Percent	Cumulative
Diseases of the heart	84,362	14.52	14.5
Eye disorders	65,469	11.27	25.8
Non-traumatic joint disorders	51,587	8.88	34.6
Complications	34,379	5.92	40.6
Abdominal hernia	24,618	4.24	44.8
Fractures	24,449	4.21	49
Spondylosis; intervertebral disc disord	22,213	3.82	52.8
Diseases of female genital organs	20,387	3.51	56.3
Diseases of arteries; arterioles; and c	20,056	3.45	59.8
Diseases of the urinary system	16,338	2.81	62.6
Cancer of urinary organs	14,540	2.5	65.1
Biliary tract disease	14,131	2.43	67.5
Cancer of breast	13,816	2.38	69.9
Cerebrovascular disease	11,506	1.98	71.9
Diseases of male genital organs	10,823	1.86	73.8

Table A.2: Participating Hospitals

Gainsharing Demonstration	Gainsharing Extension
AtlantiCare Regional Medical Center	CentraState Healthcare System
Overlook Hospital	Hunterdon Medical Center
Holy Name Hospital	Jersey Shore University Medical Center
Jersey Shore University Medical Center	JFK Medical Center
Hunterdon Medical Center	Monmouth Medical Center
Monmouth Medical Center	Overlook Hospital
St. Francis Medical Center	St. Francis Medical Center
Our Lady of Lourdes Medical Center	The Valley Hospital
The Valley Hospital	
Somerset Medical Center	
JFK Medical Center	
CentraState Healthcare System	

BPCI Model 1 Program	
Capital Health Medical Center-Hopewell	RWJ University Hospital- Rahway
Capital Health Regional Medical Center	Saint Clare's Hospital- Denville
CentraState Medical Center	Saint Clare's Hospital- Dover
Cooper Hospital/University Medical Center	Saint Michael's Medical Center
Deborah Heart and Lung Center	Saint Peter's University Hospital
Hunterdon Medical Center	South Jersey Healthcare- Elmer
Jersey Shore University Medical Center	South Jersey Healthcare- RMC
JFK Medical Center	St. Joseph's RMC
Morristown Medical Center	St. Mary's Hospital Passaic
Overlook Medical Center	The Valley Hospital
Robert Wood Johnson (RWJ) University Hospital	Underwood-Memorial Hospital
RWJ University Hospital- Hamilton	University Medical Center of Princeton

Notes: RMC stands for regional medical center.

10.4.1 ER Medicare Patients

Table A.3: Effect of Program on Baseline Admission Probability: Medicare ER Patients

	Medical Patients		Surgical Patients	
	(1)	(2)	(3)	(4)
	Baseline Adm.	Baseline Adm.	Baseline Adm.	Baseline Adm.
policy	0.003	0.002	0.000	-0.001
	(0.004)	(0.004)	(0.002)	(0.002)
high bonus		0.317***		0.497***
		(0.011)		(0.053)
policy * high bonus		0.004		0.001
		(0.003)		(0.002)
Doctor FEs	✓	✓	✓	✓
Mean	0.752	0.752	0.969	0.969
Clusters	78	78	76	76
N	1585753	1585753	69742	69742

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

Table A.4: Effect of Program on Ex-Ante Patient Health: Medicare ER Patients

	Medical Patients		Surgical Patients	
	CCI Past	Tot. Chronic	CCI Past	Tot. Chronic
policy	-0.04	-0.02	-0.05	-0.06
	(0.02)	(0.04)	(0.04)	(0.06)
Doctor FEs	✓	✓	✓	✓
Mean	2.562	5.159	1.918	4.504
Clusters	78	78	76	76
N	1592186	1592186	95542	95542

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

Table A.5: Effect of Program on Costs and Procedure Use: Medicare ER Patients

Panel A: Medical Patients

	(1) Length of Stay	(2) CT Scan	(3) MRI	(4) Any Imaging	(5) Diag. Ultra	(6) Total Costs
policy	-0.026	-0.003	-0.003	0.004	0.009*	471.649
	(0.084)	(0.005)	(0.003)	(0.007)	(0.005)	(387.532)
Doctor FEs	✓	✓	✓	✓	✓	✓
Mean	5.211	0.0797	0.0306	0.135	0.0364	8300.5
Clusters	78	78	78	78	78	65
N	1592186	1592186	1592186	1592186	1592186	1341584

[0.5em]

Panel B: Surgical Patients

	(1) Length of Stay	(2) CT Scan	(3) MRI	(4) Any Imaging	(5) Diag. Ultra	(6) Total Costs
policy	-0.249	-0.003	-0.002	0.002	0.007	249.986
	(0.236)	(0.006)	(0.005)	(0.010)	(0.007)	(584.883)
Doctor FEs	✓	✓	✓	✓	✓	✓
Mean	7.655	0.0746	0.0268	0.202	0.0485	17000.0
Clusters	76	76	76	76	76	65
N	95542	95542	95542	95542	95542	82959

Notes: Quarter, 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 6. Standard errors clustered at the hospital level. $*p < 0.1$, $** p < 0.05$, $*** p < 0.01$

10.4.2 Near Medicare Patients (Ages 50-64)

Table A.6: Effect of Program on Baseline Admission Probability: Near Medicare Patient

	Medical Patients		Surgical Patients	
	(1)	(2)	(3)	(4)
	Baseline Adm.	Baseline Adm.	Baseline Adm.	Baseline Adm.
policy	0.000	0.004	-0.004	-0.001
	(0.003)	(0.004)	(0.005)	(0.004)
high bonus		0.028		0.609***
		(0.028)		(0.020)
policy * high bonus		-0.011**		-0.006**
		(0.004)		(0.002)
Doctor FEs	✓	✓	✓	✓
Mean	0.699	0.699	0.845	0.845
Clusters	79	79	78	78
N	183259	183259	247488	247488

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

Table A.7: Effect of Program on Ex-Ante Patient Health: Near Medicare Patients

	Medical Patients		Surgical Patients	
	CCI Past	Tot. Chronic	CCI Past	Tot. Chronic
policy	-0.06	-0.05	-0.02	-0.01
	(0.04)	(0.03)	(0.02)	(0.03)
Doctor FEs	✓	✓	✓	✓
Mean	1.866	4.162	0.885	3.531
Clusters	79	79	78	78
N	187054	187054	256141	256141

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

Table A.8: Effect of Program Costs and Procedure Use: Near Medicare Patients

Panel A: Medical Patients

	(1) Length of Stay	(2) CT Scan	(3) MRI	(4) Any Imaging	(5) Diag. Ultra	(6) Total Costs
	b/se	b/se	b/se	b/se	b/se	b/se
policy	-0.020	-0.000	-0.001	0.000	0.003	612.815
	(0.162)	(0.004)	(0.003)	(0.007)	(0.004)	(527.039)
Doctor FEs	✓	✓	✓	✓	✓	✓
Mean	6.181	0.0344	0.0190	0.115	0.0422	10459.2
Clusters	79	79	79	79	79	66
N	187054	187054	187054	187054	187054	173591

Panel B: Surgical Patients

	(1) Length of Stay	(2) CT Scan	(3) MRI	(4) Any Imaging	(5) Diag. Ultra	(6) Total Costs
	b/se	b/se	b/se	b/se	b/se	b/se
policy	0.056	-0.003	-0.000	0.008	-0.002	1032.178**
	(0.103)	(0.003)	(0.001)	(0.009)	(0.004)	(458.950)
Doctor FEs	✓	✓	✓	✓	✓	✓
Mean	5.052	0.0180	0.00716	0.135	0.0458	16148.8
Clusters	78	78	78	78	78	66
N	256141	256141	256141	256141	256141	230910

Notes: Quarter, 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 6. Standard errors clustered at the hospital level. $*p < 0.1$, $**\ p < 0.05$, $***\ p < 0.01$