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

Diane Alexander

Federal Reserve Bank of Chicago

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

I. Introduction

Lowering the growth in health care costs has long been a top US 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).

I thank Janet Currie, Ilyana Kuziemko, Hank Farber, Jonathan Gruber, Tom Vogl, Anne Case, Molly Schnell, Sharon Traiberman, Thomas Winberry, and seminar participants at Princeton University, the University of Virginia, the University of Maryland, the Massachusetts Institute of Technology Sloan School of Management, Northwestern University Kellogg School of Management, the Federal Reserve Bank of Chicago, and the University of Notre Dame. In addition, I am indebted to participants of the National Bureau of Economic Research Summer Institute Health Care Workshop for their helpful comments. I also thank the staff at the New Jersey Department of Health for their help in accessing the data, and the Center for Health and Wellbeing for financial support. Programs used to generate results, as well as information for how to apply to access the data, are provided as supplementary material online.

Electronically published October 23, 2020
[*Journal of Political Economy*, 2020, vol. 128, no. 11]

Because of this uncertainty, the Patient Protection and Affordable Care Act (ACA) earmarked billions of dollars for pilot programs through the Center for Medicare and Medicare Innovation.[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 focuses on changing the financial incentives of doctors—motivated by the idea that the current system of paying doctors separately for each service provided ("fee for service") encourages them to perform unnecessary procedures. These pilot programs purport to study how doctors respond to different payment schemes, an important open question in the literature.

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

I shed new light on the importance of experimental design in health care using the New Jersey Gainsharing Demonstration, a pilot program in which hospitals paid doctors bonuses for reducing the total treatment costs for Medicare admissions. The bonuses were designed to increase when total treatment costs decreased, and thus discourage the use of treatments with low marginal benefits. Under the program, patients are divided into types by diagnosis and severity of illness categories, and a maximum bonus is assigned to each type. Doctors are then paid a fraction of this maximum bonus, depending on how close the total treatment cost is to preprogram cost benchmarks.

Under the Gainsharing Demonstration, only the treatment of an admitted Medicare patient in a participating hospital could generate a bonus for the physician. Not all patients are admitted, not all hospitals

[1] The Center for Medicare and Medicaid Innovation was established by sec. 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).

participated in the demonstration, and most doctors in New Jersey admit patients in more than one hospital. Thus, doctors could increase their expected bonuses in three ways under the program: change which patients to admit, change where to admit them, and change how patients were treated.

Doctors responded to the bonuses by reallocating admission across patients—by both changing admission thresholds and diverting healthier patients into participating hospitals. Nonsurgical patients admitted to participating hospitals had lower scores on comorbidity indexes based on previous visits, conditional on their type. As healthier patients are cheaper to treat, doctors receive higher bonuses for treating these patients, on average. Defining the bonuses within diagnosis and severity level cells was meant to serve as a type of risk adjustment. However, I find that doctors are able to identify low-cost patients even within these groups, and exploit this knowledge to increase their expected bonus payments.

Yet, conditional on admission and patient health, the bonuses did not reduce costs or change procedure use. I look at many measures of services performed: length of stay, the use of diagnostic imaging procedures labeled as overused by doctors (CT scans, MRIs, and other diagnostic imaging procedures), and total costs. I find no evidence that doctors lowered costs or changed their procedure use in response to the bonuses.

I use a difference-in-difference strategy with doctor fixed effects to measure the effect of the bonuses on doctor's admitting and treatment behavior in hospital discharge records. One critique of this doctor-level difference-in-difference specification is that doctors may respond to incentives in one hospital by changing their practice style at all hospitals in which they work. Through the lens of a within-doctor identification strategy, changing practice styles would look like a null effect. Using an alternative strategy based on doctor-level program exposure, I rule out this alternative interpretation. Consistent with the main results, there is no evidence that the bonuses are associated with lower costs; if anything, costs appear to rise with program exposure.

Changing the composition of admitted patients has the potential to negatively affect both patients and Medicare itself. For patients, admission can be the difference between intense and prolonged monitoring and being sent home after treatment. For Medicare, admission means an order of magnitude higher charges. Furthermore, while the Gainsharing Demonstration explicitly forbade increasing overall admission rates due to the bonus program, it is unclear whether this rule can be enforced in the long run. Any increases in overall admission rates would be extremely costly to both Medicare and the patients themselves.

While sorting healthier patients into participating hospitals may seem comparatively benign, this behavior can severely bias policy evaluations and result in ineffective programs being taken to scale. In an early evaluation of the Gainsharing Demonstration, the Agency for Healthcare Research and Policy (2014) reported that the bonuses reduced costs per admission by 8%. The apparent success of the first wave of the program led to its expansion. However, the initial evaluation only compared the costs of admitted patients at participating hospitals, before and after the program was implemented. I replicate this exercise and show that a simple pre- versus post-implementation comparison of admitted patients is misleading, and that the apparent cost savings disappears in a more careful evaluation. The response of physicians to the New Jersey Gainsharing Demonstration highlights the importance of program design to both the effectiveness of the payment model and the ability to generate internally valid estimates of the demonstration's efficacy.

A. *Related Literature*

This paper contributes to three main strands of literature. First, it is directly related to the literature on how doctors respond to financial incentives. There is a large body of work studying how reimbursement levels influence procedure choice, mostly focusing on the decision to perform one particular procedure (Dranove and Wehner 1994; Gruber and Owings 1996; Keeler and Fok 1996; Yip 1998; Gruber, Kim, and Mayzlin 1999; Hadley, Mitchell, and Mandelblatt 2001, 2009; Grant 2009; Coey 2013; Clemens and Gottlieb 2014; Alexander 2015).[2] These papers generally find that doctors supply more services when payment increases, as well as when the payment of a competing procedure decreases. An implication of this research is that reforms that lower the profit for performing "unnecessary" procedures could be very effective at lowering costs.[3]

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 (Dickstein 2014; Ho and Pakes 2014).[4] Unfortunately, studying the effect

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

[3] Consumer cost sharing has also been on the rise as a demand-side strategy to decrease health care expenditures by discouraging the use of low-value care. In 2016, 40% of consumers with private insurance under the age of 65 were enrolled in a high-deductible health plan, an increase of 25% from 2010 (Cohen, Martinez, and Zammitti 2016). While high-deductible plans have been shown to reduce health care expenditures, an important caveat is the now ample evidence that people who are switched to high-deductible plans tend to reduce the use of both high- and low-value treatment (Haviland et al. 2011; Brot-Goldberg et al. 2017; Fendrick and Chernew 2017; Wharam et al. 2017).

[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 changing treatment intensity and coding practices in response to diagnosis related group (DRG) specific price changes.

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

Second, doctors sending healthier patients to participating hospitals is similar to evidence that managed-care plans are able to select healthier patients into their plans, and that hospitals respond to readmission penalties with selective readmission of returning patients (Leibowitz, Buchanan, and Mann 1992; Duggan 2004; Brown et al. 2011; Duggan and Hayford 2013; Gupta 2016). 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 causes cardiac surgeons to select healthier patients, while Kolstad (2013) finds little evidence of selection. Especially with the recent popularity of cost reduction strategies that target doctor pay, it is important to know whether doctors are able to identify low-cost patients to treat.

Third, the problems and limitations of pilot programs have been widely studied in economics. An exhaustive literature review is beyond the scope of this paper.[5] These lessons, however, have generally not been applied to US health care reform. The Centers for Medicare and Medicaid Services has been running pilot programs (or "demonstrations") since the 1960s, and the Affordable Care Act appropriated $10 billion for the Center for Medicare and Medicaid Innovation, which tests "innovative health care payment and service delivery models." Furthermore, the results of these pilot programs help direct the annual spending of Medicare, a $600 billion per year program. In this paper, I point out that even when there is evidence that such programs are effective, it may be due to gaming rather than true improvements in efficiency.

B. Road Map

The rest of the paper is organized as follows. Section II describes the bonus program and the specific incentives it created for doctors. Section III develops a model of doctor decision making. The model shows that the bonuses incentivize doctors to change who is admitted and to sort patients between hospitals. The effect of the bonuses on resource use, however, is ambiguous. In the remainder of the paper, I measure the impact

[5] See, e.g., Duflo (2004), Cullen et al. (2013), and Allcott (2015).

of the bonuses empirically. Section IV describes my data and identification strategy, and results are presented in sections V and VI. Section VII discusses the implications of this study for the scaling literature, and section VIII concludes.

II. Institutional Background

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

A. How Doctors Treat Patients within Hospitals

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

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

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

When treating a patient in a hospital setting, doctors also decide whether to admit a patient and treat them, or treat the patient in the hospital on an outpatient basis. The technical definition of admission is simply that a doctor has written an order to that effect. In practice, admitted patients generally stay at least overnight and occupy a bed. Doctors considering admission weigh the benefits against the costs. While admitted patients are intensely monitored and receive more care, admission is also costly for the patient, in terms of both time and money. In addition, admitted patients spend more time in the hospital and thus face a higher risk of contracting hospital-acquired infections, which are often resistant to treatment. In nearly all diagnosis groups there are both patients treated with and without being admitted. Among Medicare patients with cardiac arrhythmia, for example, 51% of patients in my data are emergent admissions, 18% are elective admissions, 20% are treated outpatient in the emergency department, and 12% are outpatients with no emergency department revenue.

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

B. *How Doctors and Hospitals Are Paid*

For the most part, doctors in the United States are paid under a fee-for-service system, whereas hospitals are paid either a fixed amount per visit according to a broad diagnosis category or a per diem for each day spent in the hospital (Reinhardt 2006).[6] Traditional Medicare is no exception. Medicare Part A pays hospitals a fixed sum based on the patient's diagnosis (called diagnosis related groups, or "DRGs") for treating admitted patients. Conversely, physician services are paid for by Medicare Part B, which pays doctors separately for each service provided to the patient.

[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 one-third of hospital revenue (Reinhardt 2006).

Given these payment systems, the financial incentives of doctors and hospitals over how much care to provide are fundamentally at odds—pushing doctors to do more and hospitals to do less. While hospitals can theoretically constrain doctors' resource use through the threat of revoking their privileges, in reality this is difficult. Doctors can sue hospitals for loss of privileges, and hospitals are vulnerable to large suits for damages if they cannot establish that their action complied with the requirements of the Health Care Quality Improvement Act.[7] 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.[8] Medicare demonstration projects, however, are typically granted waivers from these statutes.

C. The Gainsharing Demonstration

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

[7] To qualify for immunity from liability under the Act, the hospital must establish that the action was taken (1) in the reasonable belief that the action was in the furtherance of quality health care; (2) after a reasonable effort to obtain the facts of the matter; (3) after adequate notice and hearing procedures are afforded to the physician involved or after such other procedures as are fair to the physician under the circumstances; and (4) in the reasonable belief that the action was warranted by the facts known after such reasonable effort to obtain facts and after meeting requirement 3 (42 U.S.C. § 11112).

[8] The civil money penalty (CMP) set forth in sec. 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 (sec. 1128B(b) of the Social Security Act) and the doctor self-referral prohibitions of the Act (sec. 1876 of the Social Security Act; Office of Inspector General 1999).

The Gainsharing Demonstration took place in two waves, which both applied only to doctors treating admitted Medicare patients. The initial phase took place in twelve New Jersey hospitals from July 1, 2009, to July 1, 2012. Eight of the original twelve hospitals opted to extend the program through March 31, 2013. Based on the reported success of the Gainsharing Demonstration, the New Jersey Hospital Association applied for and secured approval for a second, larger demonstration program under the ACA's Bundled Payments for Care Improvement (BPCI) initiative (Agency for Healthcare Research and Quality 2014). On April 1, 2013, the program was renamed the BPCI Model 1 program and was expanded to 23 hospitals (for simplicity, I refer to both the first and second wave as the Gainsharing Demonstration throughout the paper). Figure 1 shows that the participating hospitals in each wave are scattered around the state and are thoroughly interspersed with nonparticipating hospitals.

While I do not have data on take-up, anecdotal evidence suggests high physician participation in the demonstration. There is no reason for an eligible doctor to abstain, as there is no change in the process or form of payment, no additional paperwork, and no risk. Doctors are only rewarded for improvement and not punished for stagnation or increasing

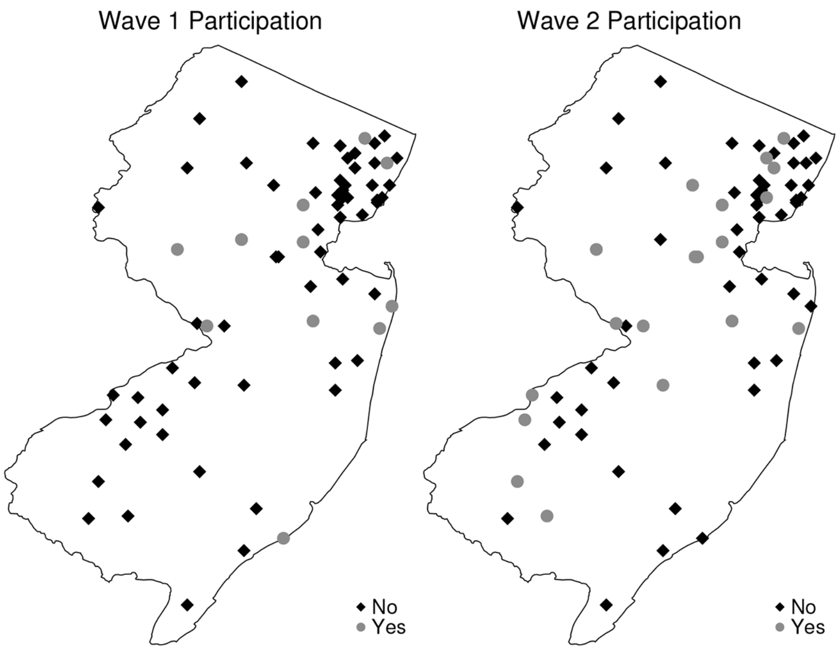

FIG. 1.—Hospital locations. Black diamonds are for hospitals that did not participate; gray circles are for hospitals that took up the bonuses. *Left*, participation in the first wave; *right*, participation in the second wave.

costs.[9] While many providers are involved in patient care, only the responsible doctor is eligible to receive a bonus under the Gainsharing Demonstration. For medical cases, this is the attending doctor, and for surgical cases, it is the surgeon. As doctors could only receive bonuses when treating admitted patients, language was included in the demonstration that total admissions could not rise under the program, though it was unclear how this would be enforced.[10]

1. Bonus Calculation

The bonus a doctor receives from the hospital through the Gainsharing Demonstration for treating an eligible (admitted and covered by Medicare) patient is calculated in three steps. First, patients are divided into types based on their diagnosis and how sick they are, using 3M's All Patient Refined Diagnosis Related Group (APR-DRG) system (for example, one type would be "hip joint replacement, severity of illness level 2"). Second, a maximum bonus is assigned to each patient type. All doctors face the same maximum bonus for treating patients of the same type. Third, this maximum bonus is scaled according to whether and how much the doctor reduces hospital costs for their patient relative to preprogram hospital costs for their patient's type in New Jersey. A hypothetical bonus calculation example is presented in figure 2. In this example, three doctors treat three patients with the same type, but receive different bonuses based on the costs of the treatment they provide.

The maximum bonuses are calculated using hospital cost data from before the program started (the base year was 2007 for the original demonstration and 2011 for the expansion).[11] The maximum bonus for

[9] While some alternative payment model pilot programs also include the potential for penalties if targets are not met, the addition of downside risk for doctors would not counteract incentives to sort healthier patients into participating hospitals, and could make these incentives even stronger.

[10] For example, the program tracked several parameters for unusual changes; a 10% increase in physician-level admissions would be considered unusual. But there is no indication of what would happen if this change were observed. Physicians with dual admitting privileges were capped at their prior-year patient volume at the participating hospital for incentives received from that hospital. However, a forward-looking doctor could increase admissions and wait a year, so it is not clear how binding this provision would be if a doctor wanted to increase admission volume.

[11] Theoretically, a participating hospital could artificially increase costs after taking up the program in order to decrease the average size of the bonus payments made to doctors. However, it is unlikely that a hospital would respond to a voluntary bonus program by changing its list charges, which could hurt its bargaining position with respect to other hospitals (fewer than half of New Jersey hospitals participated in the program). Likewise, if doctors could predict which hospitals would take up the expansion and what the base year would be, doctors could try to manipulate their charges to increase future bonus payments. However, there is no evidence that doctors at hospitals which took up the expansion proactively increased their charges in 2011, which would eventually become the rate year (see fig. A1).

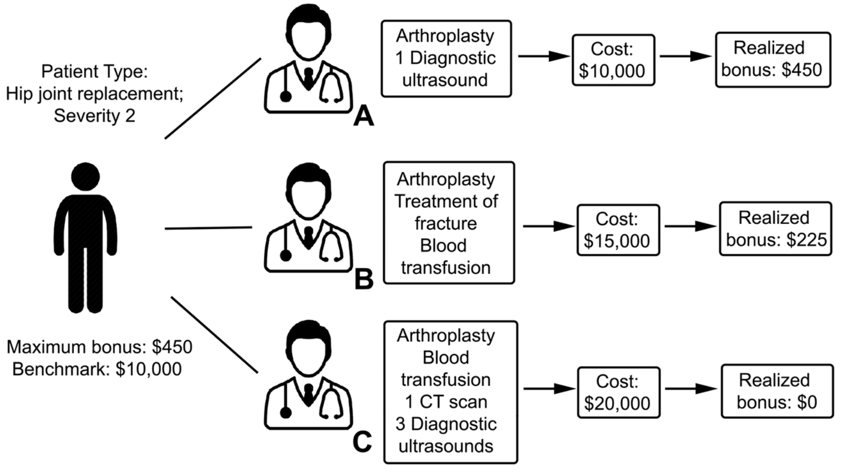

FIG. 2.—Hypothetical bonus calculation. Cost refers to the hospital treatment costs.

treating a patient type is defined as one-tenth of the average deviation from the 25th percentile of the hospital cost distribution in the state of New Jersey for that patient type in the base year. To this end, a third party calculated four maximum bonus amounts for each diagnosis (APR-DRG), depending on the severity of the patient's illness (SOI). The four severity of illness categories capture the fact that the same diagnosis (e.g., "peptic ulcer and gastritis") may be more or less serious depending on a patient's age and comorbidities.[12] 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, available online). An example of maximum bonuses for two particular APR-DRGs is given in table A1 (tables A1–A10 are available online), and the distribution of maximum bonuses is shown in figure 3.[13]

While the formulas for calculating the maximum bonuses are opaque, doctors were given quarterly "dashboards," which gave them real-time feedback on their performance and explicitly told them the amount of unearned incentive that they were leaving on the table (see fig. A2;

[12] As patient types are partially determined by the types and numbers of comorbidities recorded by the doctor, there is a potential for "upcoding"—doctors changing a patient's diagnosis to increase expected profit. I discuss this in more detail in sec. V.

[13] Hospitals described the distribution of incentive payments as a bell curve with most physicians receiving between $2,000 and $4,000 every 6 month period, with a minority of physicians receiving either small (a few hundred dollars) or relatively large ($10,000 or more) amounts. One facility self-reported payments to individual physicians of about $29,000 for a single 6 month payment period (Greenwald et al. 2014). For a comparison of my sample of the discharge data with published statistics about program participation and bonus sizes and distributions, see sec. A.3 of the appendix.

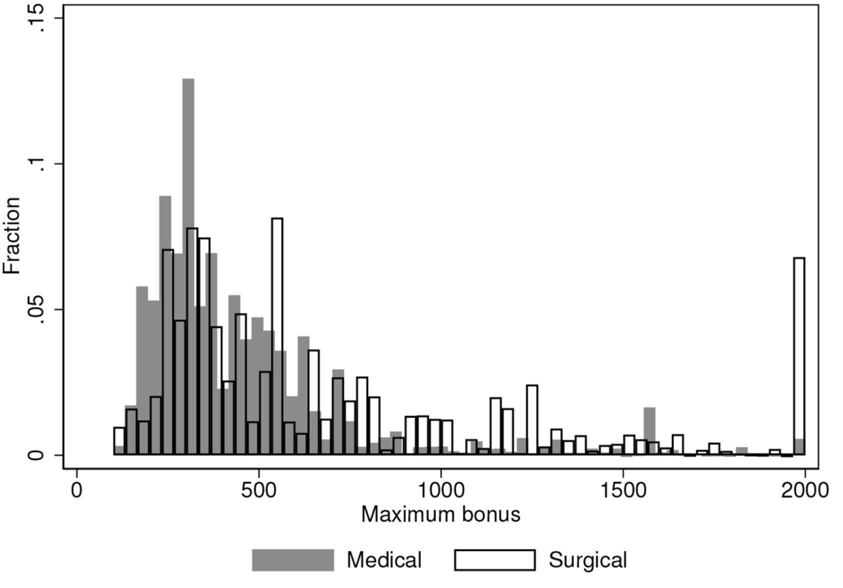

FIG. 3.—Distribution of maximum bonuses. Each observation is a Medicare beneficiary's inpatient visit to a general medical/surgical hospital in New Jersey from 2006 to 2013.

figs. A1–A10 are available online). Therefore, it is reasonable to expect doctors to quickly become familiar with which types of patients generated large bonuses.

The rationale behind the formulas used to calculate maximum bonuses is that high cost variance within a diagnosis is a red flag, and indicates the existence of high-cost patients who could be treated more cheaply. The bonuses are designed to make reducing the treatment costs of patients in diagnoses with high cost variance especially profitable for doctors. However, waste generated by unnecessary treatment is just one explanation for the underlying cause of cost variation. Alternatively, high cost variance within a group of patients could be due to disease pathophysiology, rather than doctor behavior.

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

2. Characteristics of Participating Hospitals

The hospitals that formed the demonstration and its expansion are similar to other New Jersey hospitals, on average. A cap of 12 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 12 participants to represent New Jersey hospitals as a whole.

As can be seen in table 1, the selection process appears to have been successful. The main difference between participating and nonparticipating hospitals—especially in the first wave—is that hospitals participating in the program have more Medicare patients and fewer Medicaid patients on average. In the second wave, however, more hospitals took up the program, and some of these differences disappeared. The demographics of the areas in which participating and nonparticipating hospitals are located are also similar on average, and figure A3 shows nearly indistinguishable patterns of local economic conditions between the municipalities of participating and nonparticipating hospitals.

Despite the fact that participating and nonparticipating hospitals are similar on observable characteristics, the selection of hospitals into the bonus program is nonrandom. Larger hospitals with more Medicare patients are more likely to participate, and these hospitals may be on different trajectories than nonparticipating hospitals. In the main analysis, I show graphical evidence that suggests that the two groups of hospitals are on parallel trajectories with respect to average costs. I also directly control for an important source of between-hospital variation by including doctor fixed effects.

The literature on hospital market structure suggests that nonprofit and for-profit hospitals, as well as nonprofit hospitals competing with for-profit hospitals, may respond differentially to changing financial incentives (Duggan 2000, 2002; Dafny 2005; Dafny and Dranove 2009). However, nearly all hospitals in New Jersey are private nonprofits; in 2008, just three of New Jersey's general medical and surgical hospitals were for-profit hospitals. Thus, I am unable to examine the possibility of differential responses to financial incentives by hospital organizational structure.

III. Conceptual Framework

To formalize how the bonuses should affect doctor decision making, I present a stylized model of the incentives and choices faced by doctors working in a hospital setting. I consider a doctor who works in two hospitals and must decide whether a patient is admitted, where to send the patient, and how much care to provide. The model focuses on the decisions most directly affected by the bonuses, abstracting from other potential decision margins such as the amount of effort expended or the quality of care provided. First, I describe the outcome when neither hospital offers a cost reduction bonus. Next, I introduce the cost reduction bonuses to one of the hospitals in the model. Finally, I compare how the doctor's decisions change as a result of the introduction of the bonuses.

A. The Setup

The model consists of one doctor treating a population of patients with mass 1, where all patients are within a single diagnosis–severity of illness type. I assume that the type is exogenously defined, though I will examine the validity of this assumption empirically. For each patient, the doctor must make three decisions: whether a patient is admitted, $A \in \{0, 1\}$, which hospital is attended, $H \in \{0, 1\}$, 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 $\bar{\beta}$.

The doctor is a utility maximizer, and chooses H, A, and q to maximize a weighted average of her profit from treating the patient and the patient's utility from treatment, where the weight placed on profit is λ.[14] Doctors are paid a reimbursement rate, a, for each unit of care, q, provided to the patient. The payment, a, does not depend on the hospital choice or whether the patient is admitted. Thus, the doctor's profit from treating a patient is aq. A doctor's concern for her patient's welfare can be understood as altruism on behalf of her patients, or as the doctor acting to preserve her reputation.

The patient's utility from medical treatment is concave in q, with sicker patients (those with a higher β) benefiting more from medical care. The utility a patient derives from medical care is

$$\begin{cases} \beta q - \dfrac{b}{2} q^2 & \text{if } A = 0, \\[2ex] \beta q - \dfrac{b}{2} q^2 + \gamma q - C & \text{if } A = 1. \end{cases} \tag{1}$$

The key assumption is that patients have a bliss point in q. Care provided past this preferred q need not necessarily become physically harmful, but can be interpreted as patients facing a coinsurance and the opportunity cost of their time.

A patient's utility from treatment depends additionally on whether or not they are admitted. If a patient is admitted to the hospital, there are two opposing effects. On one hand, being admitted makes treatment

[14] By considering a representative doctor, the model abstracts from potential heterogeneity over λ.

TABLE 1
HOSPITAL CHARACTERISTICS

	Wave 1				Wave 2			
	Participation				Participation			
	No	Yes	Difference	p-value	No	Yes	Difference	p-value
Hospital characteristics (AHA survey):								
Nongovernment not-for-profit	.82	.83	.01	.94	.79	.90	.12	.87
Bed size code	5.59	6.00	.41	.95	5.48	6.05	.57	.78
ER visits	52,388	54,469	2,081	1.00	49,453	59,447	9,994	.74
Hospitals in a network	.54	.58	.04	1.00	.56	.53	.04	.96
CBSA type: Metro (population 50,000+)	.20	.17	.03	.99	.17	.24	.07	1.00
Medicare inpatient discharges	6,564	8,449	1,885	.68	6,509	7,750	1,242	.89
Medicare days	40,085	50,011	9,926	.70	39,156	47,613	8,457	.77
Medicaid inpatient discharges	2,585	1,950	635	.89	2,434	2,525	91	.87
Medicaid days	14,186	8,575	5,611	.57	13,440	12,471	969	.99
Observations	51	12			42	21		
Hospital characteristics (NJ discharge data):								
Fraction Medicare (inpatient)	.51	.50	.01	1.00	.51	.52	.01	1.00
Number of doctors (Medicare/inpatient)	72	103	31	.65	68	96	28	.26
Observations	60	12			49	23		

Hospital area demographics (census):								
Fraction female	.51	.51	.00	.98	.51	.51	.00	1.00
Fraction white	.54	.49	.05	1.00	.56	.48	.08	.92
Fraction black	.16	.20	.04	.99	.16	.18	.02	1.00
Fraction Hispanic	.22	.19	.03	1.00	.20	.24	.05	.97
Fraction age 65+	.13	.12	.01	.97	.13	.11	.02	.30
Number of other hospitals in HSA	.85	.67	.18	1.00	.86	.74	.12	1.00
Observations	60	12			49	23		

NOTE.—General medical and surgical hospitals. Hospital information from the American Hospital Association annual survey data (2008) and the New Jersey hospital discharge data (2008); hospital area demographics from the 2010 census and assigned to hospitals at the county subdivision level. Medicaid/Medicare days are the total number of inpatient hospital days used by beneficiaries; the number of doctors is the number of doctors observed in the New Jersey discharge data with at least one Medicare inpatient discharge; hospital service areas (HSAs) are local health care markets for hospital care (Dartmouth Atlas). The p-values are adjusted for multiple-hypothesis testing for multiple outcomes following List, Shaikh, and Xu (2016) and using the *mhtexp* Stata command; without this correction, one difference is statistically significant at the 10% level (Medicaid days for wave 1) and two at the 5% level (fraction aged 65+ and number of doctors for wave 2).

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 patient's 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 her patient, as well as the difference in her compensation.

When a doctor is indifferent between hospitals, I assume she randomly assigns patients such that they have an equal probability of going to each hospital.[15]

B. 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:

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

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

[15] The randomization can be interpreted as patients having a slight preference for the closest hospital, and patients being evenly distributed across space. Doctors could assign patients such that any proportion goes to each hospital; I use 50-50 to keep examples simple.

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 $V_1(\beta) \geq V_0(\beta)$ and never admit patients when $V_0(\beta) > V_1(\beta)$; β^A is defined as the sickness level at which $V_0(\beta) = V_1(\beta)$. 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, $(\bar{\beta} - \beta^A)/2$ patients are admitted at each hospital. A formal proof is presented in the appendix.

C. With Bonuses

Next, I consider what happens when cost reduction bonuses of the form used in the Gainsharing Demonstration are introduced at hospital 1.

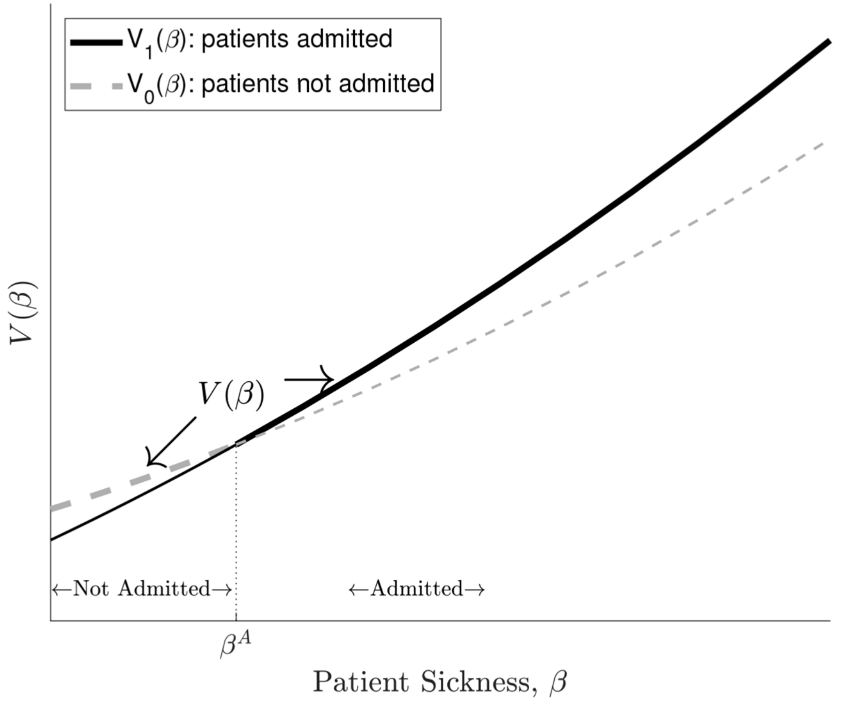

FIG. 4.—Doctor's utility as a function of β: without bonus. The bold line sections show the optimal decision rule for admission as a function of β, without bonuses.

Adding the bonuses only changes the framework described above in one way—doctors' profits change at the bonus hospital:

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

If an admitted patient is treated at the bonus hospital, the doctor is now eligible to receive a cost reduction bonus: $\max\{\alpha_0 - \alpha_1 q, 0\}$. The bonus is decreasing in the amount of care provided, q, but is never negative. The maximum bonus for the diagnosis–severity of illness group is α_0, and α_1 represents how quickly the bonus decays as q increases. Everything else remains the same, including the number of patients admitted to the bonus hospital, $\beta' = (\bar{\beta} - \beta^A)/2$.[16] Doctors are constrained by the number of patients admitted at the participating hospitals in the absence of the bonus program, as the program included language restricting doctors from increasing overall admission. Even if the rules had not mentioned admission levels, holding admission fixed is equivalent to introducing capacity constraints—assuming hospital capacity does not change in response to the program. Doctors can, however, change which patients are admitted and where they are treated. Past research has shown that patients typically accept their doctors' recommendations (Manning et al. 1987). Since all patients affected by the program are covered by Medicare and all hospitals accept Medicare, it seems reasonable to assume most patients would agree to use whichever hospital is recommended by their doctor.

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

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

[16] 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.

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

Whether or not there are bonuses, the admitted patients are always those with the largest (positive) difference between the utility a doctor receives from admitting them and not admitting them. Before the bonuses are introduced, this difference is largest for the sickest patient ($\beta = \bar{\beta}$) and is increasing in β. The introduction of the bonuses at hospital 1, however, eliminates this monotonicity. The cost reduction bonuses increase the doctor's profit from admitting healthy (low β) patients, up until the point where a patient is sick enough that quantity of care chosen is too high to generate a bonus (represented by the dash-dotted line in fig. 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.

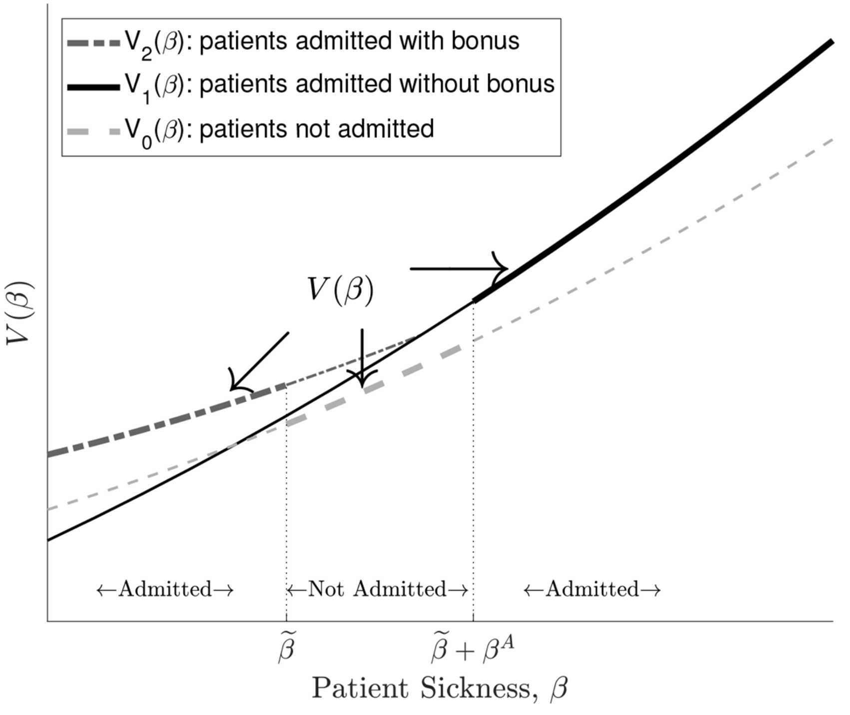

FIG. 5.—Doctor's utility as a function of β: with bonus. The bold line sections show the optimal decision rule for admission as a function of β, with bonuses.

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

After the bonuses are introduced, doctors would like to admit all patients [see fig. 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 $\beta' = (\bar{\beta} - \beta^A)/2$ 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 $\tilde{\beta}$. They will also admit the sickest $\bar{\beta} - (\tilde{\beta} + \beta^A)$ patients, randomizing over hospital choice such that they admit β' total patients at each hospital. The patients with β-values 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 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 "preperiod"), and all are admitted when the bonus is introduced (the "postperiod"). On the other hand, many "medium sick" patients with $\beta \in [\tilde{\beta}, \tilde{\beta} + \beta^A]$ are admitted in the preperiod, and are not admitted in the postperiod. Second, the bonuses cause sorting. After their introduction, doctors send the healthiest patients exclusively to the bonus hospital. Previously, the nonbonus hospital would have received some of the healthier patients, whereas now it only gets patients with $\beta > \tilde{\beta} + \beta^A$.

The bonuses' effect 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 postperiod for bonus-generating patients ($\beta \in [0, \tilde{\beta}]$) is ambiguous.

Whether the quantity of care provided for the bonus-generating patients is higher or lower than the counterfactual of neither hospital offering a bonus is determined by the relative size of γ and α_1. For more details, see the 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 fig. 6).

IV. 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—in terms of both whether and where patients are admitted. The bonuses may also impact the quantity of services provided, though the direction and magnitude are ambiguous. The relative sizes of these three effects, and whether the program ultimately decreases costs, are empirical questions which I address in the remainder of the paper.

A. 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.[17] I also create doctor identifiers using the recorded license numbers of doctors and surgeons. The final file includes unique identifiers for both patients and doctors, allowing me to track them over time and across all hospitals in New Jersey. The ability to follow both patients and doctors is often lacking in medical records and is an important strength of this paper. The discharge data also include admission and discharge dates, all diagnoses and procedure codes, payer information, patient demographic information, and list charges. To these data, I add information on hospitals from the American Hospital Association annual survey and Medicare's cost-to-charge ratio series.

The main analysis sample consists of all visits in which a Medicare beneficiary was admitted to a New Jersey hospital.[18] I restrict the sample to

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

[18] Medicare beneficiaries are those for whom fee-for-service Medicare is listed as the primary payer, to match eligible admissions in the bonus program rules. Medicare Advantage beneficiaries are excluded. No restriction is made based on the existence of secondary insurance, as the existence and type of secondary insurance could be something that doctors could sort on, or be correlated with characteristics used by doctors to sort healthier patients into participating hospitals (Medicare beneficiaries with private secondary insurance are the healthiest, Medicaid dual eligibles are the sickest, and those without secondary insurance are in between).

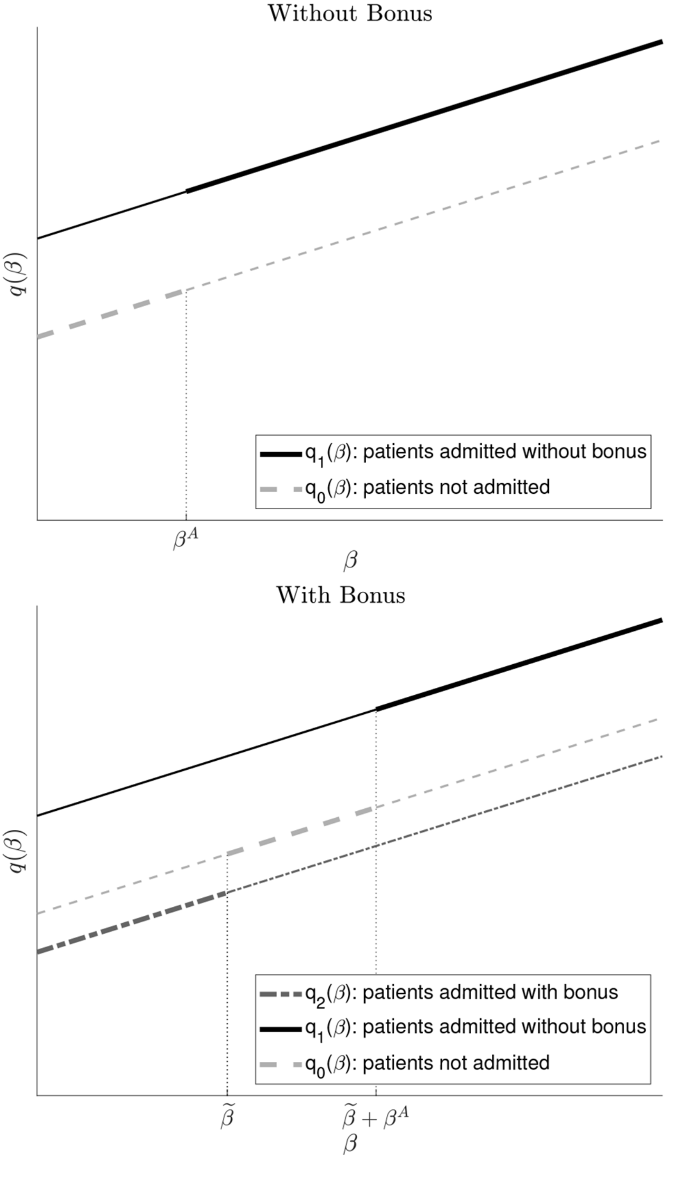

FIG. 6.—Optimal quantity of care as a function of β. The bold line sections show the optimal quantity of care provided as a function of β, both with and without bonuses.

patients seen in general medical and surgical hospitals that were open throughout the sample period. This restriction mainly excludes psychiatric and rehabilitation facilities, which were not targeted by the program. Visits to doctors with very few admitted patients over the sample period were also dropped, as these doctors likely did not have enough patients to qualify for the bonus program.

The full sample consists of approximately 1.2 million medical visits and 530,000 surgical hospital visits. Of these, 69% of medical and 27% of surgical visits were admitted through the emergency room. The route of admission is important, as doctors have no scope to send a patient already in an emergency room to a different hospital based on their latent health. They could, however, become more likely to admit healthier emergency room (ER) patients in participating hospitals. In the analysis that follows, I will use route of admission to help disentangle sorting and admission decision responses.

Summary statistics are shown in table 2, both for the whole sample of admitted Medicare beneficiaries and by whether the hospital ever took up the bonus program. Columns 1 and 2 of panel A show that the admitted Medicare beneficiaries are predominately white, with an average age of 75, and are slightly over half female. The doctors treating these patients worked at 2.3 hospitals on average (with a median of two hospitals), with 37% ever working in both a participating and a nonparticipating hospital (col. 1 of panel B). The average maximum bonus a doctor could earn for treating a surgical patient was $702, versus $444 for a medical patient. While few doctors receive the whole maximum bonus, even taking home half of these amounts would be a windfall (for comparison, in 2012 Medicare paid doctors $675.99 to repair a knee ligament; see Smith 2012).[19]

Patients treated in hospitals that did and did not take up the policy look similar for the most part, with the main difference being that patients in hospitals that did not take up the policy are more likely to have gone through the ER, on average. These nonparticipating hospitals also performed fewer total diagnostic tests. To the extent that I can infer from the discharge data, doctors working in participating and nonparticipating hospitals look very similar.[20] Doctors working in both hospital types differ mostly in that they work in slightly more hospitals on average, which is not surprising, given that in order to work in both hospital types they have to work in at least two hospitals.

My primary measure of latent patient health is the Charlson comorbidity index (CCI). This index is designed to predict the 1 year mortality for hospital inpatients based on the presence of comorbid conditions, with a higher score on the index denoting a sicker patient. The index is computed on the basis of the presence of 17 conditions, each weighted by

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

[20] All I observe about the physician in the discharge data is a unique identifier, so unfortunately I cannot look at more detailed physician-level characteristics, such as specialty. It is very difficult to back out information about the doctor by looking at the characteristics of their patients, as only one doctor (or a doctor and a surgeon) is listed on the discharge record, despite the fact that many doctors are likely involved in caring for the patient.

TABLE 2
MAIN SAMPLE CHARACTERISTICS

	A. Medicare Patients					
	All		Ever Bonus Hospital		Never Bonus Hospital	
	Medical (1)	Surgical (2)	Medical (3)	Surgical (4)	Medical (5)	Surgical (6)
Characteristics:						
Age	75.37	74.51	75.47	74.50	75.28	74.52
White	.785	.828	.795	.843	.777	.810
Black	.136	.099	.139	.090	.134	.108
ER revenue	.692	.268	.605	.222	.767	.322
Woman	.572	.536	.560	.521	.581	.554
In policy hospital	.463	.536	1.000	1.000	.000	.000
Outcomes:						
Latent health:						
Charlson index	2.67	1.81	2.64	1.80	2.70	1.81
Charlson index = 0	.294	.441	.294	.435	.293	.448
Surgical risk index	.653	.385	.645	.386	.660	.383
Surgical risk index = 0	.601	.742	.602	.741	.599	.743
Costs and quantity:						
Length of stay	5.78	6.92	5.83	6.73	5.74	7.14
CT scan	.056	.036	.067	.041	.046	.031
MRI	.023	.013	.023	.013	.023	.012
Diagnostic ultrasound	.035	.058	.046	.067	.025	.047
Any imaging	.118	.191	.138	.203	.101	.178
Total costs	9,360	18,213	9,868	18,874	8,823	17,290
Observations	1,184,413	533,227	548,320	285,554	636,093	247,673

	B. Doctors Treating These Patients			
	All (1)	Works in Participating Hospital (2)	Works in Not Participating Hospital (3)	Works in Both Hospital Types (4)
Average number of patients	472	482	498	538
Average number of hospitals	2.32	2.61	2.66	3.49
Median number of hospitals	2.00	2.00	2.00	3.00
Ever in bonus hospital	.69	1.00	.55	1.00
Ever in both types	.37	.54	.55	1.00
Average maximum bonus (medical)	444	448	446	452
Average maximum bonus (surgical)	702	676	711	684
Fraction surgeons	.60	.60	.62	.65
Observations	3,639	2,514	2,475	1,350

Note.—Admitted Medicare patients in general medical/surgical hospitals (2006–2013). “In policy hospital” is fraction treated at hospitals that ever participate in the program.

the associated risk of death, and has been widely validated.[21] Furthermore, the Charlson index has been shown to be strongly predictive of hospital resource utilization, which makes it uniquely well suited for measuring latent health as it relates to cost of treatment (Charlson et al. 2008).

In order to measure a patient's latent health (rather than the acute event that brought them to the hospital), I construct a "leave out" version of the Charlson comorbidity index, which exploits the time-series dimension of the data. The leave-out index uses data on all prior hospital visits made by each patient (regardless of admission), excluding any diagnoses recorded during the current visit.[22] Thus, the index is a measure of the latent health of patients that are known to the doctor (or at least correlated with information known to the doctor), but is not used in the bonus formula.

While the Charlson comorbidity index is a useful summary measure of patient health, it was developed to measure the mortality risks of medical inpatients, and may be less sensitive as a measure of preoperative physical status. Therefore, I supplement the Charlson index with an index based on general surgical risk factors. Specifically, I use all conditions in the universal surgical risk calculator that can be identified via ICD-9-CM diagnosis codes (Best et al. 2002; Bilimoria et al. 2013).[23] I construct a surgical risk factor index that is the sum of these indicators, again using prior hospital visits, and leaving out diagnoses recorded during the index visit. A higher score on the surgical risk factor index represents a sicker patient.

Patients admitted under a medical diagnosis have a higher disease burden on average compared to surgical visits, as measured by both the average Charlson index and surgical risk factor index, as well as the proportion scoring a zero on each index. Patients admitted under a medical diagnosis are about a year older, more likely to be black, and much more likely to be admitted through the emergency room. These patients are

[21] The Charlson comorbidity index is a weighted sum over the following conditions (weights are in parentheses): acute myocardial infarction (1), congestive heart failure (1), peripheral vascular disease (1), cerebrovascular disease (1), dementia (1), chronic pulmonary disease (1), rheumatologic disease (connective tissue disease; 1), peptic ulcer disease (1), mild liver disease (1), diabetes without complications (1), diabetes with chronic complications (1), hemiplegia or paraplegia (2), renal disease (2), cancer (2), moderate or severe liver disease (3), metastatic carcinoma (6), AIDS/HIV (6).

[22] If a patient appears just once in the data, he or she is assigned a zero. This strategy will introduce some measurement error, as a patient could have a serious disease but not a previous hospital visit. However, given the seriousness of the conditions used in the Charlson comorbidity index and the high disease burden of Medicare patients, I expect the vast majority of patients with one of these conditions will visit a hospital multiple times. Of visits eligible to receive bonuses, just 3.4% of medical visits and 5.6% of surgical visits are to patients seen only once in the data. Table A4 repeats the main analysis for a sample excluding these single-visit patients.

[23] These conditions are disseminated cancer, diabetes, hypertension, dyspnea, COPD, acute renal failure, and ascites or congestive heart failure within 30 days preoperatively. The exact ICD-9-CM codes used to define these indicators are listed in table A5.

also more varied. There are 161 medical diagnosis groups, compared to 117 surgical diagnosis groups, and just 10 diagnosis groups make up half of all patients admitted under surgical codes (table A2 lists the most common diagnosis groups for medical and surgical patients in the sample).

To examine whether the bonuses changed procedure use or lowered costs, I also look across several measures. The first two are summary measures of resource use: length of stay and total costs. Length of stay is defined as the number of nights spent in the hospital, and is often used to proxy for the intensity of care provided during the visit. The total hospital costs incurred during a visit are estimated using the total list charges reported in the discharge data, deflating them by Medicare's hospital-year-level cost-to-charge ratio, and then converting them to real 2010 dollars. The Medicare cost-to-charge ratio is explicitly designed to translate list charges into an estimate of the resource cost of inpatient care. Surgical cases are more resource intensive on both measures, spending on average an extra day in the hospital and incurring nearly twice the costs of medical cases.

In addition to summary measures of resource use, I look specifically at the use of diagnostic imaging to proxy for the use of unnecessary procedures. While it is difficult to pinpoint any specific test as unnecessary, there is widespread agreement that diagnostic imaging is overused (Hillman and Goldsmith 2010; Abaluck et al. 2015).[24] 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.[25]

B. Empirical Strategy

The main challenge in identifying the effect of the cost reduction bonuses on doctor decision making is that participating hospitals are different from hospitals that did not take up the program. In particular, hospitals that seek out a program designed to reduce wasteful spending may be on different treatment cost trajectories than those that do not. Figure 7

[24] 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).

[25] 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).

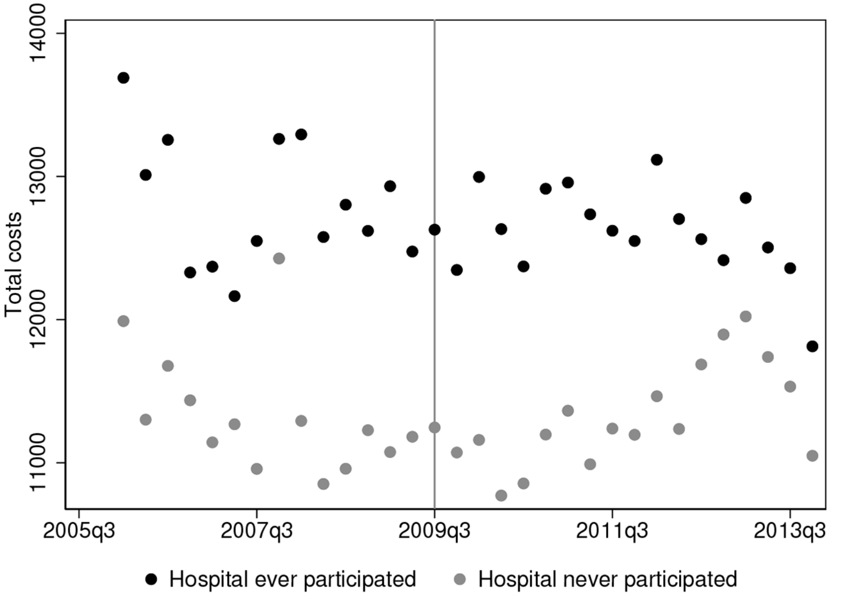

Fig. 7.—Average costs by hospital participation in demonstration. Average quarterly costs for admitted Medicare patients, averaged across hospitals that ever participated in the Gainsharing Demonstration, versus hospitals that never participated. Costs are deflated by Medicare cost-to-charge ratios.

plots average treatment costs for admitted Medicare patients across hospitals by future program participation, and shows that hospitals that took up the program had higher average costs than those that did not. The level difference in costs reflects the fact that participation was not randomly assigned; hospitals with higher average costs initially had more to gain from a program incentivizing doctors to reduce costs. However, there is no evidence of differential pretrends between the two groups with respect to costs.

Despite no evidence of differential pretrends in costs across the hospital groups, I include doctor fixed effects to further control for time-invariant physician-level characteristics. An important strength of this specification is that I do not have to worry about different compositions of surgeons and physicians across hospitals driving the results. However, a weakness of this strategy is that it will not be able to detect any responses to the bonuses that occur in both hospitals, for example if exposure to the bonuses causes doctors to change their practice style in all hospitals in which they work. For now, I rule out this type of behavior. I return to this issue with an alternative identification strategy in section VI, which examines total resource use as a function of program exposure.

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

$$\text{outcome}_{idht} = \beta_0 + \beta_1 \text{policy}_{ht} + \beta_2 X_{it} + \lambda_t + \lambda_h + \lambda_d + \epsilon_{idht}, \qquad (3)$$

where i stands for individual, d for doctor, h for hospital, and t for time (in quarters). Here policy_{ht} is an indicator for whether the visit occurred in a participating hospital when the bonus program was in effect, and the coefficient of interest is β_1. The patient characteristics included in X_{it} vary by specification. They are omitted when looking at the effect of the bonuses on latent health, as the goal is to measure the effect of the bonuses on the composition of admitted patients. When considering the effect of the bonuses on costs and procedure use, however, I want to control for changes in patient composition. Thus, these regressions include age, sex, race, and measures of latent health. Hospital, quarter-by-year, and doctor fixed effects are also included in all regressions (λ_h, λ_t, and λ_d). In some specifications, patient-type (APR-DRG by SOI pairs) fixed effects are also included. Standard errors are clustered at the hospital level.

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

My empirical strategy takes as given physician-hospital relationships. In order to test this assumption, I plot the fraction of doctors in each quarter that are either first seen practicing in a participating hospital, or last seen practicing at a nonparticipating hospital. As can be seen in figure A7, there is no evidence that the bonus program caused physicians to initiate contracts with certain hospitals or to terminate them with others.[26]

V. Results

A. *Effects of Bonuses on Latent Health of Admitted Patients*

Doctors admit healthier bonus-eligible patients in participating hospitals in response to the program. Figure 8 displays the effect of the bonus

[26] While there was a waiting period for new doctors before they were eligible to receive bonuses, this lack of response may still be somewhat surprising if doctors are forward looking and expected the policy to continue. Even if this was the case, however, there may not have been much scope for doctors to respond to the bonuses on this dimension, as many likely already have privileges at the hospitals relevant to their practice.

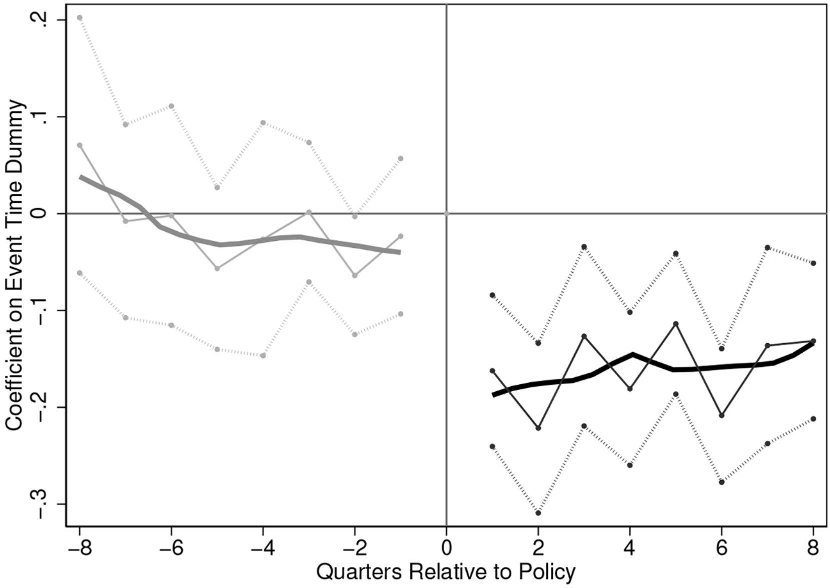

FIG. 8.—Healthier patients sent to participating hospitals: Charlson comorbidity index for medical patients. Event study plot created by regressing the Charlson comorbidity index on a full set of event time indicators, as well as hospital, quarter, type (APR-DRG by SOI), and doctor fixed effects. Plotted are the coefficients on the event time indicators, which show the time path of the Charlson comorbidity index of patients admitted at participating hospitals, relative to nonparticipating hospitals, before and after the program went into effect. The dotted lines show 95% confidence intervals, where standard errors are clustered at the hospital level. Time is normalized relative to the quarter in which the hospital took up the bonus program.

policy on the average Charlson comorbidity index of medical patients in event time, where the implementation of the policy is normalized to $t = 1$. The event time specification is identical to equation (3), except the binary policy variable is replaced with quarterly event time dummies denoting the number of quarters before and after a hospital took up the policy.[27] After the policy is introduced, there is a clear drop in the average comorbidity burden of Medicare patients with nonsurgical diagnoses.[28] The medical patients admitted at participating hospitals under the Gainsharing Demonstration became healthier, relative to the patients at nonparticipating hospitals.

[27] The model is fully saturated; hospitals that never participated are assigned an event time of −8.

[28] As a placebo check, fig. A5 shows that there is no analogous improvement in latent health for "near Medicare" patients (aged 50 to 64) treated under the program. These patients have many of the same health problems as the Medicare population, but are too young to quality for Medicare coverage.

The event time result in figure 8 is presented in regression form in table 3. The bonuses are associated with a decrease in the average Charlson comorbidity index of medical patients of 0.09. To put the magnitude of this change in perspective, decreasing the index by a tenth is associated with decreases in in-hospital mortality of 3% to 6% across seven OECD countries (Quan, Li, and Couris 2011).[29] The bonuses are also associated with an increase in the probability that admitted medical patients have a Charlson score of zero; these patients can be thought of as the ones in the best overall health. The results are similar with and without type fixed effects (APR-DRG by SOI), implying that the change in latent health occurs within the types over which the bonuses are defined.

The effect of the bonuses on latent health is much larger and more precisely measured for patients admitted under medical diagnoses, compared to surgical cases. Columns 1 and 2 of panel B show little effect of the bonuses on the average latent health of surgical patients treated under the program; the coefficients are a quarter of those for medical patients in columns 1 and 2 of panel A. However, the point estimates in columns 3 and 4 of panel B suggest that surgical patients treated in participating hospitals are somewhat more likely to have a Charlson score of zero.

The results in table 3 suggest that doctors were able to game the bonuses more effectively for medical patients than for surgical patients. An alternative interpretation, however, is that the diagnoses included in the Charlson index are more closely tied to the costs of medical conditions than surgical ones. While it is difficult to completely rule out this alternative interpretation, table A6 uses a surgical risk index as an alternative measure of latent health, and the results are similar to those in table 3. Finally, table A7 looks at the effect of the bonuses on a wide range of individual comorbidities associated with more complex and expensive patients. Medical patients admitted at participating hospitals are significantly less likely to have a wide range of individual chronic conditions, but again this is not the case for surgical patients; the point estimates are uniformly negative, but small and not statistically distinguishable from zero.

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

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

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

	Charlson (1)	Charlson (2)	CCI = 0 (3)	CCI = 0 (4)
	A. Medical Patients			
Policy	−.094***	−.093***	.005*	.006*
	(.028)	(.027)	(.003)	(.003)
Type fixed effects	No	Yes	No	Yes
Mean dependent variable	2.673	2.673	.294	.294
Clusters	74	74	74	74
Observations	1,184,413	1,184,413	1,184,413	1,184,413
	B. Surgical Patients			
Policy	−.023	−.023	.006	.007*
	(.037)	(.032)	(.004)	(.004)
Type fixed effects	No	Yes	No	Yes
Mean dependent variable	1.807	1.807	.441	.441
Clusters	73	73	73	73
Observations	533,227	533,227	533,227	533,227

Note.—Quarter-by-year, doctor, and hospital fixed effects included in all regressions. Standard errors clustered at the hospital level.

* $p < .1$.

*** $p < .01$.

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

1. Admission

In order to isolate changes in which patients are admitted from changes in where patients are admitted, I look at two subsamples of the data for which sorting is not possible: patients who were admitted through the emergency room, and patients whose doctors work in just one hospital. Patients admitted through the ER cannot be sorted in response to the bonuses, as emergency room doctors cannot send a healthier than average ER patient to a different hospital. Similarly, doctors cannot sort patients between hospitals if they only have admitting privileges in one.

Table 4 shows that even when doctors are unable to sort patients between hospitals, admitted Medicare patients with medical diagnoses are healthier when admitted under the program. Thus, doctors must be admitting healthier patients in response to the cost reduction bonuses. For medical patients, the magnitudes are similar across the two samples,

TABLE 4

Effect of Bonuses on Latent Health: No Sorting

	Medical Patients				Surgical Patients			
	Charlson (1)	Charlson (2)	CCI = 0 (3)	CCI = 0 (4)	Charlson (5)	Charlson (6)	CCI = 0 (7)	CCI = 0 (8)
	A. Admitted through Emergency Room							
Policy	−.074**	−.082***	.004	.005	−.067	−.050	.014*	.012*
	(.033)	(.023)	(.003)	(.004)	(.046)	(.039)	(.008)	(.007)
Type fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
Mean dependent variable	2.622	2.622	.302	.302	1.887	1.887	.446	.446
Clusters	72	72	72	72	71	71	71	71
Observations	819,569	819,569	819,569	819,569	143,020	143,020	143,020	143,020
	B. Single-Hospital Doctors							
Policy	−.122***	−.093**	.013***	.009**	−.070	−.063	.020**	.017*
	(.037)	(.037)	(.005)	(.005)	(.046)	(.040)	(.009)	(.009)
Type fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
Mean dependent variable	2.456	2.456	.315	.315	1.669	1.669	.463	.463
Clusters	58	58	58	58	58	58	58	58
Observations	247,039	247,039	247,039	247,039	91,716	91,716	91,716	91,716

Note.—Quarter-by-year, doctor, and hospital fixed effects included in all regressions. Standard errors clustered at the hospital level.

* $p < .1$.

** $p < .05$.

*** $p < .01$.

as well as to the main results in table 3. For surgical patients, the magnitudes are larger and more precisely estimated when the sorting margin is eliminated, suggesting that surgeons may primarily respond to the bonuses by changing the margin for admission rather than sorting patients between hospitals.

2. Sorting

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

It is still possible, however, that some of the change in patient health in this sample is due to changes in which patients are admitted, rather than sorting. To focus more tightly on the sorting channel, I further narrow the sample to patients in diagnoses that are nearly always admitted.[30] In these diagnoses, any changes in the latent health must be due to doctors sorting patients between hospitals. While the results are less precisely estimated, the point estimates in panel B of table 5 suggest that doctors do respond to the bonuses by sorting healthier patients into participating hospitals. However, these results are concentrated among doctors treating patients with medical diagnoses; there is no evidence that healthier surgical patients are sorted into participating hospitals after the bonuses go into effect.

While I cannot calculate the relative importance of each channel, it appears that doctors treating nonsurgical cases respond to the bonuses both by newly admitting healthier patients and by sorting patients across hospitals. Although the costs of doctors sorting patients between hospitals may be minimal (at least to patients and doctors), changing the pool of patients seen at each hospital may have important implications for hospital profitability. Furthermore, the fact that a healthier mix of patients is admitted when doctors are given bonuses for low-cost admissions is

[30] These are the top quarter of the main sample of medical and surgical patients, based on the average admission rate in each APR-DRG in the Medicare population. The average admission rate in the medical high admission rate sample is 93.3. The average admission rate in the surgical high admission rate sample is 99.8. Table A3 lists the high-admission APR-DRGs. Nonadmitted patients include both those that were seen in the ER and not admitted and those who got non-ER outpatient treatment without being admitted.

TABLE 5

EFFECT OF BONUSES ON LATENT HEALTH: EXCLUDING EMERGENCY ROOM

	MEDICAL PATIENTS				SURGICAL PATIENTS			
	Charlson (1)	Charlson (2)	CCI = 0 (3)	CCI = 0 (4)	Charlson (5)	Charlson (6)	CCI = 0 (7)	CCI = 0 (8)
	A. Excluding Emergency Room							
Policy	−.128**	−.092**	.008	.004	.003	−.005	.002	.004
	(.048)	(.043)	(.005)	(.004)	(.038)	(.033)	(.005)	(.004)
Type fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
Mean dependent variable	2.785	2.785	.275	.275	1.778	1.778	.439	.439
Clusters	73	73	73	73	72	72	72	72
Observations	364,844	364,844	364,844	364,844	390,207	390,207	390,207	390,207
	B. Excluding Emergency Room, High-Admission DRGs							
Policy	−.128*	−.103*	−.001	−.003	−.011	−.012	.003	.003
	(.065)	(.059)	(.006)	(.006)	(.032)	(.027)	(.008)	(.008)
Type fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
Mean dependent variable	3.322	3.322	.212	.212	1.364	1.364	.518	.518
Clusters	72	72	72	72	72	72	72	72
Observations	134,808	134,808	134,808	134,808	147,621	147,621	147,621	147,621

NOTE.—Quarter-by-year, doctor, and hospital fixed effects included in all regressions. Standard errors clustered at the hospital level.

* $p < .1$.

** $p < .05$.

worrying for two reasons. First, the healthier patients who are admitted under the program but not otherwise face substantial costs (both in time and money) from admission. Second, if capacity constraints bind, some sicker patients will not be admitted who would have been otherwise. For the sicker patients, the welfare effect of not being admitted depends on whether the admissions decisions were optimal in the preperiod, which is beyond the scope of this paper. However, if the admissions decisions were either optimal or too low in the absence of the bonuses, these patients could also be worse off on average from receiving less intensive treatment.

3. Upcoding

Both the model and the empirical results assume that the assignment of a patient into a particular type (characterized by a diagnosis and a severity of illness pair) is unaffected by the Gainsharing Demonstration. One might be worried about this assumption, as during the 1980s and 1990s many hospitals were accused of upcoding—exaggerating a patient's diagnosis to extract a higher reimbursement from Medicare. Dafny (2005) found that hospitals, and in particular for-profit hospitals, responded to a 1988 Medicare policy change that generated large diagnosis related group (DRG) price changes by upcoding patients in diagnosis codes with the largest price increases. Similarly, Silverman and Skinner (2004) found that between 1989 and 1996, the percentage point share of the most generous diagnosis groups for pneumonia and respiratory infections rose precipitously.

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

While doctors may not be able to change the diagnosis group, it is possible that doctors could respond to the Gainsharing Demonstration by trying to move their patients into higher severity of illness bins.[31] Influencing the severity of illness (SOI) designation should theoretically be

[31] However, upcoding would be more difficult here, as 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.

difficult (or at least indirect), as it is imputed by software and not recorded by the doctor. The only way doctors can affect the severity of illness is to change which secondary diagnoses are recorded on a patient's chart. While the link between any one comorbidity and the designation generated by the software is not clear, adding additional diagnoses to all patients could lead to higher average SOI designations. If doctors record more comorbidities 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. Upcoding, therefore, could generate similar patterns in the data to sorting.

There is no association in the data between the bonuses and the average severity of illness within APR-DRGs, however, suggesting that upcoding is not a concern in this context. The regressions reported in table 6 use the same empirical strategy outlined in equation (3), but with severity of illness as the dependent variable and APR-DRG fixed effects, rather than APR-DRG by severity of illness fixed effects. If upcoding occurred in response to the cost reduction bonuses, then within each APR-DRG, the patients admitted under the bonus policy should have a higher severity, on average. The introduction of the bonus program appears to have no effect on the average severity of illness, either in the full sample or in the subgroups used in the previous sections. Not only are the point estimates insignificant, they are very small, and not even consistently positive.

One interpretation of this null result is that doctors did not have sufficient information about how the software translated comorbidities into

TABLE 6
EFFECT OF PROGRAM ON POPULATION-LEVEL SEVERITY OF ILLNESS

	Full Sample (1)	ER Revenue (2)	High Admission (3)
	A. SOI: Medical Patients		
Policy	.001	.006	−.022
	(.020)	(.030)	(.030)
Mean dependent variable	2.291	2.259	2.689
Clusters	74	72	72
Observations	1,184,413	819,569	134,808
	B. SOI: Surgical Patients		
Policy	.009	−.004	.010
	(.015)	(.026)	(.021)
Mean dependent variable	2.052	2.204	2.066
Clusters	73	71	72
Observations	533,227	143,020	147,621

NOTE.—Quarter-by-year, doctor, hospital, and APR-DRG fixed effects also included, as well as dummies for age categories, sex, and race. Standard errors clustered at the hospital level.

severity of illness levels to successfully upcode. Another is that the proximity of the payer (the hospital) to the recipient (the doctor) in the Gainsharing Demonstration differs substantially from earlier settings in which upcoding 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. Either way, it does not appear that changes in the composition of APR-DRG cells as a result of upcoding are likely to be driving the observed changes in latent health.

4. Placebo Tests

To confirm that the association between the bonuses and latent health is not spurious, I conduct two placebo tests. First, I hold fixed the true hospital participation in the Gainsharing Demonstration, but randomly assign start dates for the program, and repeat the main within-doctor regressions using randomly assigned dates. Second, I randomly assign New Jersey hospitals to participate in the program, holding constant both the number of participating hospitals and the timing of the program. I repeat the main difference-in-difference regressions using randomly assigned participation. The cumulative distribution functions of the coefficients from both simulations (each based on 100 repetitions) are presented in figure A6.

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

B. Effects of Bonuses on Costs and Quantity of Services Provided

Despite the fact that the program was explicitly designed to reduce costs, the bonuses have no effect on costs or resource use, conditional on patient health. Table 7 looks at the effect of the bonuses on measures of procedure use and total treatment costs. All regressions in table 7 control for an array of measures of latent health: the Charlson index, the surgical risk factor index, and the individual chronic conditions included in table A7. Conditional on these latent health measures, the bonuses are not associated with significant decreases in length of stay, the use of diagnostic imaging tests, or costs. Even taking the point estimates at face value, the magnitudes are small, and the signs are not consistently negative. In addition, there is no evidence of substitution between high-tech (MRIs and CT scans) and low-tech (diagnostic ultrasounds) imaging,

TABLE 7
Bonuses Do Not Reduce Costs or Change Procedure Use

	Length of Stay (1)	CT Scan (2)	MRI (3)	Diagnostic Ultrasound (4)	Any Imaging (5)	Total Costs (6)
			A. Medical Patients			
Policy	−.091	−.001	−.003	−.001	−.002	335
	(.111)	(.003)	(.002)	(.004)	(.005)	(380)
Mean dependent variable	5.782	.056	.023	.035	.118	9,360
Clusters	74	74	74	74	74	58
Observations	1,184,413	1,184,413	1,184,413	1,184,413	1,184,413	975,970
			B. Surgical Patients			
Policy	−.053	−.003	−.001	.003	.008	880
	(.156)	(.002)	(.001)	(.006)	(.011)	(596)
Mean dependent variable	6.917	.036	.013	.058	.191	18,213
Clusters	73	73	73	73	73	58
Observations	533,227	533,227	533,227	533,227	533,227	457,606

Note.—Quarter-by-year, doctor, hospital, and diagnosis by severity of illness fixed effects also included, as well as dummies for age categories, sex, race, and variables measuring underlying health: the Charlson index, the surgical risk factor index, indicators for scoring zero on each index, and the individual chronic conditions in table A7. Standard errors clustered at the hospital level. The sample is smaller when looking at costs, as the cost-to-charge ratio is not available for all hospitals.

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).

While regressions reported in table 7 control for a number of latent health measures, it is likely that I cannot completely control for differences in underlying health status. Given the fact that healthier patients were admitted into participating hospitals, these patients may have required fewer resources from the start. Thus, the small decreases reported in some measures in table 7 should be considered an upper bound on the true overall effect.

As the measure of costs based on deflated list charges from the hospital discharge data is quite noisy, my preferred measures of resource use are those reflecting actual procedure use and length of stay. However, the null effect of the policy on costs remains when the top 1% of costs are trimmed (see table A9). Finally, not only was there no effect on the level of costs; there was also no effect of the policy on the standard deviation of costs (see table A10).

Despite not reducing either the level or dispersion of costs overall, one might expect the bonuses to be more successful if doctors are not able to manipulate where patients are treated or whether they were admitted. However, there is no evidence that the bonuses reduced treatment costs for surgical patients, who displayed little evidence of manipulation on either the sorting or admissions margin. Table A8 narrows the sample to single-hospital doctors for whom the sorting margin is eliminated, and again finds no evidence that the bonuses are associated with lower costs. Surgeons and doctors working in a single hospital provide an interesting glimpse into what could be expected if the program were scaled up. Given the null results in both groups it is unlikely that the bonuses would be more successful if all hospital participated, or even if both the sorting and admission margins were eliminated.

Given these null results, how did the initial evaluation conclude that the program succeeded in decreasing costs? In column 1 of table 8, I replicate the initial evaluation of the first wave of the program for medical patients. As in the initial study, I first only include hospitals that eventually take up the initial demonstration, with no controls for latent health or doctor fixed effects. Here, the policy appears to decrease costs, and this decrease is statistically significant. In column 2, however, I show that clustering standard errors at the hospital level already renders the decrease in costs insignificant.[32] In columns 3–5, I add health controls, comparison hospitals, and doctor fixed effects, and show that the sign

[32] In most cases, 11 clusters is thought to be too few for clustered standard errors, which is a worry primarily because having few clusters generally leads to overrejection. However, this overrejection is less of a concern in table 8, as cols. 2 and 3 are only meant to be illustrative of the lack of power of the initial evaluation.

TABLE 8
REPLICATING THE INITIAL EVALUATION

	Total Costs				
	(1)	(2)	(3)	(4)	(5)
Policy	−420***	−420	−578	588	616
	(137)	(790)	(810)	(548)	(533)
Health controls	No	No	Yes	Yes	Yes
Comparison hospitals	No	No	No	Yes	Yes
Doctor fixed effects	No	No	No	No	Yes
Mean dependent variable	9,431	9,431	9,431	9,331	9,331
Clusters		11	11	58	58
Observations	194,349	194,349	194,349	809,982	809,982

NOTE.—Quarter-by-year, hospital, and diagnosis by severity of illness fixed effects included in all regressions. Health controls are the Charlson index, the surgical risk factor index, indicators for scoring zero on each index, indicators for the chronic conditions in table A7, and for age categories, sex, and race. Comparison hospitals are those where the gainsharing policy was not implemented.
*** $p < .01$.

flips from negative to positive. Table 8 also demonstrates the extent to which low-hanging fruit is not being utilized from a data analysis perspective. When evaluating the first wave of the Gainsharing Demonstration, simply including comparison hospitals would have given essentially the correct result.[33] While it is possible to conclude that the bonus program lowered costs, this conclusion does not hold up to a more thorough investigation.

VI. Alternative Strategy: Doctor-Level Program Exposure

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

In order to isolate the effect of the bonuses on total costs and procedure use, I use an alternative identification strategy based on doctor-level ex ante program exposure. Program exposure is zero in the preperiod,

[33] The Bundled Payments for Care Improvement (BPCI) Model 1 program was an expanded version of the New Jersey Gainsharing program (Coates 2014), which officially began in April 2013. Applications to participate in BPCI Model 1, however, were due in late 2011. The first interim evaluation report on the New Jersey Gainsharing Demonstration was written in mid-2012, and the final report (which did use comparison hospitals) was not published until September 2014—at which point the BCPI Model 1 had already been in operation for over a year.

when no doctors are working under the bonus scheme, and then rises to the preprogram fraction of a doctor's caseload treated at participating hospitals. In particular, the exposure variable measures the fraction of a doctor's Medicare patients that would have been affected by the program if the distribution of patients across hospitals had been fixed in the preperiod (2006–2008).

For a doctor who only works in participating hospitals, the exposure variable is zero before the program and 1 when the program goes into effect. For a doctor whose caseload in the preperiod is split evenly between two hospitals, one of which participates, the exposure variable goes from zero to 0.5. This exposure variable captures the fact that some doctors only admit patients to participating hospitals, others are not exposed at all, and many doctors are in between. And by construction, the exposure measure reflects only ex ante exposure, and will not be affected by doctors sorting patients in response to the bonus program.

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

$$\text{outcome}_{dt} = \beta_0 + \beta_1 \text{exposure}_{dt} + \lambda_d + \lambda_t + \epsilon_{dt}, \tag{4}$$

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

There is no evidence of any cost-saving response to the Gainsharing Demonstration as a result of program exposure; if anything, exposure is associated with higher costs. Columns 1 and 2 of table 9 support the (at least short-run) effectiveness of program rules prohibiting increases in the number of admitted patients; doctors with more exposure do not increase overall admissions or admission rates. However, the exposure to the Gainsharing Demonstration is associated with higher average costs (col. 3 shows the effect on the total costs incurred over a quarter, and col. 4 the average costs incurred per patient). In addition, there is no evidence that program exposure decreases the number of imaging tests performed. Instead, exposure is associated with insignificant increases in the number of CT scans and MRIs performed (cols. 5 and 6), as well as in the overall number of diagnostic imaging tests (col. 8). There is also no impact of program exposure on the total number of days spent in the hospital (col. 9). However, physician-level program exposure is again associated with healthier admitted patients (col. 10), with a magnitude very similar to that found in the previous section.

TABLE 9

Simulated Share Treated on Costs and Procedure Use

	Admitted Patients (1)	Admission Rate (2)	Total Costs (3)	Average Costs (4)	CT Scans (5)	MRIs (6)	Diagnostic Ultrasound (7)	Diagnostic Imaging (8)	Hospital Days (9)	CCI of Admitted (10)
Simulated share	−.126	−.005	7,738	268**	.098	.037	−.060	.075	−.694	−.118***
	(1.756)	(.005)	(18,937)	(119)	(.217)	(.090)	(.174)	(.350)	(10.795)	(.017)
Mean dependent variable	47.415	.851	459,896	10,571	3.327	1.207	1.896	7.355	282.031	2.404
Clusters	3,466	3,466	3,322	3,322	3,466	3,466	3,466	3,466	3,466	3,466
Observations	86,985	86,985	79,674	79,674	86,985	86,985	86,985	86,985	86,985	86,985

Note.—Sample includes all Medicare patients seen by doctors in the main analysis. Doctor and quarter-by-year fixed effects included. Total admissions, total costs, CT scans, MRIs, diagnosic ultrasounds, and diagnostic imaging are totals at the doctor-quarter level. Hospital days refers to the total number of days per quarter patients stayed in the hospital. Admission rate and average costs are averaged across patients at the doctor-quarter level. Regressions and means are weighted by the number of admitted Medicare patients.

** $p < .05$.

*** $p < .01$.

The results of the doctor-level exposure analysis rule out systematic changes in practice style in response to the bonuses. One interpretation of table 9 is that some doctors respond to the program by performing additional tests in order to justify admission for patients who otherwise would not have been admitted. While the estimates are too noisy to pinpoint the source of the cost increase, there is no evidence from either identification strategy that the Gainsharing Demonstration resulted in lower costs. There is also no evidence that the program increased the number of admitted patients or admission rates, suggesting that hospitals were successful at limiting overall changes in the volume of admissions. As was predicted by the model, the program resulted in a healthier pool of admitted patients, without a concurrent increase in the overall number admitted—implying that some sicker patients who would have been admitted previously received less intensive care under the program.

VII. Discussion: Implications for Scaling

Policy makers considering interventions often seek to ground their decisions in research, and look to pilot programs for insight on the potential efficacy of a new policy. It is common, however, for treatment effects found in pilot programs to substantially diminish when the intervention is applied at a larger scale, and these predictable changes tend not to be accounted for in cost-benefit analyses (Al-Ubaydli, List, and Suskind 2017). In response, the experimental literature is currently exploring how to ensure that pilot programs are successful when taken to scale, what can go wrong when scaling up rigorously evaluated, internally valid experiments, and ways to mitigate these issues (Al-Ubaydli et al. 2017; Al-Ubaydli, List, and Suskind 2017, 2019; Banerjee et al. 2017; Muralidharan and Niehaus 2017).

In Al-Ubaydli, List, and Suskind (2019), the scaling problem is modeled in a world where the government monitors experiments designed and run by scientists, and must decide which pilot programs to scale up and when the evidence is sufficient to act. My results suggest that this model of policy making, which is a good description of heavily supervised and carefully evaluated randomized controlled trials, is difficult to apply to the New Jersey Gainsharing Demonstration. There are two main concerns. First, replicating the case study will not lead to more precise causal inference because the false-positive outcome was created by gaming, not by chance. Second, the pilot program's implementation created opportunities for gaming in ways in which a fully scaled version of the program would not. In particular, the implementation of the pilot program created the appearance of treatment effects that were in fact a result of incentives to change which patients were admitted and to sort patients between hospitals, rather than a fundamental relationship between the treatment and the desired outcome.

One potential explanation for the existence of these design issues is that the program relied heavily on industry insiders for both development and implementation. These industry stakeholders may be able to provide valuable institutional knowledge, which is particularly helpful when studying complex institutions such as US health care. On the other hand, they may also be incentivized to design programs in ways that benefit members of their organization, rather than to credibly inform policy or gain scientific understanding about how doctors respond to financial incentives.

What lessons about policy experimentation can we learn from this program? First, it is important to more explicitly acknowledge the differing incentives faced by policy makers and industry. Once this step has been taken, a framework for experimentation can be designed to involve outsiders more formally, both initially in the program design and later in the evaluation. The current process of external evaluation is a first step, but the fact that the program was expanded before the formal evaluation was complete suggests there is more work to be done.

Abstracting from issues with the process, how could the program itself have been designed to be more successful? Two natural alternatives to a hospital-based design are market- and doctor-level treatment. Market-level treatment would shut down the sorting channel by encompassing the entire network. The main downside, however, is power; New Jersey only has seven hospital referral regions, which represent regional health care markets for tertiary medical care. Alternatively, applying the treatment at the doctor or practice level would also shut down the sorting channel, and provide for a higher-powered evaluation of the policy. Unless participation was mandatory, however, there would likely be sorting into the program—for example, very motivated doctors, or doctors with the highest bonus potential—which could also limit the usefulness of the demonstration for considering a scaled-up version of the program.[34]

Neither of these alternative treatment definitions, however, would shut off the incentives to manipulate the admission margin. While this channel could be easily eliminated by not requiring bonus-eligible patients to be admitted, it is perhaps telling that this requirement ended up in the demonstration. Broadening the program to include outpatients could make the program less attractive to hospitals, who have more to gain from decreasing costs for admitted patients. And without buy-in from the hospitals, the demonstration never would have existed in the first place.

[34] In nearly all Medicare payment model demonstrations, and all such demonstrations at the practice level, participation has been voluntary (United States Government Accountability Office 2018; Robeznieks 2019).

When the institution under study is very sophisticated, situations can arise in which industry insiders have a better understanding of the institution than experimenters and policy makers. The New Jersey Gainsharing Demonstration can be read as a case study in issues that pilot programs experience in settings such as US health care, where networks are complex and agents are mobile and well informed. As experimental economics increasingly moves from relatively small-scale lab and field experiments into collaborations with governments and working in complex institutional settings, these considerations will likely become increasingly important.

VIII. Conclusion

In this paper, I show that a pilot program that paid doctors bonuses for reducing costs was unsuccessful; doctors changed which patients were admitted and sorted healthier patients into participating hospitals, but did not reduce costs. The results of this program provide two main takeaways for policy. First, doctors are able to identify high- and low-cost patients within narrowly defined bins, and use this information to their advantage. Adjusting payments for patient severity is a ubiquitous feature of health care compensation systems, and policy makers should be wary of doctors manipulating these margins. Second, and more broadly, I provide yet another reason why pilot programs may not be well suited for comparing the effectiveness of different health policy reforms. Not only did the sorting and selection behavior of providers undermine the pilot program; the program was expanded before external evaluations were complete. While external validity is always a concern when deciding whether to expand a pilot program, this paper calls into question the internal validity of pilot program evaluations in US health care.

In early 2018, the Trump administration announced plans for a new, voluntary bundled payment model, "Bundled Payments for Care Improvement Advanced." Under the new model, providers will receive a bonus if the spending for admitted Medicare patients with certain diagnoses is below a target, and will pay a fine if spending is too high. The new demonstration shares two key features with the New Jersey Gainsharing Demonstration: eligibility for the bonus is conditional on admission, and participation is voluntary. The addition of penalties for doctors may further complicate the program's evaluation, as doctors can forecast their spending, and those who predict receiving a penalty may choose not to participate. My analysis of the New Jersey Gainsharing Demonstration suggests that while this type of incentive program is unlikely to reduce costs, it may appear to be effective in simple evaluations due the gaming behavior of physicians.

References

Abaluck, Jason, Leila Agha, Christopher Kabrhel, Ali Raja, and Arjun Venkatesh. 2015. "Negative Tests and the Efficiency of Medical Care: What Determines Heterogeneity in Imaging Behavior?" Working Paper no. 19956 (August), NBER, Cambridge, MA.

Agency for Healthcare Research and Quality. 2014. "Hospital Gain-Sharing Program Offers Incentives to Physicians Based on Their Efficiency, Producing Significant Cost Savings without Decline in Quality." Technical report, AHRQ Innovations Exchange, Rockville, MD.

Alexander, Diane. 2015. "Does Physician Pay Affect Procedure Choice and Patient Health? Evidence from Medicaid C-Section Use." Working paper, Dept. Econ., Princeton Univ.

Allcott, Hunt. "Site Selection Bias in Program Evaluation." 2015. *Q.J.E.* 130 (3): 1117–65.

Al-Ubaydli, Omar, John A. List, Danielle LoRe, and Dana Suskind. 2017. "Scaling for Economists: Lessons from the Non-Adherence Problem in the Medical Literature." *J. Econ. Perspectives* 31 (4): 125–43.

Al-Ubaydli, Omar, John A. List, and Dana L. Suskind. 2017. "What Can We Learn from Experiments? Understanding the Threats to the Scalability of Experimental Results." *A.E.R.* 107 (5): 282–86.

———. 2019. "The Science of Using Science: Towards an Understanding of the Threats to Scaling Experiments." Working Paper no. 25848 (May), NBER, Cambridge, MA.

Banerjee, Abhijit, Rukmini Banerji, James Berry, et al. 2017. "From Proof of Concept to Scalable Policies: Challenges and Solutions, with an Application." *J. Econ. Perspectives* 31 (4): 73–102.

Best, William R., Shukri F. Khuri, Maureen Phelan, et al. 2002. "Identifying Patient Preoperative Risk Factors and Postoperative Adverse Events in Administrative Databases: Results from the Department of Veterans Affairs National Surgical Quality Improvement Program." *J. American Coll. Surgeons* 194 (3): 257–66.

Bilimoria, Karl Y., Yaoming Liu, Jennifer L. Paruch, et al. 2013. "Development and Evaluation of the Universal ACS NSQIP Surgical Risk Calculator: A Decision Aid and Informed Consent Tool for Patients and Surgeons." *J. American Coll. Surgeons* 217 (5): 833–42.

Brot-Goldberg, Zarek, Amitabh Chandra, Benjamin Handel, and Jonathan Kolstad. 2017. "What Does a Deductible Do? The Impact of Cost-Sharing on Health Care Prices, Quantities, and Spending Dynamics." *Q.J.E.* 132:1261–1318.

Brown, Jason, Mark Duggan, Ilyana Kuziemko, and William Woolston. 2011. "Will Risk-Adjustment Decrease Health Care Costs? New Evidence from the Medicare Advantage Programs." Working paper, Dept. Econ., Princeton Univ.

Chan, David. 2015. "The Efficiency of Slacking Off: Evidence from the Emergency Room." Working Paper no. 21002 (June), NBER, Cambridge, MA.

Charlson, Mary, Robert Charlson, Janey Peterson, Spyridon Marinopoulos, William Briggs, and James Hollenberg. 2008. "The Charlson Comorbidity Index Is Adapted to Predict Costs of Chronic Disease in Primary Care Patients." *J. Clinical Epidemiology* 61 (12): 1234–40.

Clemens, Jeffrey, and Joshua Gottlieb. 2014. "Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?" *A.E.R.* 104 (4): 1320–49.

Coates, Robert. 2014. "The New Jersey Gainsharing Experience." *Physician Executive* 40 (1): 46–51.

Coey, Dominic. 2013. "Physician Incentives and Treatment Choices in Heart Attack Management." Discussion Paper no. 12-027, Stanford Inst. Econ. Policy Res.

Cohen, Robin A., Michael E. Martinez, and Emily P. Zammitti. 2016. "Health Insurance Coverage: Early Release of Estimates from the National Health Interview Survey, January–March 2016." Technical report, Nat. Center Health Statis., Atlanta, GA.

Cullen, Julie Berry, Steven Levitt, Erin Robertson, and Sally Sadoff. 2013. "What Can Be Done to Improve Struggling High Schools?" *J. Econ. Perspectives* 27 (2): 133–52.

Cutler, David. 1990. "Empirical Evidence on Hospital Delivery under Prospective Payments." Working paper, Dept. Econ., Massachusetts Inst. Tech.

———. 1995. "The Incidence of Adverse Medical Outcomes under Prospective Payment." *Econometrica* 63 (1): 29–50.

Dafny, Leemore. 2005. "How Do Hospitals Respond to Price Changes?" *A.E.R.* 95 (5): 1525–47.

Dafny, Leemore, and David Dranove. 2009. "Regulatory Exploitation and Management Changes: Upcoding in the Hospital Industry." *J. Law and Econ.* 52 (2): 223–50.

Dickstein, Michael. 2014. "Physician vs. Patient Incentives in Prescription Drug Choice." Working paper, Stern School Bus., New York Univ.

Dranove, David, Daniel Kessler, Mark McClellan, and Mark Satterthwaite. 2003. "Is More Information Better? The Effects of 'Report Cards' on Health Care Providers." *J.P.E.* 111 (3): 555–88.

Dranove, David, and Paul Wehner. 1994. "Physician-Induced Demand for Childbirths." *J. Health Econ.* 13 (1): 61–73.

Duflo, Esther. 2004. "Scaling Up and Evaluation." Technical report, Ann. World Bank Conference Development Econ., World Bank, Washington, DC.

Duggan, Mark. 2000. "Hospital Ownership and Public Medical Spending." *Q.J.E.* 115 (4): 1343–73.

———. 2002. "Hospital Market Structure and the Behavior of Not-for-Profit Hospitals." *RAND J. Econ.* 33 (3): 433–46.

———. 2004. "Does Contracting Out Increase the Efficiency of Government Programs? Evidence from Medicaid HMOs." *J. Public Econ.* 88 (12): 2549–72.

Duggan, Mark, and Tamara Hayford. 2013. "Has the Shift to Managed Care Reduced Medicaid Expenditures? Evidence from State and Local-Level Mandates." *J. Policy Analysis and Management* 32 (3): 505–35.

Ellis, Randall, and Thomas McGuire. 1996. "Hospital Response to Prospective Payment: Moral Hazard, Selection, and Practice-Style Effects." *J. Health Econ.* 15:257–77.

Fendrick, Mark, and Michael Chernew. 2017. "Precision Benefit Design—Using 'Smarter' Deductibles to Better Engage Consumers and Mitigate Cost-Related Nonadherence." *JAMA Internal Medicine* 177 (3): 368–70.

Grant, Darren. 2009. "Physician Financial Incentives and Cesarean Delivery: New Conclusions from the Healthcare Cost and Utilization Project." *J. Health Econ.* 28 (1): 244–50.

Greenwald, Leslie, Walter Adamache, Maggie Cole-Beebe, Peter Amico, Elizabeth Hunter, and Bob Baker. 2014. "Evaluation of the Medicare Physician Hospital Collaboration Demonstration: Final Report." Technical report, RTI Internat., Raleigh, NC.

Gruber, Jonathan. 2008. "Covering the Uninsured in the United States." *J. Econ. Literature* 46 (3): 51–68.

———. 2010. “The Cost Implications of Health Care Reform.” *New England J. Medicine* 362:2050–51.
Gruber, Jonathan, John Kim, and Dina Mayzlin. 1999. “Physician Fees and Procedure Intensity: The Case of Cesarean Delivery.” *J. Health Econ.* 18 (4): 473–90.
Gruber, Jonathan, and Maria Owings. 1996. “Physician Financial Incentives and Cesarean Section Delivery.” *RAND J. Econ.* 27 (1): 99–123.
Gupta, Atul. 2016. “Impact of Performance Pay for Hospitals: The Readmissions Reduction Program.” Working paper, Dept. Econ., Stanford Univ.
Hadley, Jack, Jean Mitchell, and Jeanne Mandelblatt. 2001. “Medicare Fees and Small Area Variations in Breast-Conserving Surgery among Elderly Women.” *Medical Care Res. and Rev.* 58 (3): 334–60.
Hadley, Jack, James Reschovsky, Catherine Corey, and Stephen Zuckerman. 2009. “Medicare Fees and the Volume of Physicians’ Services.” *Inquiry* 46 (4): 372–90.
Haviland, Amelia, Neeraj Sood, Roland McDevitt, and M. Susan Marquis. 2011. “How Do Consumer-Directed Health Plans Affect Vulnerable Populations?” *Forum Health Econ. and Policy* 14 (2): 1–12.
Hillman, Bruce, and Jeff Goldsmith. 2010. “The Uncritical Use of High-Tech Medical Imaging.” *New England J. Medicine* 363:4–6.
Ho, Kate, and Ariel Pakes. 2014. “Hospital Choices, Hospital Prices, and Financial Incentives to Physicians.” Working Paper no. 19333 (March), NBER, Cambridge, MA.
Keeler, Emmett, and Thomas Fok. 1996. “Equalizing Physician Fees Had Little Effect on Cesarean Rates.” *Medical Care Res. and Rev.* 53:465–71.
Kolstad, Jonathan. 2013. “Information and Quality When Motivation Is Intrinsic: Evidence from Surgeon Report Cards.” *A.E.R.* 103 (7): 2875–2910.
Leibowitz, Arleen, Joan Buchanan, and Joyce Mann. 1992. “A Randomized Trial to Evaluate the Effectiveness of a Medicaid HMO.” *J. Health Econ.* 11 (3): 235–57.
Lessler, Adam, Joshua Isserman, Rajan Agarwal, Harold Palevsky, and Jesse Pines. 2010. “Testing Low-Risk Patients for Suspected Pulmonary Embolism: A Decision Analysis.” *Ann. Emergency Medicine* 55 (4): 316–26.
List, John, Azeem Shaikh, and Yang Xu. 2016. “Multiple Hypothesis Testing in Experimental Economics.” Working Paper no. 21875 (January), NBER, Cambridge, MA.
Manning, Willard, Joseph Newhouse, Naihua Duan, Emmett Keeler, and Arleen Leibowitz. 1987. “Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment.” *A.E.R.* 77 (3): 251–77.
Muralidharan, Karthik, and Paul Niehaus. 2017. “Experimentation at Scale.” *J. Econ. Perspectives* 31 (4): 103–24.
Office of Inspector General. 1999. “Gainsharing Arrangements and CMPs for Hospital Payments to Physicians to Reduce or Limit Services to Beneficiaries.” Special Advisory Bull., Office Inspector General, Washington, DC.
Orszag, Peter, and Ezekiel Emanuel. 2010. “Health Care Reform and Cost Control.” *New England J. Medicine* 363 (7): 601–3.
Quan, Hude, Bing Li, and Chantal Couris. 2011. “Updating and Validating the Charlson Comorbidity Index and Score for Risk Adjustment in Hospital Discharge Abstracts Using Data from 6 Countries.” *American J. Epidemiology* 173 (6): 676–82.
Rao, Vijay, and David Levin. 2012. “The Overuse of Diagnostic Imaging and the Choosing Wisely Initiative.” *Ann. Internal Medicine* 157 (8): 574–76.
Reinhardt, Uwe. 2006. “The Pricing of U.S. Hospital Services: Chaos Behind a Veil of Secrecy.” *Health Affairs* 25:57–69.

Robeznieks, Andis. 2019. “Physician Proposals Inspire New HHS Pay Models for Primary Care.” https://www.ama-assn.org/practice-management/payment-delivery-models/physician-proposals-inspire-new-hhs-pay-models-primary.

Silverman, Elaine, and Jonathan Skinner. 2004. “Medicare Upcoding and Hospital Ownership.” *J. Health Econ.* 23:369–89.

Smith, Sherry, ed. 2012. *Medicare RBRVS 2012: The Physician Guide*. Chicago: American Medical Assoc.

Smith-Bindman, Rebecca. 2010. “Is Computed Tomography Safe?” *New England J. Medicine* 363 (1): 1–4.

Timbie, Justin, Peter Hussey, Lane Burgette, et al. 2014. “Medicare Imaging Demonstration Final Evaluation: Report to Congress.” Technical report, RAND, Santa Monica, CA.

United States Government Accountability Office. 2018. “Medicare: Voluntary and Mandatory Episode Based Payment Models and Their Participants.” Technical report, US Government Accountability Office, Washington, DC.

Verma, Seema. 2017. “Medicare and Medicaid Need Innovation.” *Wall Street Journal*, September 19.

Wharam, J. Frank, Fang Zhang, Emma M. Eggleston, Christine Y. Lu, Stephen Soumerai, and Dennis Ross-Degnan. 2017. “Outpatient Care and Acute Complications Before and After High-Deductible Insurance Enrollment: A Natural Experiment for Translation in Diabetes (NEXT-D) Study.” *JAMA Internal Medicine* 177 (3): 358–68.

Yip, Winnie. 1998. “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.” *J. Health Econ.* 17:675–99.