Adoption and Learning Across Hospitals: The Case of a Revenue-Generating Practice

Adam Sacarny[*]

April 1, 2018

Abstract

Performance-raising practices tend to diffuse slowly in the health care sector. To understand how incentives drive adoption, I study a practice that generates revenue for hospitals: submitting detailed documentation about patients. After a 2008 reform, hospitals could raise their Medicare revenue over 2% by always specifying a patient's type of heart failure. Hospitals only captured around half of this revenue, indicating that large frictions impeded takeup. Exploiting the fact that many doctors practice at multiple hospitals, I find that four-fifths of the dispersion in adoption reflects differences in the ability of hospitals to extract documentation from physicians. Hospital adoption is robustly correlated with generating survival for heart attack patients and using inexpensive survival-raising standards of care. Hospital-physician integration and electronic medical records also influence adoption. These findings highlight the potential for institution-level frictions, including agency conflicts, to explain variations in health care performance across providers.

JEL Classification: D22; I1; O31; O33; L2

Keywords: Hospitals; Healthcare; Technology adoption; Firm Performance; Upcoding

1 Introduction

A classic finding of studies of technology is that new, performance-raising forms of production are adopted slowly and incompletely. For example, Griliches (1957) observed this pattern in the takeup of hybrid corn across states; more recent research has studied adoption patterns in agriculture in the developing world, manufacturing in advanced economies, management practices internationally, and a host of other examples (Conley and Udry, 2010; Foster and Rosenzweig, 1995; Collard-Wexler and De Loecker, 2015; Bloom et al., 2012). In the health-care sector, clinical quality-improving practices, including checklists, hand-washing, and drugs like β-blockers provide analogous examples of slow adoption. Disparities in the use of these practices are a leading explanation for health care productivity variations across providers and regions (Skinner and Staiger, 2015; Baicker and Chandra, 2004; Chandra et al., 2013). Given the enormous potential for new forms of production to improve patient outcomes in the health care sector and to raise output in the economy more generally, the nearly ubiquitous finding of delayed takeup is particularly vexing.

In this paper, I study a health care practice that raises revenue for the hospital: the detailed reporting of heart failure patients to Medicare. A 2008 Medicare policy change created a financial incentive for hospitals to provide more detail about their patients in insurance reimbursement claims.[1] Yet hospitals could only provide these details if they were documented by physicians. The incentive for hospitals to report the information was large: this policy put over 2% of hospital Medicare incomes on the line in 2009 – about \$2 billion – though it did not directly affect the pay of physicians. By tracking the spread of the reporting practice across hospitals, this study examines the role of financial incentives and agency conflicts in the adoption of new practices. While improved heart failure billing is a revenue-raising but not survival-raising practice, and is thus less influenced by physicians' intrinsic motivations for clinical quality, it is a test case of how financial incentives drive takeup in the presence of firm-level barriers to adoption.

Figure 1 shows that the change in incentives triggered a rapid but incomplete response by hospitals: in just weeks following the reform, hospitals started capturing 30% of the revenue made available; by the end of 2010 they were capturing about 52%. This finding is consistent with existing work showing that hospitals respond to incentives by changing how they code their patients

---

[1] All years are federal fiscal years unless otherwise noted. A federal fiscal year begins on October 1 of the previous calendar year, i.e. three months prior to the calendar year.

(Dafny, 2005; Silverman and Skinner, 2004). Yet presented inversely, in spite of the reform being announced earlier that year, 70% of the extra heart failure revenue was not captured shortly after implementation and nearly half was still not being realized after several years.

I show that substantial hospital-level heterogeneity underlies the national takeup of detailed heart failure codes. Mirroring the literature that has demonstrated large differences in productivity across seemingly similar firms (Fox and Smeets, 2011; Syverson, 2011; Bartelsman et al., 2013), I find dispersion in the takeup of detailed billing codes across hospitals. This dispersion exists even after accounting for disparities in the types of patients that different hospitals treat. For example, 55% of heart failure patients received a detailed code at the average hospital in 2010, and with the full set of patient controls the standard deviation of that share was 15 percentage points. A hospital two standard deviations below the mean provided detailed heart failure codes for 24% of its heart failure patients, while a hospital two standard deviations above the mean did so for 85% of its patients. While Song et al. (2010) and Finkelstein et al. (2017) find evidence of disparities in regional coding styles, this study is the first to isolate the hospital-specific component of coding adoption and study its distribution (I also find disparities in coding across regions, but regions leave unexplained at least three-quarters of the variation in hospital coding styles).

My findings suggest that hospitals were aware of the financial incentive to use the detailed codes, but that this awareness was tempered by significant frictions. I note two key potential drivers of incomplete and varied adoption of the codes across hospitals. First, an agency problem arises because physicians supply the extra information about the heart failure, but Medicare does not pay them for the detailed codes. Second, hospitals' health information management staff and systems may have been differentially effective at translating the information that physicians provided into the high-value codes.

To study the role of these frictions, I consider adoption rates that isolate the role of hospitals above and beyond their patients and physicians. Because doctors practice at multiple hospitals, it is possible to decompose the practice of detailed documentation into hospital- and physician-specific components. This decomposition is an application of a labor economics technique that has been frequently used in the context of workers and firms (Abowd et al., 1999; Card et al., 2013); to the author's knowledge this study is among the first, alongside Finkelstein et al. (2016)'s decomposition of health spending across regions, to apply this approach in health care.

Isolating the hospital contribution addresses the concern that some hospitals might work with physicians who would be more willing to supply the documentation wherever they practice. Yet dispersion is, if anything, slightly increased when the hospital component is isolated: the standard deviation of the detailed documentation rate across hospitals rises from 15 percentage points with rich patient controls to 16 percentage points with patient and physician controls. The residual variation means that even if facilities had the same doctors, some would be more capable of extracting specific documentation from their physicians than others (I also study the physician contribution to adoption, where dispersion is of a similar magnitude). These results are consistent with firm-level disparities in resolving frictions.

I next consider the correlation between hospital adoption and hospital characteristics. The most powerful predictors of hospital adoption are the measures of clinical quality: heart attack survival and use of survival-raising processes of care. High clinical quality facilities are also more likely to be early adopters. Under the view that extracting the revenue-generating codes from physicians makes a hospital revenue-productive, these results show that treatment and revenue productivity are positively correlated. This result also touches on a key policy implication of this study: that financial incentives that push providers to raise treatment quality may be relatively ineffective on the low quality facilities most in need of improvement. Adoption is correlated with hospital-physician integration, suggesting that a key tool for hospitals to resolve takeup frictions is contractual arrangements that align the two parties. Electronic medical records also influence adoption, suggesting that health information systems can help to resolve the frictions – though this finding is estimated imprecisely in my preferred specification.

I contribute to the literature on health care provider performance variations in several ways. First, by focusing on whether hospitals are able to modify their billing techniques to extract revenue, I isolate disparities in a context where it is plausible they might be small or nonexistent. These disparities reflect differences in hospitals' basic ability to respond to incentives. Second, using decomposition techniques adapted from studies of labor markets, I show that four-fifths of the variation in adoption is driven by some hospitals being able to extract more high-revenue codes from their patients and physicians than others. Third, I correlate the adoption of revenue-generating codes with the use of high quality standards of care in treatment to find that a common factor may drive both outcomes. Fourth, I show that facilities that more closely integrate with their physicians

are also more likely to adopt, hinting that principal-agent problems may play a role in productivity dispersion more generally – inside and outside the health care sector.

A key caveat of these analyses is that they are descriptive, and thus only suggestive of causal relationships. For example, this study shows that clinical performance and coding are correlated; this relationship could be driven by unobserved institution-level factors like the quality of hospital staff (though not physicians, which I control for). Likewise, while hospitals with better coding are more likely to be integrated with physicians, this integration could be the result of other factors, like management practices, that exert their own influences on coding.

The paper proceeds as follows. Section 2 discusses the heart failure billing reform, the data I use to study it, and provides a simple analytical framework. Section 3 describes the econometric strategy and identification. Section 4 presents results on dispersion in takeup, then shows how takeup relates to hospital and physician characteristics. Section 5 provides a discussion of the results. Section 6 concludes.

2 Setting and Data

Heart failure (HF) is a syndrome defined as the inability of the heart's pumping action to meet the body's metabolic needs. It is uniquely prevalent and expensive among medical conditions. There are about 5 million active cases in the United States; about 500,000 cases are newly diagnosed each year. Medicare, the health insurance program that covers nearly all Americans age 65 and over, spends approximately 43% of its hospital and supplementary insurance dollars treating patients with HF (Linden and Adler-Milstein, 2008).

The classic economic literature on health care eschews studying HF in favor of less common conditions like acute myocardial infarctions (AMIs), or heart attacks (see e.g. McClellan et al., 1994; Cutler et al., 1998, 2000; Skinner et al., 2006 and Chandra and Staiger, 2007). The literature has focused on these conditions because they are thought to be sensitive to treatment quality, are well observed in most administrative data, and almost always result in a hospitalization, removing the issue of selection into treatment. Since this paper concerns how hospitals learn to improve their billing practices, not the effect of treatment on health, the endogenous selection of patients into the inpatient setting is not a central econometric barrier. Rather, the great deal of revenue at stake for

the reimbursement of heart failure patients makes it a condition that is well suited for this study's aim of understanding how hospitals respond to documentation and coding incentives.

The hospitals I study are paid through Medicare's Acute Inpatient Prospective Payment System (IPPS), the $112 billion program that pays for most Medicare beneficiaries who are admitted as inpatients to most hospitals in the United States MEDPAC (2015). As part of a 2008 overhaul of the IPPS – the most significant change to the program since its inception – the relative payment for unspecified type (vaguely documented) and specified type (specifically documented) HF changed. This element of the reform made the documentation valuable and provided the financial incentive for the spread of the practice.

2.1 Payment Reform and Patient Documentation

The 2008 overhaul was a redesign of the IPPS risk-adjustment system, the process that adjusts payments to hospitals depending on the severity, or level of illness, of a patient. Medicare assigns a severity level to every potential condition a patient might have. A patient's severity is the highest-severity condition listed on his hospital's reimbursement claim. The reform created three levels of severity (low, medium, or high) where there had been two (low or high), shuffling the severity level of the many heart failure codes in the process.[2]

By the eve of the reform, Medicare policymakers had come to believe that the risk-adjustment system had broken down, with nearly 80% of inpatients crowded into the high-severity category (GPO, 2007). The reporting of HF had been a primary cause of the breakdown: there were many codes describing different types of HF, and all of them had been considered high-severity. Patients with HF accounted for about one-fourth of high-severity patients (or one-fifth of patients overall) in the final year before the reform.

Risk adjustment relies on detailed reporting of patients by providers, but according to the Centers for Medicare & Medicaid Services (CMS), which administers Medicare, the overwhelmingly most common of the HF codes – 428.0, "congestive heart failure, unspecified" – was vague. Patients with this code did not have greater treatment costs than average (GPO, 2007). A set of heart failure codes that gave more information about the nature of the condition was found to predict

---

[2]The new severity system's levels in order from low to medium to high were called Non-CC (no complication or comorbidity), CC (complication or comorbidity), and MCC (major complication or comorbidity). The old system's levels were only Non-CC and CC.

treatment cost and, representing specifically identified illnesses, was medically consistent with the agency's definitions of medium and high severity. These codes were in the block 428.xx, with two digits after the decimal point to provide the extra information. The vague code was moved to the low-severity list, but each of the detailed codes was put on either the medium- or the high-severity list (Table 1).

The detailed codes were exhaustive over the types of heart failure, so with the right documentation, a hospital could continue to raise its HF patients to at least a medium level of severity following the reform. The specific HF codes indicate whether the systolic and/or diastolic part of the cardiac cycle is affected and, optionally, whether the condition is acute and/or chronic. Submitting them is a practice that requires effort from both physicians and hospital staff and coordination between the two. In this way it is similar to technologies that have been the focus of researchers and policy-makers, including the use of β-blockers (an inexpensive class of drugs that have been shown to raise survival following AMI; see e.g. Skinner and Staiger, 2015) in health care and the implementation of best managerial practices in firms (e.g. Bloom et al., 2012; McConnell et al., 2013; Bloom et al., 2016).

2.2 Analytical Approach

The framework for analyzing adoption views the decision to use a specific HF code $code \in \{0, 1\}$ as a function of the propensity of the hospital and the doctor to favor putting down the code or documentation thereof. I let hospitals be indexed by h, doctors by d, and patients by p. Under additive separability, hospitals can be represented by a hospital type α_h and doctors by a doctor type α_d. Patient observables are X_p and the remaining heterogeneity, which accounts for unobserved determinants of coding behavior, is ϵ_{ph}:

$$code_{ph} = \alpha_h + \alpha_d + X_p\beta + \epsilon_{ph} \tag{1}$$

The hospital's type can be thought of as its underlying propensity to identify and extract the codes independently of the types of physicians who practice at the hospital. The doctor type reflects that some physicians are more or less prone to document the kind of HF that their patients have due to their own practice styles and the incentives of the physician payment system. In this framework,

doctors carry their types across hospitals. Finally, the patient component accounts for observed differences that, in a way that is common across facilities, affect the cost or benefit of providing a specific code.

The dispersion of the hospital types is the first focus of the empirical analysis. A hospital's type can be thought of as its revenue productivity – its residual ability to extract revenue from Medicare after accounting for the observable inputs to the coding production process, like patient and doctor types. A wide literature has documented persistent productivity differentials in the manufacturing sector (see Syverson, 2011 for a review), and work is ongoing to develop documentation of similar facts in the service and health care sectors (Fox and Smeets, 2011; Chandra et al., 2016a,b). Dispersion in hospital types is therefore a form of productivity dispersion. In Section 2.5 I discuss potential drivers of this dispersion and in Section 4.3 I estimate it.

The second element of the empirical analysis focuses on describing the kinds of hospitals that are most effective at responding to the incentives for detailed coding. These analyses look at the relationships between hospital types and characteristics of the hospital. The first set of characteristics, called C_h, comprises the hospital's size, ownership, location, teaching status, and *ex-ante* per-patient revenue put at stake by the reform. The second set, called I_h, contains factors related to potential facility-level frictions that might improve revenue extraction, like EMRs and hospital-physician integration. The final set Z_h includes measures of the hospital's clinical performance – defined here as its ability to use evidence-based medical inputs and to generate survival.

In the key hospital-level analysis, I regress the hospital type on these three sets of characteristics:

$$\alpha_h = \gamma + C_h\rho + I_h\gamma + Z_h\theta + \eta_h \tag{2}$$

The signs of ρ, γ, and θ are not obvious, both because the causal relationships between hospital characteristics and the takeup of revenue-generating technology are not well known and because other, unobserved factors may be correlated with C_h, I_h, and Z_h and drive takeup. I discuss these potential relationships and estimate this equation in Section 4.4.

The final component of the empirical analysis applies these methods to the physician types, analyzing their dispersion and correlates. From the perspective of revenue generation, physician types are a form of productivity; in practice, they embody the physician's willingness to supply

8

the detailed documentation about their patients. A physician type thus may reflect her alignment with hospitals' aim to generate revenue or her desire to supply information in medical records. Since supplying the documentation may have clinical payoffs as well, types may reflect differences in clinical practice patterns. Section 4.6 studies the dispersion and correlates of these types.

2.3 Data

My data is primarily drawn from the MEDPAR and Inpatient Research Information Files (RIFs), 100% samples of all inpatient stays by Medicare beneficiaries with hospital care coverage through the government-run Original Medicare program. Each row in this file is a reimbursement claim that a hospital sent Medicare. I use data on heart failure hospital stays from the calendar year 2006-2010 files, yielding fiscal year 2007-2010 data (in some secondary analyses I use files back to 2002). These stays are identified as those with a principal or secondary ICD-9 diagnosis code of 428.x, 398.91, 402.x1, 404.x1, or 404.x3.[3] I source additional information about patients from the enrollment and chronic conditions files.

I eliminate those who lacked full Medicare coverage at any point during their hospital stay, were covered by a private plan, were under age 65, or had an exceptionally long hospital stay (longer than 180 days). To focus on hospitals that were subject to the reform, I include only inpatient acute care facilities that are paid according to the IPPS. As a result, I drop stays that occur at critical access hospitals (these hospitals number about 1,300 but are very small and have opted to be paid on a different basis) and Maryland hospitals (which are exempt from the IPPS). The result is a grand sample of all 7.9 million HF claims for 2007 through 2010, 7.3 million of which (93%) also have information about the chronic conditions of the patients.

2.4 Revenue at Stake from Reform

Since HF was so common and the payment for medium- or high-severity patients was so much greater than for low-severity patients, hospitals had an incentive to use detailed codes when possible. Before the reform, the gain from these detailed codes relative to the vague code was zero because they were

---

[3]The codes outside the 428.x block indicate HF combined with or due to other conditions. Patients with these codes can also receive 428.x codes to make the claim more specific about the HF acuity and the part of the cardiac cycle affected – and to raise the hospital's payment. See Table 1 and Appendix Table A1.

effectively identical in the Medicare payment calculation. Consistent with these incentives, fewer than 15% of HF patients received a detailed code in the year before the reform.

Following the reform, the gain was always weakly positive and could be as high as tens of thousands of dollars; the exact amount depended on the patient's main diagnosis and whether the patient had other medium- or high-severity conditions. For patients with other medium-severity conditions, hospitals could gain revenue if they could find documentation of a high-severity form of HF. For patients with other high-severity conditions, finding evidence of high-severity HF would not change Medicare payments, but using the detailed codes was still beneficial to the hospital because it would help to keep payments from being reduced if the claim were audited and the other high-severity conditions were found to be poorly supported.

The reform was phased in over two years and incentives reached full strength in 2009. By then, the average gain per HF patient from using a detailed HF code instead of a vague one was $227 if the code indicated chronic HF (a medium-severity condition) and $2,143 if it indicated acute HF (a high-severity condition).[4] As a point of comparison, Medicare paid hospitals about $9,700 for the average patient and $10,400 for the average HF patient in 2009.[5] Looking at the grand sample of all HF patients from 2007 through 2010, the evolution of the gain to specific coding is shown in Figure 2 and the corresponding takeup in revenue is shown in Figure 1 (Appendix Figure A1 plots the raw takeup of the detailed codes).

For each hospital, the gain to taking up the revenue-raising practice – the revenue at stake from the reform – depended on its patient mix. Hospitals with more HF patients, and more acute (high-severity) HF patients, had more to gain from adopting specific HF coding. To get a sense of how this gain varied across hospitals, I predict each hospital's *ex ante* revenue put at stake by the reform. This prediction takes the hospital's 2007 HF patients and probabilistically fills in the detailed HF codes the patients would have received under full adoption. It then processes the patient under the new payment rules to calculate the expected gain in payment from these codes. Heart failure codes are predicted using the relationship between coding and patient characteristics in hospitals that were relatively specific coders in 2010 (see Appendix Section A.1.2).

---

[4] These averages are calculated on the grand sample of HF patients in 2009. They include the patients for whom the detailed codes do not raise payments because, for example, they already had another medium- or high-severity condition. This calculation is described in greater detail in Appendix Section A.1.1.

[5] All hospital payment calculations in this section refer to DRG prices, the base unit of payment for hospitals in the IPPS system, and exclude other special payments like outlier payments. They are given in constant 2009 dollars.

Figure 3 shows the distribution of *ex ante* revenue put at stake by the reform across hospitals; the average hospital would have expected to gain $1,007 per HF patient in 2009 by giving all of its HF patients specific HF codes rather than vague ones. The standard deviation of the revenue at stake per HF patient was $230. Appendix Figure A3 shows the distribution of the gain when it is spread across all Medicare admissions, which follows a similar but attenuated (as expected) pattern.

To provide a sense of scale, one can consider these amounts relative to hospital operating margins. The 2010 Medicare inpatient margin, which equals hospitals' aggregate inpatient Medicare revenues less costs, divided by revenues, was -1.7% (MEDPAC, 2015). This negative operating margin has been cited by the American Hospital Association as evidence that Medicare does not pay hospitals adequately (American Hospital Association, 2005). The gains from detailed coding for HF were even larger than this margin: pricing the pre-reform patients under the 2009 rules shows that hospitals could have expected to raise their Medicare revenues by 2.9% by giving all of their HF patients specific HF codes.

2.5 Organizational Processes and Takeup Frictions

Figure 1 shows that the reform induced an almost instantaneous partial adoption of the detailed coding practice. Over the following years the takeup continued, though it remained far from 100% even by the end of 2010. The finding of incomplete takeup raises the question of what costs must be incurred by the hospital to adopt.

For a hospital to legally submit a detailed code, a doctor must state the details about the HF in the patient's medical chart.[6] As the physician treats a patient, she inputs information about diagnoses, tests, and treatments in the patient's medical chart. When the patient is discharged, the physician summarizes the patient's encounter, including the key medical diagnoses that were confirmed or ruled out during the stay. This discharge summary provides the primary evidence that the hospital's health information staff (often called coders) and computer systems use when processing the chart (Youngstrom, 2013). The staff can review the chart and send it back to

---

[6]The chart is a file, physical or electronic, containing the patient's test results, comments by providers of treatment, and ultimately a set of primary and secondary diagnoses. Its role is to provide a record of the patient's stay for the purposes of treatment continuity and coordination, but the chart also serves as documentation supporting the hospital's claims from payers like Medicare (Kuhn et al., 2015). CMS and its contractors frequently review charts to ensure that providers are not "upcoding", or submitting high-paying codes that are not indicated by the documentation.

the doctor with a request for more information – this process is called querying. Then, the staff must work with coding software to convert the descriptions of diagnoses into the proper numeric diagnosis codes, which become a part of the inpatient reimbursement claim. A concise description of the coding process can be found in O'Malley et al. (2005).

Both physicians and staff needed to revise old habits and learn new definitions; they also needed to work together to clarify ambiguous documentation. Coding staff might query a physician to specify which part of the cardiac cycle was affected by the HF, and other staff might review patient charts and instruct physicians on how to provide more detailed descriptions (Rosenbaum et al., 2014). Hospitals could also provide clinicians with scorecards on whether their documentation translated into high-value codes, or update their medical record forms and software to make it quicker to document high-value conditions (Richter et al., 2007; Payne, 2010).

A potential friction comes from a principal-agent problem that pitted a hospital interest in detailed documentation against physicians who had little to gain financially from providing the information. Although this documentation may seem nearly costless to produce, physicians face competing demands on their time when they edit medical charts. HF is often just one condition among many that are relevant to the patient's treatment. A doctor's first-order concern may be documenting aspects of the patient that are crucial for clinical care, making documentation that matters solely for the hospital's billing a secondary issue, a view expressed, for example, by the American College of Physicians (Kuhn et al., 2015).

Hospitals also face significant constraints on using incentive pay to resolve the potential conflicts in their aims and those of physicians. The Stark Law and the Anti-Kickback Statute (AKS) both make it illegal for hospitals to incentivize their physicians to refer patients to the facility (regulations are less strict for physicians employed by the hospital, who are broadly exempted from the AKS). Both laws implicate hospital payments to physicians that reward documentation because these payments would incentivize physicians to refer certain groups of patients to the hospital. Such arrangements would pay physicians depending on the "volume or value" of referrals, violating exemptions and safe harbor provisions of both laws (BNA, 2017).

Verifying that hospitals follow these rules in practice is difficult due to the confidential nature of hospital-physician contracts. One approach to reach into the "black box" of hospital practices is to survey hospital managers directly, as in Bloom et al. (2012) and McConnell et al. (2013). In

preliminary work comprising 18 interviews on documentation and coding practices with hospital chief financial officers (CFOs) in a large for-profit hospital chain, all stated that they did not use financial incentives to encourage coding. Generating systematic evidence on the managerial practices underlying coding intensity will be an important avenue for future research.

Taking up the revenue-generating practice required hospitals to pay a variety of fixed and variable costs that could encourage better physician documentation as well as their ability to translate documentation into high-value codes. Examples of these costs include training hospital staff to prompt doctors for more information when a patient's chart lacks details, training coding staff to more effectively read documentation, and hiring coders with more experience. Hospitals could purchase health information technology that automatically suggests high-value codes and that prompts staff to look for and query doctors about these codes. Hospitals also could expend resources creating ordeals for physicians who fail to provide detailed documentation. The view that physician habits are expensive for the hospital to change matches accounts of quality improvement efforts that sought to make reluctant physicians prescribe evidence-based medicines, wash their hands, and perform other tasks to improve clinical outcomes (Voss and Widmer, 1997; Stafford and Radley, 2003; Pittet et al., 1999).

2.6 Clinical Costs of Takeup

One possibility is that taking up the reform requires medical testing of HF patients to confirm the details of their conditions. The minimum information needed to use a specific code is a statement of whether there is systolic or diastolic dysfunction. Echocardiograms are non-invasive diagnostic tests that are the gold standard to confirm these dysfunctions. Some observers proposed that the reform put pressure on physicians to perform echocardiograms that they had not considered medically necessary (Leppert, 2012). If these concerns were realized, one could interpret the adoption friction as not one of documentation, but rather the refusal of doctors and hospital staff to provide costly treatment that they perceived to lack clinical benefit.

Official coding guidelines indicate that more detailed HF coding did not have to involve changes in real medical treatment. The coding guidelines state that "if a diagnosis documented at the time of discharge is qualified as 'probable,' 'suspected,' 'likely,' 'questionable,' 'possible,' or 'rule out,' the condition should be coded as if it existed or was established" (Prophet, 2000). Clinically, the

information to diagnose and submit a vague HF code typically enables the submission of a specific HF code – a patient's medical history and symptoms are predictive of the type of HF – and time series evidence is consistent with this view. Appendix Figure A2 shows no perceptible change in heart testing rates (echocardiograms) around the reform..

A more systematic test of the correlation between HF coding and treatment suggests that heart testing can account for only a small fraction of the rise in detailed coding. In Appendix Table A2 I partition patients into 25 groups using major diagnostic categories (MDCs), an output of the DRG classification system that is based on the patient's principal diagnosis. For each group, I calculate its *ex ante* HF rate using 2003-2004 patients and analyze how its detailed HF coding and echocardiogram rates grew between the pre-reform (2005-2007) and post-reform (2008-2010) eras. Unsurprisingly, groups with a greater fraction of HF patients *ex ante* were more likely to grow their detailed HF coding rates: for each additional 10 percentage points of HF *ex ante*, detailed coding later rose by 4.5 percentage points. These groups were also more likely to grow their echocardiogram rates, but the growth was one-eighth that of detailed HF coding, with an additional 10 percentage points of HF *ex ante* associated with a 0.6 percentage point higher echocardiogram rate later.

3   Econometric Strategy

In this section, I describe my approach for analyzing the roles that hospitals and physicians played in the adoption of the revenue generating practice. I decompose coding into the component that is due to the facility and the component that is due to its doctors. The notion of outcomes being due to a hospital and doctor component follows a common econometric model of wages that decomposes them into firm and worker effects (Abowd et al., 1999; Card et al., 2013).

This approach enables two key hospital analyses. First, it uncovers the dispersion in the adoption of detailed HF coding among observably similar hospitals and shows whether it is robust to removing the physician component of coding – that is, it tests whether dispersion would persist even if hospitals had the same doctors. Second, it admits a study of the relationship between adoption and hospital factors like EMRs, financial integration with physicians, and clinical quality. Later, I apply the same approaches to study the dispersion in and correlates of physician coding.

3.1 Specification

The key analyses describe the distribution of the adoption of the coding practice with two-step methods. The first step extracts a measure of adoption at the hospital level, which is the hospital effect given in equation 1. This fixed effect is the probability that a HF patient in the hospital receives a detailed HF code, after adjusting for patient observables and physician effects. In the second step, I analyze the distribution of the fixed effects by calculating their standard deviation (to look for variations among seemingly similar enterprises) and by regressing them on hospital characteristics and clinical performance (to see which facilities are most likely to adopt).

3.1.1 First Step: Estimating Hospital Fixed Effects

In the first step, I run the regression given in equation 1. I consider versions of this regression with patient controls of varying degrees of richness, and run these regressions both with and without physician fixed effects. I then extract estimates of the hospital fixed effects $\hat{\alpha}_h$. These estimates equal the share of HF patients at the hospital who received a specific code $(\overline{code}_h)$ less the contribution of the hospital's average patient $(\overline{X}_h\hat{\beta})$ and the patient-weighted average physician effect $(\frac{1}{N_h}\sum_{p\in P_h}\hat{\alpha}_{d(p)}$, where N_h is the number of HF patients at the hospital, P_h indexes the patients, and $d\,(p)$ indicates the doctor that attended to patient p):

$$\hat{\alpha}_h = \overline{code}_h - \overline{X}_h\hat{\beta} - \frac{1}{N_h}\sum_{p\in P_h}\hat{\alpha}_{d(p)}$$

In the simplest specification, which includes no patient controls nor physician fixed effects, the estimates of the hospital fixed effects $\hat{\alpha}_h$ become the shares of HF patients in hospital h who receive a specific HF code:

$$\hat{\alpha}_h^{simple} = \overline{code}_h \tag{3}$$

There are two caveats to using this measure, both of which can be seen by taking the difference between $\hat{\alpha}_h^{simple}$ and $\hat{\alpha}_h$:

$$\hat{\alpha}_h^{simple} - \hat{\alpha}_h = \overline{X}_h\hat{\beta} + \frac{1}{N_h}\sum_{p\in P_h}\hat{\alpha}_{d(p)}$$

15

One is that heterogeneity in $\hat{\alpha}_h^{simple}$ may be due to patient-level factors $\overline{X}_h\hat{\beta}$ that have been shifted to the error term of the simple measure. For example, dispersion in coding could reflect that some hospitals have patients who are difficult or less profitable to code. The specifications with rich sets of patient observables aim to address this concern. When patient-level factors are included, the use of hospital (and potentially physician) fixed effects means that the coefficients on patient characteristics are estimated from the within-hospital (and potentially within-physician) relationships between these characteristics and coding.

The second caveat is that dispersion could also reflect the role of physicians in coding, $\frac{1}{N_h}\sum_{p\in P_h}\hat{\alpha}_{d(p)}$ – some hospitals may have doctors who are particularly willing or unwilling to provide detailed documentation of their patients. Whether the physician component should be removed depends on the aim of the analysis, since the physician's actions inside the hospital are a component of the hospital's overall response to the reform. For example, hospitals with much to gain from the reform may be more likely to teach their physicians how to recognize the signs and symptoms of HF. These physicians would then be more likely to document specific HF in any hospital. Controlling for the physician effects would sweep out this improvement. Still, the extent to which the response to the reform is driven by changes in hospital behavior above and beyond the actions of its physicians is of interest in identifying the performance of the facility itself, which could reflect the performance of its own coding systems as well as how it resolves agency issues.

3.1.2 Second Step: Describing the Distribution of the Hospital Fixed Effects

This section explains the analyses of the $\hat{\alpha}_h$ and how they account for estimation error due to sampling variance.

Dispersion among Similar Hospitals The first key analysis of this paper studies the dispersion of the hospital fixed effects. However, the objects $\hat{\alpha}_h$ are noisy – though unbiased – estimates of α_h, meaning that their dispersion will be greater than the true dispersion of α_h. This noise comes from small samples at the hospital level (some hospitals treat few HF patients) and imprecision in the estimates of the other coefficients in the model. When the specification lacks physician fixed effects, the only other coefficients in the model are at the patient level, and are estimated from millions of observations. These coefficients are estimated precisely, reducing the role for this noise.

When the specification includes physician fixed effects, the imprecision of the hospital effects grows as the variation available to identify the hospital component is reduced. In a simple specification with no patient-level characteristics, the hospital effects are identified only by patients who were treated by mobile doctors, and one component of the measurement error in the hospital effect is an average of the measurement error of those physicians' effects. As these coefficients become estimated more precisely, for example as the number of patients treated by the mobile doctors rises, the estimation error falls (for more discussion of the identification conditions see Abowd et al., 2002 and Andrews et al., 2008).

Estimates of the variance of α_h must account for measurement error in order to avoid overstating dispersion. To produce these estimates, I adopt the Empirical Bayes procedure described in Appendix C of Chandra et al. (2016a). This procedure uses the diagonals of the variance-covariance matrix from the first-step regression as estimates of the variance of the hospital fixed effect measurement error. I generate a consistent estimate of the variance of α_h by taking the variance of $\hat{\alpha}_h$ and subtracting the average squared standard error of the hospital fixed effects (i.e. the average value of the diagonals of the variance-covariance matrix).[7]

Describing the Adopters   The other key hospital analysis describes the adopters by placing the hospital fixed effect estimates on the left-hand side of regressions of the form of equation 2. The measurement error in the $\hat{\alpha}_h$ therefore moves into the error term where its primary effect is to reduce the precision of the estimates of the coefficients ρ, γ, and θ. Since the measurement error is due to sampling variance in the first step, it is not correlated with the characteristics and performance measures that are found on the right-hand side of the key regressions, and it does not bias the estimates of the coefficients.

3.2   Separate Identification of Hospital and Physician

The HF context allows the separate identification of hospital and physician contributions to takeup. The key insight behind the decomposition is that physicians are frequently observed treating patients

---

[7]The Chandra et al. (2016b) procedure uses an iterative approach to develop optimal weights and then uses these weights when taking averages. The optimal weights would favor hospitals with more precisely estimated fixed effects, i.e. those with more patients treated by mobile physicians. To prevent bias in the estimated standard deviation that would occur if underlying dispersion is correlated with the volume of identifying patients, in these estimates I give each hospital equal weight and take simple averages.

at multiple hospitals, since doctors may have admitting privileges at several facilities. When the same physician practices in two hospitals, her propensity to provide detailed documentation at each facility identifies the hospital effects relative to each other. Likewise, when two physicians practice at the same hospital, their outcomes at that hospital identify the physician effects relative to each other.

The hospital and physician effects can be separately identified within a mobility group – the set of doctors and hospitals that are connected to each other by shared patients. Consider the graph of doctors and hospitals, in which each doctor and hospital is represented by a point (i.e. a node in graph theory). In the graph, a doctor and hospital have a line (i.e. an edge) drawn between their nodes if the doctor treats a patient at that hospital. Two hospitals or doctors are connected if there exists any unbroken sequence of lines (i.e. a path) going from one to the other in the graph. A mobility group starts with a doctor or hospital and includes all other doctors and hospitals that are connected to her or it (a maximal connected subgraph).

The key assumption of the econometric model here is that the probability that a patient receives a specific code must approximate a linear probability model with additive effects from the patient, hospital, and doctor such that:

$$\mathbb{E}\left[code_{ph}\right] = \alpha_h + \alpha_d + X_p\beta$$

Though the idea the three levels are linear and additively separable is clearly an approximation given the binary nature of the outcome, the additivity assumption can be tested by estimating a match effects model (Card et al., 2013). This model replaces the hospital and physician fixed effects with a set of effects at the hospital-physician level (i.e. $\alpha_{h,d}$), allowing any arbitrary relationship between hospital and physician types. The match effects model improves explanatory power minimally, suggesting that additivity is not a restrictive assumption in this context.[8]

The additive model does not structure the matching process between hospital types and physician types. Hospitals drawn from one part of the hospital type distribution may systematically

---

[8]Specifically, the adjusted R^2 of the first-step regression with hospital fixed effects, physician fixed effects, and the full set of patient controls is 0.369, while the adjusted R^2 of the same regression with the two sets of fixed effects replaced by one level of hospital-physician match effects is 0.372. The match effects model is also inferior by the Bayesian information criterion (BIC), a test of overfitting that values model explanatory power but penalizes complexity in the form of the number of coefficients being estimated.

match with physicians drawn from any part of their type distribution. Likewise, the model makes no assumption about the relationship between physician type and mobility status: mobile physicians may be drawn from a different part of the physician type distribution than their non-mobile counterparts.

Instead, the principal threats to identification are twofold. First, the conditional expectation equation implies that patients do not select hospitals or doctors on the basis of unobserved costs of coding. If such selection were to occur, the fixed effect of a hospital with unobservably more costly to code patients would, for example, be estimated with negative bias. In practice, I test this assumption by including a rich set of patient characteristics as controls. Adding controls yields qualitatively similar, albeit somewhat attenuated, results.

A second identification requirement is that the assignment of doctors to hospitals must not reflect match-specific synergies in the coding outcome. Though there may be an unobserved component of coding that is due to the quality of the match, the matching of doctors and hospitals must not systematically depend on this component (Card et al., 2013). For example, one hospital might demand more specificity in HF coding from physicians who were directly employed by the facility. These physicians would have positive match effects with that hospital. If they tended to practice at the hospital, the match effects would load onto the hospital effect, biasing it upward. The role of match-specific synergies is bounded by the match effects model described in footnote 8 – the low explanatory improvement of that model indicates that the size of these synergies must be small, limiting the scope for endogeneity from this source.

4   Analysis and Results

4.1   Analysis Sample

I use the grand sample described in Section 2.3 to construct an analysis sample of hospitals' claims to Medicare for their HF patients. I start with the 1.9 million HF patients across 3,414 hospitals from 2010. For 1.6 million (84%) of these stays across 3,381 hospitals and 136,067 physicians, I observe the patient's history of chronic conditions as well as the attending physician, who was primarily in charge of taking care of the patient in the hospital and thus most responsible for the

final diagnoses that were coded and submitted on the hospital's claim.[9] Hospital and physician types are only separately identified within the mobility group described in Section 3.2. I call the first-step analysis sample the set of 1.5 million patient claims that occur within the largest mobility group of hospitals and physicians – 80% of the grand sample of HF claims in 2010.

This sample is described in Table 2. There are 2,831 hospitals and 130,487 doctors in the sample. The average hospital sees 534 HF patients in 2010 and its HF patients are treated by 57 distinct doctors. At the average hospital, 19 of these doctors are mobile, which means that they are observed treating at least one HF patient at another hospital. In this sample, the average doctor sees 12 HF patients in a given year and works at 1.23 distinct hospitals. About one-fifth of doctors are mobile.

Table 3 provides additional information about the doctors by mobility status using data from the AMA Masterfile.[10] The average mobile physician treats about twice as many patients as a non-mobile physician.[11] Mobile physicians are more likely to be primary physicians like internists or medical specialists like cardiologists, and they are less likely to be women. Mobile physicians have about 8 months more training – but about 8 months less experience since completing training – than their non-mobile counterparts, and they are also more likely to have received their medical training outside the U.S. The difference in characteristics between mobile and non-mobile physicians does not invalidate the econometric model, which allows physician types to vary flexibly with mobility status. The relevant identification assumption, described in more detail in Section 3.2, is instead that physicians and hospitals do not match based on unobserved coding synergies.

---

[9]I use the attending physician identifier from the Medicare Inpatient RIF. To ensure that only valid individual physicians are included, I drop patients with physician identifiers that could not be found in the AMA Masterfile, a census of all physicians.

The small literature on identifying the attending physician in Medicare claims has suggested looking at physician claims (found in the Medicare Carrier RIF) and choosing the physician who bills Medicare for the most evaluation and management services, rather than the physician indicated by the hospital on its inpatient claim (Trude, 1992; Trude et al., 1993; Virnig, 2012). There are two advantages to using the hospital's report, however. First, the hospital's report of the attending physician may more accurately reflect the physician with whom the facility was communicating to determine the patient's diagnosis codes. The literature on identifying the physician is more concerned with the most medically responsible physician, not the one most responsible for billing and coding. Second, I only observe physician claims for a 20% random sample of patients, dramatically restricting the set of patients for whom I observe the physician when using the physician claim method.

[10]About one-fifth of the physicians are "singletons" observed treating only one patient. These physicians do not contribute to identification of the hospital effects (see e.g. Correia, 2015) and make the algorithm later used to estimate the dispersion in physician effects unstable. For symmetry with the later physician analysis, this table excludes singleton physicians.

[11]Specialties are grouped according to the Dartmouth Atlas definitions. See Table 2 of the document found at http://www.dartmouthatlas.org/downloads/methods/research_methods.pdf

4.2 Hospital Characteristics

Table 4 shows summary statistics for the 2,341 hospitals in the main analyses for which I observe complete information on all covariates – the second-step analysis sample. Hospital size (beds) and ownership are taken from the Medicare Provider of Services file. Hospital location and teaching status are taken from the 2010 Medicare IPPS Impact file. The location definition is the one used by Medicare: a large urban area is any Metropolitan Statistical Area (MSA) with a population of at least 1 million, an other urban area is any other MSA, and the rest of the country is considered rural. Only 22% of the hospitals in the sample are rural – many rural hospitals are classified as critical-access facilities exempt from this reform, and they are excluded from my analyses. Teaching hospitals are defined as those with any residents; major teaching facilities are the 10% with a resident-to-bed ratio of at least 0.25; minor teaching facilities are the 28% that have a resident-to-bed ratio greater than zero but less than 0.25.

I define the *ex ante* revenue at stake as the expected value of giving all of the hospital's pre-reform (2007) HF patients a specific code according to post-reform (2009) reimbursement rules. The revenue at stake is scaled by the total number of patients at the hospital, making it the per-patient expected gain from fully taking up the reform (see Appendix Section A.1.2).

Hospital EMR adoption comes from Healthcare Information and Management Systems Society (HIMSS) data and is classified into basic and advanced according to the approach in Dranove et al. (2014). 6% of hospitals do not have an EMR, half have EMRs with only a basic feature (clinical decision support, clinical data repository, or order entry), and 43% have EMRs with an advanced feature (computerized practitioner order entry or physician documentation). To measure hospital-physician integration, I use the American Hospital Association (AHA) hospital survey data and follow Scott et al. (2016) to group hospitals by the tightest form of integration that they report. I divide arrangements into no relationship (31%), contract (17%), and employment (35%); if the hospital did not respond to the question or described integration using a freeform text field, I classify the hospital as having an unknown or other relationship (18%).

The standards of care measures were collected by CMS under its Hospital Compare program and are described in greater detail in Appendix Section A.2.1. They indicate the shares of times that standards of care were followed for AMI, HF, pneumonia, and high-risk surgery patients in 2006.

21

These standards of care are inexpensive, evidence-based treatments that were selected because they had been shown to improve patient outcomes and aligned with clinical practice guidelines (Williams et al., 2005; Jencks et al., 2000). When productivity is defined as the amount of survival a hospital can generate for a fixed set of inputs, these scores measure the takeup of productivity-raising technologies. They notably include β-blockers, a class of inexpensive drugs that dramatically improve survival following AMI and have been the subject of several studies of technology diffusion (see e.g. Skinner and Staiger, 2007, 2015).

Adjusted AMI survival is based on the sample and methods of Chandra et al. (2013) and its construction is described in Appendix Section A.2.2. A form of treatment performance, a hospital's adjusted survival is the average 30-day survival rate of AMI patients treated at the hospital in 2000-2006, after controlling for the inputs used to treat the patient and a rich set of patient observables. An increase in the rate of 1 percentage point means that, at the same level of inputs and for the same patient characteristics, the hospital is able to produce a 1 percentage point greater probability that the patient survives 30 days. This rate is adjusted to account for measurement error using an Empirical Bayes shrinkage procedure. The survival rate at the average hospital is 81%, and the standard deviation of that rate across facilities, after accounting for differences in patient characteristics, input utilization, and measurement error, is 3 percentage points.

4.3 Dispersion across Hospitals

I now assess dispersion in hospital adoption and its sensitivity to patient and physician controls (for dispersion at the hospital system and geographic region levels, see Appendix Section A.3). To provide a sense of the time series of adoption, Figure 4 shows the distribution of raw $\hat{\alpha}_h^{simple}$, the share of HF patients at hospital h who received a detailed HF code, in each year from 2007 to 2010. Takeup across hospitals occurred rapidly after the reform. By 2010, the third year after the reform, the median hospital used detailed codes 55% of the time. Variation was substantial: a mass of hospitals used the codes for the vast majority of their HF patients while a nontrivial number of hospitals almost never used them.[12]

Table 5 shows the standard deviation of adoption overall and among homogeneous categories

---

[12]To limit the scope for measurement error, the figure uses all patients in the grand sample from 2007 to 2010 from hospitals with at least 50 HF patients in that year. Appendix Table A3 shows that the standard deviation of coding across hospitals is similar among the sample definitions.

of hospitals. I divide the space of hospitals on the basis of characteristics that have been the focus of literature on hospital quality. The left three columns estimate dispersion using varying sets of patient controls and no physician controls; in these results, the hospital effects include the component of coding that is due to the physicians. The final column adds first-step physician effects, which subtracts the physician component.

The controls are described briefly here and in full detail in Appendix Section A.4. Column 1 uses no patient-level controls. Column 2 controls for observables about the patient's hospital admission found in the hospital's billing claim: age, race, and sex interactions; whether the patient was admitted through the emergency department; and finely grained categories for the patient's primary diagnosis. Richly controlling for the patient's principal diagnosis also helps to account for the patient-level return to the detailed codes, which will vary depending on the patient's DRG. Column 3 adds indicators for a broad set of chronic conditions. To improve comparability across analyses, the table only includes hospitals for which all covariates are observed.

Among all hospitals, the standard deviation of the coding scores with no controls is 0.20 (column 1), meaning that a hospital with one standard deviation greater adoption gives 20 percentage points more of its HF patients a specific HF code. This measure does not account for differences in patient or doctor mix across hospitals. With patient observables on admission included, the standard deviation falls to 0.15 (column 2). Additionally controlling for patient illness histories has little further effect (column 3). This dispersion is the standard deviation across hospitals of the probability a HF patient gets a specific code, holding fixed the patient's observed characteristics. It calculates adoption across hospitals after removing the component that can be explained by within-hospital relationships between patient observables and coding. Further adding physician fixed effects raises the standard deviation slightly to 0.16 (column 4). This result is the dispersion across hospitals in the probability a specific code is used, given a HF patient with a fixed set of characteristics and a fixed physician. With these controls, a hospital with one standard deviation greater adoption is 16 percentage points more likely to give a patient a specific code.

Within key groups of hospitals, dispersion tends to decline with the inclusion of patient characteristics in the first step; the additional inclusion of physician fixed effects yields smaller changes in magnitude of varying sign. Specifically, dispersion declines by 4-6 percentage points with the inclusion of patient characteristics; the additional inclusion of physician effects yields changes ranging

from a decline of 2 percentage points to a rise of 3 percentage points.

While it may seem counterintuitive that disparities in adoption sometimes increase with the addition of physician controls, this finding is possible if high type hospitals tend to match with low type physicians. When physician controls are omitted, the hospital's adoption includes both the facility component and an average physician component. If dispersion in adoption rises after removing the physician component, it indicates that the average physician component was negatively correlated with the hospital component – evidence of negative assortative matching. While the econometric model assumes additivity of hospital and physician effects, it is agnostic about the matching process, permitting assortative or non-assortative matching.

4.4 Describing the Adopters

Having found evidence of disparities in adoption even after accounting for patients and physicians, in this section I turn to the characteristics that are associated with adoption. That is, I estimate equation 2 by regressing the hospital adoption measures (estimated with varying patient and physician controls) on the hospital characteristics. I first discuss what existing literature on hospital performance suggests for the *ex ante* relationships one might expect between hospital covariates and HF coding. I then show how these correlations are borne out in my data.

In Appendix Section A.3, I consider two additional explanatory factors: hospital system and geographic region. While most variation in hospital adoption is not explained by either level, the explanatory power of each is non-trivial: system and region fixed effects account for as much as one-fourth of variation with physician effects *not* swept out, and one-fifth of variation with physician effects removed.

4.4.1 Potential Roles of Hospital Characteristics

Size (Number of Beds)   A long line of research has documented a relationship between hospital size and quality, though with an unclear causal link. Epstein (2002) provides a critical review of this association, called the volume-outcomes hypothesis. Likewise, a scale-coding relationship could be the result of several factors. It could derive from features of the code production process. As with clinical quality, it could reflect that hospitals learn by doing, and large hospitals have more patients to learn from. Larger hospitals would also be more likely to adopt detailed HF coding if

there were fixed costs of adoption – the return on these fixed costs is greater when they yield better coding on a bigger patient population. In this context, fixed costs could include health information technology software (though I study this possibility directly by looking at EMRs). Lastly, a scale-coding gradient could be the incidental result of an omitted third factor, though the correlation between size and coding could still be of interest for policymakers seeking to understand the effects of the reform on distribution and which facilities are likely to respond in the future.

Ownership   While there is no consensus on whether non-profit or for-profit hospitals provide superior quality of care (see e.g. McClellan and Staiger, 2000; Sloan et al., 2003; Joynt et al., 2014), the disparities have been clearer in studies of billing and coding, which have found that for-profit hospitals exploited revenue-making opportunities more aggressively than their non-profit and government-run counterparts (Dafny, 2005; Silverman and Skinner, 2004). Earlier work has typically focused on upcoding, or the exaggeration of patient severity to raise payments. Here, a hospital can provide a detailed HF code for all its HF patients with detailed documentation but no upcoding (upcoding would entail submitting a detailed code that lacked supporting documentation).

Location   Research has considered differences in clinical performance between urban and rural facilities, but whether rural hospitals should be more effective at adopting the revenue-raising technology than urban hospitals holding scale fixed is unclear *ex ante*. Evidence on outcomes and processes along the dimension of hospital location may be suggestive. Most of the literature has found that health care outcomes and clinical quality are lower in rural hospitals relative to their urban counterparts, a finding that persists even conditional on hospital size (MEDPAC, 2012; Baldwin et al., 2010; Goldman and Dudley, 2008).

Teaching Status   Teaching hospitals have better outcomes and higher quality processes of care than non-teaching hospitals (Ayanian and Weissman, 2002; Mueller et al., 2013; Burke et al., 2017). Beyond the academic literature, teaching hospitals appear to be regarded in conventional wisdom as purveyors of the frontier of high quality care (see, for example, *U.S. News and World Report* rankings of hospitals). Whether this conventional wisdom is true, and whether it translates into more responsiveness to incentives in the form of takeup of the revenue-generating practice, is an open question – for example, the presence of residents who lack prior experience with hospital

documentation and billing needs may act as a drag on a hospital's coding, while the need to document extensively for training purposes could improve coding.

Revenue at Stake   A hospital with more revenue at stake from the reform, all else equal, would have a greater incentive to buy software that improves specific coding and to coax its doctors to provide detailed documentation. The revenue at stake depends on the hospital's patient mix – hospitals with more HF patients and hospitals with more acute HF patients have more to gain. However, even after controlling for a host of observables about the hospitals, unobserved characteristics may still exert an effect on adoption along this gradient, since patient mix and acuity may be correlated with other attributes about the hospital that independently affect its coding (for example, after conditioning on characteristics, having revenue at stake could be correlated with having a safety net role and thus other associated but unobserved factors). In the regressions, I calculate revenue at stake per patient, rather than total revenue at stake or total number of HF patients. The total revenue measures are closely related to hospital size (in practice, the total measures are not conditionally correlated with adoption).

Electronic Medical Records   Hospital EMRs may facilitate detailed coding by reducing the cost for physicians of providing additional information. EMRs can also prompt physicians to provide documentation or copy it over automatically from older records (Abelson et al., 2012). While most of the literature on EMRs centers on their potential to improve the quality of care, some quasi-experimental work considers its effects on documentation and coding. Two studies find that EMRs raise coding intensity, albeit with different magnitudes: Li (2014) shows a significant increase in the fraction of patients in high-severity DRGs as hospitals submit more diagnosis codes in their claims, while Agha (2014) shows that hospital payments rise due to EMR adoption, but that increased coding intensity explains only 7% of the change.

Hospital-Physician Integration   Physicians traditionally practiced at hospitals without formal contractual or employment relationships, billing insurers directly for the care they provided. In recent years hospitals and physicians have come to integrate more closely (Scott et al., 2016). The tightest form of integration occurs when physicians are directly employed by hospitals; an intermediate form occurs when physicians or their group practices sign contracts with the hospital establishing

a relationship. Multiple studies have shown that such integration raises the prices providers receive from private insurers, either by increasing the bargaining power of the integrated unit or because Medicare's administrative pricing rules favor integrated entities (Baker et al., 2014; Neprash et al., 2015).

Tighter hospital-physician integration has the potential to increase coding rates by aligning the revenue objectives of physicians with those of the hospital. In addition, the federal Anti-Kickback Statute, which restricts how hospitals pay physicians, does not apply to employment relationships (BNA, 2017). While there is little evidence on how hospital-physician integration affects documentation and coding, integration in other areas of health care can improve coding: Geruso and Layton (2015) show that private insurance plans in Medicare that integrate with health care providers raise the coded severity of their patients more than unintegrated plans, leading to increased federal capitation payments.

Clinical Performance and Quality Whether high treatment performance hospitals are more likely to adopt the coding practice is not obvious. High quality hospitals may have good managers who effectively work with physicians to incorporate consensus standards of care – a correlation that has been observed in U.S. hospital cardiac care units (McConnell et al., 2013). These managers may use the same techniques to extract more detailed descriptions from their physicians. The managers could also use their treatment performance-raising techniques to ensure that coding staff does not miss revenue-making opportunities.

On the other hand, a negative correlation between treatment quality and revenue productivity is also plausible. To the extent that productivity depends on managerial quality, the relationship between revenue productivity and treatment quality could reflect whether one is a substitute for another in the hospital management production process. In the substitutes view, managers specialize in either coaxing physicians and staff to extract revenue from payers or in encouraging them to treat patients well.

4.4.2 Results

Table 6 displays estimates of the correlation between hospital characteristics and takeup of detailed HF coding. The columns of this table show the results when different sets of first-step controls are

included. These specifications match those used in the dispersion analysis. The hospital effects are estimated with noise, adding left-hand side measurement error to the regressions. This measurement error comes from sampling variance, so it does not bias the coefficients.

Without Physician Controls Columns 1 to 3 depict the correlations with increasingly rich first-step patient controls but no physician controls. They establish several relationships of interest. There is a coding-scale relationship: hospitals that are 10% larger give 0.18 percentage points more of their HF patients a specific code. Adding patient controls reduces this effect to 0.11 percentage points (significant at the 10% but not 5% level) – some of the raw relationship between size and coding can be accounted by larger hospitals tending to have patients that are more likely to receive a detailed code at any hospital. Hospital ownership matters: controlling for patients, non-profit hospitals give 2.8 percentage points more of their patients a specific code than government-run facilities. There is no statistically significant difference between the takeup rates of for-profit and government-run hospitals. Major (but not minor) teaching hospitals are significantly more likely to provide detailed codes than non-teaching facilities, a difference of 2.8 percentage points with patient controls.

Hospitals with basic EMRs are significantly less likely to provide detailed codes than hospitals with advanced EMRs or no EMRs at all. By 2010 only 6% of hospitals lack an EMR, so the comparison between basic and advanced may be of greater relevance: I find that hospitals with advanced EMRs provide a detailed code for 2.4 percentage points more patients than hospitals with basic EMRs, and this difference is highly significant. Employment of physicians is also a significant predictor of detailed coding, with these hospitals providing the code for 1.8 percentage points more patients than hospitals without formal relationships with their doctors.

Hospitals that appear to be higher quality in their treatment are also more likely to use these high-revenue billing codes. With the full set of patient controls, for each standard deviation rise in the use of standards of care, 1.9 percentage points more HF patients tend to get a specific code. The effect for each standard deviation rise in AMI survival is 2.5 percentage points.

With Physician Controls Column 4 repeats the results of column 3 with first-step physician controls, slightly changing the interpretation of the coefficients. In these columns, a positive (neg-

ative) relationship between a hospital characteristic and coding indicates that the facility was able to extract more (less) detailed coding, holding physicians fixed.

The gradient between hospital size and extraction of detailed HF codes is positive without first-step physician controls, but its point estimate becomes negative when the physician component of adoption is removed. Though statistical power is low, this finding suggests that larger hospitals outperform smaller hospitals in column 3 because they utilize physicians that provide more documentation wherever they treat patients.

Non-profit and for-profit hospitals were 2.0 and 3.1 percentage points, respectively, more likely to extract specific codes from doctors than their government-run counterparts, though these coefficients were not significant at the 5% level (the latter is significant at the 10% level). Compared to the differential unconditional on physicians, the point estimate for non-profit hospitals is reduced by about one-third and no longer significant. Since removing the physician component of adoption removes the coding advantage of non-profit facilities, my results imply that the physicians who work at non-profit hospitals are somewhat more likely to provide the detailed documentation wherever they practice. On the other hand, the for-profit effect expands, suggesting that these facilities have physicians that are less likely to provide detailed documentation wherever they work, but that the low physician contribution is counteracted by the hospitals' ability to extract codes from their doctors.

Hospitals in large urban and other urban areas – areas of high and intermediate population, respectively – extract specific codes from their doctors for about 3 percentage points more of their patients than hospitals in rural areas, though only the effect for other urban areas is significant at the 10% level. This relationship is muted without the physician controls, which indicates that urban hospitals, like for-profit hospitals, are more effective at extracting the codes but are held back by their physicians. For major teaching facilities, the gradient expands from 2.8 to 4.5 percentage points with the removal of the physician component, suggesting that physicians also hold back these hospitals, but here the hospital effect is so big that these hospitals still outperform their non-teaching peers unconditional on physicians.

The result for EMRs attenuates with the addition of physician controls; hospitals with advanced EMRs submit detailed codes for 1.5 percentage points more of their patients than hospitals with basic EMRs, but the difference is not significant. Hospital-physician relationships are still associated

29

with adoption; the magnitudes grow but are estimated with more imprecision. The point estimate for employment affiliation is qualitatively similar but is now significant at the 10% level; physician employment is associated with a 2.4 percentage point higher coding rate. The effect of contract affiliation grows under this specification to a meaningful but statistically insignificant 2.1 percentage points.

Finally, the use of detailed HF codes is correlated with both AMI survival and the use of consensus standards of care: conditional on patients and doctors, hospitals with one standard deviation greater use of standards of care or one standard deviation greater treatment performance use specific codes for 1.7 and 3.0 percentage points (respectively) more of their patients. A similar gradient was also observed unconditional on the doctors (in columns 1 to 3) – these results indicate that it cannot be explained by high treatment quality hospitals simply having physicians that provide detailed documentation wherever they practice. Instead, these results indicate that these hospitals are more likely to extract the codes from their physicians than their lower treatment quality peers.

4.5 Dynamics of Adoption

In Table 7, I present evidence on the dynamics of adoption by showing how coding at different points in time correlates with hospital characteristics. Columns 1-4 show the results looking at 2008Q1, the full year 2008, the full year 2010 (i.e. the main analysis sample), and 2010Q4, respectively. Each column regresses hospital coding scores from the time period on the full set of characteristics, but estimates coding using patients from different time periods. The regressors are also taken from 2008 and 2010 data to match the regressands, except *ex ante* revenue and the clinical measures which are as described in Section 4.2. In the first-step estimation of hospital coding scores, I do not control for the physician because this would further cut the sample for the single quarter analyses: some physicians are mobile during the year but not the quarter, breaking the connections between hospitals that regression requires to separately identify the physician and the hospital.[13]

Hospitals with high clinical quality respond more quickly to the reform. The clinical quality measures are strongly and significantly associated with adoption in all time periods presented. A one standard deviation increase in use of standards of care is predicted to raise adoption by 1.7

---

[13]To construct the 2008 scores, I apply the same selection criteria I used on the 2010 analysis sample to 2008 patient data, extracting the largest connected subgraph of hospitals in that year. I then estimate equation 1, using the full set of patient controls but omitting physician fixed effects.

percentage points (2008Q1) to 1.9 percentage points (2010Q4). Likewise, a one standard deviation increase in AMI survival is associated with a 1.3 percentage point (2008Q1) to 2.6 percentage point (2010Q4) rise in adoption.

The clinical measures are the only significant coefficients at the 5% level in 2008Q1 or 2008 presented in the table. During this time, advanced EMRs are also significantly associated with adoption relative to basic EMRs (the coefficient in the table is relative to no EMRs, where the difference is not significant). Otherwise, the initial associations between coding and hospital characteristics in 2008Q1 and 2008 tend to be attenuated relative to the associations that develop by 2010, and the results for 2010 and 2010Q4 tend to be similar. Hospital size, for example, is positively associated with adoption in the very early period (and the only remaining covariate that is significant at the 10% level) but with a smaller magnitude than that observed in 2010 and 2010Q4.

4.6   Physicians

Here I study the dispersion and determinants of the physician effects in equation 1 in the same way as for the hospital effects. Table 8 presents the results.[14] In columns 1-3, the first-step model regresses coding on physician fixed effects but no hospital effects. Going from left to right, I control for patient characteristics increasingly richly, matching the approach in Table 6. Since there are no first-step hospital effects, these models allow the physician effects to also embody the effect of the average hospital at which the physician practices – just as the earlier results that did not control for physicians allowed the hospital effects to also embody the average physician practicing at the facility. Column 4 adds hospital fixed effects to the model, mimicking the first-step model of column 4 of Table 6. This model subtracts the hospital component of adoption from the physician effects.

Panel A at the top of the table shows the standard deviation of the physician effects. Since the average physician treats 14.6 patients, the raw physician effects are estimated with significant measurement error. To avoid overstating dispersion, I adjust for measurement error using a procedure adapted with minor changes from Gaure (2014) (see Appendix A.5). This procedure is similar to that used for the hospital effects, but avoids the computationally intensive process of directly calculating the diagonals of the variance-covariance matrix of the 101,370 physician effects. As

---

[14]The table omits the one-fifth of physicians who are "singletons" observed treating only one patient, who do not contribute to identification of the hospital effects and who make the dispersion estimator unstable (Correia, 2015).

with the hospital effects, dispersion shrinks with the additional of patient controls observable upon admission but is not sensitive to additional controls for chronic conditions. After full patient controls, the standard deviation is 18 percentage points. Further subtracting the hospital component of adoption shrinks the standard deviation to 15 percentage points.

Panel B regresses the physician effects on characteristics from the AMA Masterfile. Perhaps unsurprisingly given the large sample of physicians, most of the covariates are significant predictors. The analysis reveals a strong association between coding and volume: each 10% increase in volume is associated with a 0.3 percentage point increase in the share of patients getting a specific code (both with and without hospital controls). This gradient could reflect a learning-by-doing effect – or physicians who code better could simply attract more patients for other reasons. The analysis also shows that mobile physicians are less likely to code. Again, this result is consistent with several hypotheses, one of which relates to agency issues: mobile doctors, who are less attached to any particular hospital, may be more reluctant to conform to documentation practices that benefit those hospitals.

Specialization is also associated with coding. Relative to primary care physicians, surgeons are less likely to code (with and without hospital controls). Medical (i.e. non-surgical) specialists are more likely to code, but only when the hospital component of adoption is not subtracted out (column 3); isolating the physician component of adoption erases the association (column 4), suggesting that medical specialists tended to benefit from practicing at high-type hospitals. A similar result obtains for years in training. Female sex, younger age, and training in the U.S. are all associated with more coding; the effects remain but attenuate when the hospital component of adoption is subtracted out in column 4.

5 Discussion

5.1 Variations in Adoption in Context

The adoption of the coding practice was incomplete at the national level, but the national time series masks enormous heterogeneity at the level of the hospital and physician. Looking across hospitals, the rate of detailed coding has wide dispersion, with some hospitals almost never using specific codes and other hospitals almost always using them. A perhaps natural view is that in comparison

32

to other sectors of the economy, some health care providers are uniquely unable or unwilling to respond to incentives. Yet dispersion alone is not enough to make health care providers exceptional – this finding is nearly universal in the adoption of new practices and technologies.

The hallmark features of a new practices are wide variations in the level of adoption at a point in time and variation in adoption over time as takeup slowly occurs. This pattern is found in hybrid corn (e.g. Griliches, 1957), and it has also been found in health care, for example in the use of β-blockers and other evidence-based therapies (see e.g. Bradley et al., 2005 and Peterson et al., 2008). Likewise, there is persistent dispersion in productivity within narrowly defined (non-health care) industries (Fox and Smeets, 2011; Syverson, 2011) as well as the health care sector (Chandra et al., 2016b). I have shown that adoption of the HF coding technology across hospitals follows the established pattern. An important distinction between the coding practice and the use of practices like β-blockers is that the latter have clinical payoffs and may diffuse purely from the intrinsic motivations of health care providers.

Some hospitals may submit detailed codes because their doctors are likely to provide specific documentation wherever they treat patients. Other hospitals might take up the revenue generating practice by counteracting the poor documentation habits of their physicians with facility-specific techniques, like aggressively reviewing patients' medical charts. Uniquely in the HF coding setting I can observe the component of adoption that is specific to the hospital – the extent to which a hospital can extract more details out of a constant set of physicians and patients than other hospitals.

The dispersion that I find in the hospital component of adoption is about four-fifths the raw level of dispersion. Progressively adding patient controls shows that the attenuation is entirely due to characteristics observable on admission like principal diagnosis. These controls help to account for the patient-level return to detailed coding inherent in the hospital payment system. Further controlling for patient illness histories (the remaining controls) does not affect dispersion. These results show that at least some of the raw dispersion in coding is due to selection of patients across hospitals – though even after accounting for a wide array of patient factors and physician fixed effects, the vast majority of dispersion remains.

This residual dispersion across hospitals has a standard deviation of 16 percentage points. One point of comparison is the standard deviation of the consensus clinical standards of care scores,

which measure adherence rates to evidence-based treatment guidelines. The measures of coding of HF are also effectively rates of adherence to the revenue-generating practice. To the extent that there are substantial disparities across hospitals in their adherence to clinical standards, the disparities in coding are at least as substantial. According to Table 4, the four standards of care scores have standard deviations ranging from 6 to 11 percentage points. The dispersion in the hospital component of HF coding adoption is above the top end of this range.

The magnitude of dispersion across physicians is similar to that across hospitals, with a standard deviation of 15 percentage points after removing the hospital contribution. Thus, some physicians are much more likely to support the detailed documentation of patients than others, even within the same facility. Given the restrictions hospitals face on financially incentivizing coding, such variations could naturally occur due to several sources of heterogeneity, including physicians' intrinsic motivation to provide clinically relevant documentation, their responsiveness to hospitals' non-financial efforts to encourage coding (e.g. meetings and trainings), and their internalization of hospitals' revenue objectives.

5.2 Correlates of Adoption, Agency Issues, and Other Frictions

The hospital component of adoption is robustly correlated with clinical quality – high clinical quality hospitals are able to extract more specific documentation from a fixed set of physicians than other hospitals. Moreover, hospitals that integrate with their physicians, particularly through direct employment relationships (where the result is positive and significant at the 10% level) but also through contractual relationships (where the result is positive but estimated more imprecisely), are more likely to extract the documentation. These correlations suggest – though do not prove – that agency problems could play a role in the adoption of a variety of technologies in the facility.

Incentive misalignments owing to principal-agent problems have been proposed as impediments to the adoption of new technology and to making organizational change more generally. One notable example of this view is found in Gibbons and Henderson (2012), who adopt a typology of managerial pathologies, focusing in particular on the many failures of organizations to take up practices that were widely known to be beneficial. These failures, they argue, are consistent with poor *implementation*: managers "know they're behind, they know what to do, and they're trying hard to do it, but they nonetheless cannot get the organization to get it done." (p. 34)

Implementation difficulties are particularly acute in the health care setting because facilities (in this view, the principals) and physicians (the agents) tend to be paid separately and on different bases. In the case of heart failure, physician payments from Medicare do not depend on whether a reimbursement claim uses vague or detailed diagnosis codes because by default, physicians are paid for each procedure they perform. Though hospitals might want to encourage detailed coding by paying doctors for it, doing so risks running afoul of federal laws that heavily restrict the incentives that hospitals pass on to their physicians (BNA, 2017).

My results are consistent with frictions beyond agency issues at the hospital level. Some hospitals may have outdated health IT or poorly trained billing staff who miss opportunities to detect high-value diagnosis codes; hospitals with better staff and computer systems could find it less costly to extract and submit the detailed codes. Indeed, I find evidence that EMRs play a role in adoption, particularly when the physician component is included. Though cross-sectional, this result aligns with the quasi-experimental literature showing that EMRs can raise coding intensity (Agha, 2014; Li, 2014).

As public insurers move to incentivize the adoption of consensus health care treatments, the effects that these incentives will have remain unclear. Looking at the relationships between HF coding and hospital characteristics sheds light both on the likely effects of future incentives as well as the mechanisms that drive incomplete takeup. In particular, these correlates offer evidence on which providers are likely to be responsive to financial incentives for other processes of care. To get a sense of the responsiveness, it is useful to look at the correlation between takeup and characteristics without removing the effect of the physician, since the overall response of the hospital is of interest. I have shown that bigger, non-profit, major teaching, vertically integrated, and higher treatment performance hospitals are more responsive. Likewise, hospitals with advanced EMRs are more likely to adopt than hospitals with only basic EMRs.

One reason to incentivize the use of evidence-based practices is to push lagging hospitals to take them up. Quality disparities have been a key focus of health care literature (see e.g. Fisher et al., 2003), and policymakers are increasingly using direct financial incentives with the hope of improving outcomes at low-performing hospitals. For example, the Medicare Value-Based Purchasing program is now reducing payments to hospitals that fail to use consensus standards of care or whose patients report low satisfaction with their experiences. Yet debate continues to rage over whether these

policies are having their intended effect of raising quality; according to my findings, responsive providers tend to be getting better results from treatment and are more likely to follow consensus standards of care already. Lower performance providers – i.e. those that produce less survival for a given patient and level of inputs, or those less likely to follow best practices – are less responsive. These results suggest that hospitals that are behind the curve on medical standards are also less attuned to financial incentives, which means that policies to incentivize takeup could have their least effect on the providers that need the most improvement. In turn, these programs could serve to widen disparities in the quality of care across providers.

6 Conclusion

This paper has examined the takeup of a revenue-generating practice – the use of specific, detailed codes to describe heart failure on inpatient claims – that was incentivized following a 2008 reform. I have shown that hospitals responded by rapidly changing the coding of patients in their claims. Yet this improvement was incomplete and uneven, a characteristic feature of the adoption of new technologies and practices. I have also decomposed the takeup of the practice into a component that is due to the hospital and a component that is due to its doctors. The decomposition exercise shows that, among other predictors, hospitals that had high treatment performance and followed consensus standards of care were better able to extract detailed documentation.

My results have implications for future research and policy. First, my finding that hospital-physician integration is associated with coding opens another channel through which consolidation can raise health care prices. Existing literature has demonstrated that vertical integration increases the negotiated rates that providers receive from private insurers for a given unit of billed service (Baker et al., 2014; Neprash et al., 2015). Integration could also change the coded intensity of billing for the same real unit of health care. The standard decomposition of spending into prices and quantities would attribute changes in coding intensity to the latter category. Given the association I find between vertical integration and coding, the large recent increase in consolidation, and the potential for small changes in inpatient coding intensity to raise spending substantially, this broader effect of integration is worthy of future study.

These results are also relevant as public and private insurers seek to directly raise hospital

performance by reforming health care payment systems. Principal-agent problems owing to a bifurcated system that pays doctors and hospitals on separate bases may impede the further adoption of techniques and practices that raise clinical performance. For example, when Medicare opts to pay hospitals to use evidence-based clinical practices like giving AMI patients aspirin, it trusts that the facilities will recognize the financial gains to changing their processes of care and successfully transmit the incentives to the physicians who prescribe the drugs. Yet some facilities appear much more able to recognize and transmit these incentives than others.

One potential policy to obviate the incentive transmission problem is modify the physician payment system. Provisions of the Affordable Care Act and the Medicare Access and CHIP Reauthorization Act (MACRA) require Medicare's physician payments to incentivize standards of care much as it already does for hospital payments. Yet the physician incentives to date do not necessarily target the same metrics as hospital incentives, leaving agency issues unaddressed. Aligning incentives for both hospitals and doctors could improve the effectiveness of value-based payment reforms.

A key topic for further study is obtaining direct evidence on which factors underly the variation uncovered in this research, perhaps by surveying hospitals about the potential factors. Opening the "black box" of how hospitals interact with their employees and their physicians to achieve their objectives is an important topic for future research – much as these questions are central to ongoing work in organizational economics studying firms in other sectors of the economy.

References

Abelson, Reed, Julie Creswell, and Griff Palmer. 2012. "Medicare Bills Rise as Records Turn Electronic." *The New York Times.*

Abowd, John M., Robert H. Creecy, and Francis Kramarz. 2002. "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data." Longitudinal Employer-Household Dynamics Technical Papers 2002-06, Center for Economic Studies, U.S. Census Bureau.

Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. "High Wage Workers and High Wage Firms." *Econometrica*, 67(2): 251–333.

Agha, Leila. 2014. "The effects of health information technology on the costs and quality of medical care." *Journal of Health Economics*, 34 19–30, URL: http://dx.doi.org/10.1016/j.jhealeco.2013.12.005.

American Hospital Association. 2005. "The Fragile State of Hospital Finances." January, URL: http://www.aha.org/content/00-10/05fragilehosps.pdf.

Andrews, M. J., L. Gill, T. Schank, and R. Upward. 2008. "High wage workers and low wage firms: negative assortative matching or limited mobility bias?" *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(3): 673–697.

Ayanian, John Z., and Joel S. Weissman. 2002. "Teaching Hospitals and Quality of Care: A Review of the Literature." *Milbank Quarterly*, 80(3): 569–593.

Baicker, Katherine, and Amitabh Chandra. 2004. "Medicare Spending, the Physician Workforce, and Beneficiaries' Quality of Care." *Health Affairs*, Web Exclusive W4–184–97.

Baker, L. C., M. K. Bundorf, and D. P. Kessler. 2014. "Vertical Integration: Hospital Ownership Of Physician Practices Is Associated With Higher Prices And Spending." *Health Affairs*, 33(5): 756–763, URL: http://dx.doi.org/10.1377/hlthaff.2013.1279.

Baldwin, Laura-Mae, Leighton Chan, C .Holly A. Andrilla, Edwin D. Huff, and L. Gary

Hart. 2010. "Quality of Care for Myocardial Infarction in Rural and Urban Hospitals." *The Journal Of Rural Health*, 26(1): 51 – 57.

Bartelsman, Eric J., John C. Haltiwanger, and Stefano Scarpetta. 2013. "Cross-Country Differences in Productivity: The Role of Allocation and Selection." *The American Economic Review*, 103(1): 305–334.

Bloom, Nicholas, Christos Genakos, Raffaella Sadun, and John Van Reenen. 2012. "Management Practices Across Firms and Countries." *The Academy of Management Perspectives*, 26(1): 12–33.

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen. 2016. "Management as a Technology?" Working Paper 22327, National Bureau of Economic Research.

BNA, Bloomberg. 2017. "Health Care Program Compliance Guide, Chapter 1805 Hospital Incentives to Physicians."Technical report.

Bradley, Elizabeth H., Jeph Herrin, Jennifer A. Mattera, Eric S. Holmboe, Yongfei Wang, Paul Frederick, Sarah A. Roumanis, Martha J. Radford, and Harlan M. Krumholz. 2005. "Quality Improvement Efforts and Hospital Performance: Rates of Beta-Blocker Prescription After Acute Myocardial Infarction." *Medical Care*, 43(3): 282–292.

Burke, Laura G., Austin B. Frakt, Dhruv Khullar, E. John Orav, and Ashish K. Jha. 2017. "Association Between Teaching Status and Mortality in US Hospitals." *JAMA*, 317(20): , p. 2105, URL: http://dx.doi.org/10.1001/jama.2017.5702.

Card, David, Jörg Heining, and Patrick Kline. 2013. "Workplace Heterogeneity and the Rise of West German Wage Inequality." *The Quarterly Journal of Economics*, 128(3): 967–1015.

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson. 2013. "Healthcare Exceptionalism? Productivity and Allocation in the U.S. Healthcare Sector." Working Paper 19200, National Bureau of Economic Research.

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson. 2016a. "Healthcare Exceptionalism? Performance and Allocation in the U.S. Healthcare Sector." *American Economic Review*, Forthcoming.

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson. 2016b. "Productivity Dispersion in Medicine and Manufacturing." *American Economic Review Papers & Proceedings*, 106(5): 99–103.

Chandra, Amitabh, and Douglas O. Staiger. 2007. "Productivity Spillovers in Health Care: Evidence from the Treatment of Heart Attacks." *Journal of Political Economy*, 115(1): 103–140.

Collard-Wexler, Allan, and Jan De Loecker. 2015. "Reallocation and Technology: Evidence from the US Steel Industry." *American Economic Review*, 105(1): 131–71.

Conley, Timothy G., and Christopher R. Udry. 2010. "Learning about a New Technology: Pineapple in Ghana." *American Economic Review*, 100(1): 35–69.

Correia, Sergio. 2015. "Singletons, cluster-robust standard errors and fixed effects: A bad mix."Technical report.

Correia, Sergio. 2016. "Linear Models with High-Dimensional Fixed Effects: An Efficient and Feasible Estimator."Technical report, Working Paper.

Cutler, David M., Mark McClellan, and Joseph P. Newhouse. 2000. "How Does Managed Care Do It?" *The RAND Journal of Economics*, 31(3): pp. 526–548.

Cutler, David M., Mark McClellan, Joseph P. Newhouse, and Dahlia Remler. 1998. "Are Medical Prices Declining? Evidence from Heart Attack Treatments." *The Quarterly Journal of Economics*, 113(4): 991–1024.

Dafny, Leemore S. 2005. "How Do Hospitals Respond to Price Changes?" *The American Economic Review*, 95(5): 1525–1547.

Dranove, David, Chris Forman, Avi Goldfarb, and Shane Greenstein. 2014. "The Trillion Dollar Conundrum: Complementarities and Health Information Technology." *American Economic Journal: Economic Policy*, 6(4): 239–270, URL: http://dx.doi.org/10.1257/pol.6.4.239.

Epstein, Arnold M. 2002. "Volume and Outcome–It Is Time to Move Ahead." *The New England Journal of Medicine*, 346(15): 1161–4.

Finkelstein, Amy, Matthew Gentzkow, Peter Hull, and Heidi Williams. 2017. "Adjusting Risk Adjustment - Accounting for Variation in Diagnostic Intensity." *The New England journal of medicine*, 376(7): 608–610.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2016. "Sources of Geographic Variation in Health Care: Evidence from Patient Migration." *Quarterly Journal of Economics*, Forthcoming.

Fisher, Elliott S., David E. Wennberg, Thérèse A. Stukel, Daniel J. Gottlieb, F. L. Lucas, and Etoile L. Pinder. 2003. "The Implications of Regional Variations in Medicare Spending. Part 2: Health Outcomes and Satisfaction with Care." *Annals of Internal Medicine*, 138(4): 288–98.

Foster, Andrew D., and Mark R. Rosenzweig. 1995. "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture." *Journal of Political Economy*, 103(6): 1176–1209.

Fox, Jeremy T., and Valérie Smeets. 2011. "Does Input Quality Drive Measured Differences In Firm Productivity?" *International Economic Review*, 52(4): 961–989, URL: http://dx.doi.org/10.1111/j.1468-2354.2011.00656.x.

Gaure, Simen. 2014. "Correlation bias correction in two-way fixed-effects linear regression." *Stat*, 3(1): 379–390, URL: http://dx.doi.org/10.1002/sta4.68.

Geruso, Michael, and Timothy Layton. 2015. "Upcoding: Evidence from Medicare on Squishy Risk Adjustment." URL: http://dx.doi.org/10.3386/w21222.

Gibbons, Robert, and Rebecca Henderson. 2012. "What Do Managers Do? Exploring Persistent Performance Differences among Seemingly Similar Enterprises." In *The Handbook of Organizational Economics.* eds. by Robert Gibbons, and John Roberts, Chap. 17, Princeton: Princeton University Press, 680–731.

Goldman, L. Elizabeth, and R. Adams Dudley. 2008. "United States rural hospital quality in the Hospital Compare database-accounting for hospital characteristics." *Health Policy*, 87(1): 112–27.

GPO. 2007. "Changes to the Hospital Inpatient Prospective Payment Systems and Fiscal Year 2008 Rates." *Federal Register*, 72(162), August 27, 2007 Edition.

Griliches, Zvi. 1957. "Hybrid Corn: An Exploration in the Economics of Technological Change." *Econometrica*, 25(4): 501–522.

Jencks, Stephen F., Timothy Cuerdon, Dale R. Burwen, Barbara Fleming, Peter M. Houck, Annette E. Kussmaul, David S. Nilasena, Diana L. Ordin, and David R. Arday. 2000. "Quality of medical care delivered to Medicare beneficiaries: A profile at state and national levels." *The Journal of the American Medical Association*, 284(13): 1670–6.

Joynt, Karen E, E John Orav, and Ashish K Jha. 2014. "Association between hospital conversions to for-profit status and clinical and economic outcomes." *The Journal of the American Medical Association*, 312(16): 1644–52.

Kuhn, Thomson, Peter Basch, Michael Barr, Thomas Yackel, and Medical Informatics Committee of the American College of Physicians. 2015. "Clinical documentation in the 21st century: executive summary of a policy position paper from the American College of Physicians." *Annals of Internal Medicine*, 162(4): 301–3.

Leppert, Michelle A. 2012. "CMS releases 2013 IPPS Final Rule." URL: http://www.hcpro.com/HOM-283017-6962/CMS-releases-2013-IPPS-Final-Rule.html.

Li, Bingyang. 2014. "Cracking the Codes: Do Electronic Medical Records Facilitate Hospital Revenue Enhancement?" January, URL: http://www.kellogg.northwestern.edu/faculty/b-li/JMP.pdf.

Linden, Ariel, and Julia Adler-Milstein. 2008. "Medicare Disease Management in Policy Context." *Health Care Financing Review*, 29(3): 1–11.

McClellan, Mark B., and Douglas O. Staiger. 2000. "Comparing Hospital Quality at For-Profit and Not-for-Profit Hospitals." In *The Changing Hospital Industry: Comparing For-Profit and Not-for-Profit Institutions.* ed. by David M. Cutler: University of Chicago Press, 93–112.

McClellan, Mark, Barbara J. McNeil, and Joseph P. Newhouse. 1994. "Does More Intensive Treatment of Acute Myocardial Infarction in the Elderly Reduce Mortality? Analysis Using Instrumental Variables." *The Journal of the American Medical Association*, 272(11): 859–66.

McConnell, K John, Richard C Lindrooth, Douglas R Wholey, Thomas M Maddox, and Nick Bloom. 2013. "Management practices and the quality of care in cardiac units." *JAMA Internal Medicine*, 173(8): 684–92.

MEDPAC. 2012. "Medicare and the Health Care Delivery System." June 2012, URL: http://www.medpac.gov/documents/Jun12_EntireReport.pdf, Chapter: Serving Rural Medicare Beneficiaries.

MEDPAC. 2015. "Health Care Spending and the Medicare Program." June 2015, URL: http://www.medpac.gov/documents/data-book/june-2015-databook-health-care-spending-and-the-medicare-pr

Mihaly, Kata, Daniel F. McCaffrey, J. R. Lockwood, and Tim R. Sass. 2010. "Centering and reference groups for estimates of fixed effects: Modifications to felsdvreg." *Stata Journal*, 10(1): 82–103, URL: https://ideas.repec.org/a/tsj/stataj/v10y2010i1p82-103.html.

Mueller, Stephanie K, Stuart Lipsitz, and Leroi S Hicks. 2013. "Impact of hospital teaching intensity on quality of care and patient outcomes." *Medical Care*, 51(7): 567–74.

Neprash, Hannah T., Michael E. Chernew, Andrew L. Hicks, Teresa Gibson, and J. Michael McWilliams. 2015. "Association of Financial Integration Between Physicians and Hospitals With Commercial Health Care Prices." *JAMA Internal Medicine*, 175(12): , p. 1932, URL: http://dx.doi.org/10.1001/jamainternmed.2015.4610.

Nichols, Austin. 2008. "FESE: Stata module to calculate standard errors for fixed effects." Statistical Software Components, Boston College Department of Economics, February, URL: https://ideas.repec.org/c/boc/bocode/s456914.html.

O'Malley, Kimberly J., Karon F. Cook, Matt D. Price, Kimberly Raiford Wildes, John F. Hurdle, and Carol M. Ashton. 2005. "Measuring Diagnoses: ICD Code Accuracy." *Health Services Research*, 40(5 Pt 2): 1620–39.

Payne, Thomas. 2010. "Improving clinical documentation in an EMR world." *Healthcare Financial Management*, 64(2): 70–4.

Peterson, Eric D., Bimal R. Shah, Lori Parsons, Charles V. Pollack, Jr., William J. French, John G. Canto, C. Michael Gibson, and William J. Rogers. 2008. "Trends in Quality of Care for Patients With Acute Myocardial Infarction in the National Registry of Myocardial Infarction from 1990 to 2006." *American Heart Journal*, 156(6): 1045–55.

Pittet, Didier, Philippe Mourouga, Thomas V. Perneger, and the Members of the Infection Control Program. 1999. "Compliance with Handwashing in a Teaching Hospital." *Annals of Internal Medicine*, 130(2): 126–30.

Prophet, Sue. 2000. "How to Code Symptoms and Definitive Diagnoses." *Journal of AHIMA*, 71(6): 68–70.

Richter, Erin, Andrew Shelton, and Ying Yu. 2007. "Best practices for improving revenue capture through documentation." *Healthcare Financial Management*, 61(6): 44–7.

Rosenbaum, Benjamin P, Robert R Lorenz, Ralph B Luther, Lisa Knowles-Ward, Dianne L Kelly, and Robert J Weil. 2014. "Improving and measuring inpatient documentation of medical care within the MS-DRG system: education, monitoring, and normalized case mix index." *Perspectives in Health Information Management*, 11, p. 1c.

Scott, Kirstin W., E. John Orav, David M. Cutler, and Ashish K. Jha. 2016. "Changes in Hospital–Physician Affiliations in U.S. Hospitals and Their Effect on Quality of Care." *Annals of Internal Medicine*, 166(1): , p. 1, URL: http://dx.doi.org/10.7326/M16-0125.

Silverman, Elaine M., and Jonathan S. Skinner. 2004. "Medicare Upcoding and Hospital Ownership." *Journal of Health Economics*, 23(2): 369–89.

Skinner, Jonathan S., and Douglas O. Staiger. 2007. "Technology Adoption from Hybrid Corn to Beta-Blockers." In *Hard-to-Measure Goods and Services: Essays in Honor of Zvi Griliches.* eds. by Ernst R. Berndt, and Charles R. Hulten: University of Chicago Press, 545–570.

Skinner, Jonathan S, Douglas O Staiger, and Elliott S Fisher. 2006. "Is technological change in medicine always worth it? The case of acute myocardial infarction." *Health affairs (Project Hope)*, 25(2): w34–47.

Skinner, Jonathan, and Douglas Staiger. 2015. "Technology Diffusion and Productivity Growth in Health Care." *Review of Economics and Statistics*, 97(5): 951–964.

Sloan, Frank A., Justin G. Trogdon, Lesley H. Curtis, and Kevin A. Schulman. 2003. "Does the Ownership of the Admitting Hospital Make a Difference? Outcomes and Process of Care of Medicare Beneficiaries Admitted with Acute Myocardial Infarction." *Medical Care*, 41(10): 1193–1205.

Song, Yunjie, Jonathan Skinner, Julie Bynum, Jason Sutherland, John E. Wennberg, and Elliott S. Fisher. 2010. "Regional Variations in Diagnostic Practices." *The New England Journal of Medicine*, 363(1): 45–53.

Stafford, Randall S, and David C Radley. 2003. "The Underutilization of Cardiac Medications of Proven Benefit, 1990 to 2002." *Journal of the American College of Cardiology*, 41(1): 56–61.

Syverson, Chad. 2011. "What Determines Productivity?" *Journal of Economic Literature*, 49(2): 326–65.

Trude, Sally. 1992. "Physicians' Hospital Visits: A Description of Rates of Visits Under Medicare." A Rand Note N-3523-HCFA, RAND, URL: http://www.rand.org/pubs/notes/N3523.html.

Trude, Sally, Grace M. Carter, and Carolinda Douglass. 1993. "Measuring Physician Practice Patterns with Medicare Data." URL: https://archive.org/details/measuringphysici00trud.

Virnig, Beth. 2012. "Using Medicare Hospitalization Information and the MedPAR." URL: http://www.resdac.org/media/using-medicare-hospitalization-information-and-medpar.

Voss, Andreas, and Andreas F. Widmer. 1997. "No Time for Handwashing!? Handwashing versus Alcoholic Rub: Can We Afford 100% Compliance?" *Infection Control and Hospital Epidemiology*, 18(3): 205–208.

Williams, Scott C., Stephen P. Schmaltz, David J. Morton, Richard G. Koss, and Jerod M. Loeb. 2005. "Quality of care in U.S. hospitals as reflected by standardized measures, 2002-2004." *The New England Journal of Medicine*, 353(3): 255–64.

Youngstrom, Nina. 2013. "Discharge Summaries Take Center Stage; Risks Grow with Electronic Health Records." URL: http://racmonitor.com/news/27-rac-enews/1351-discharge-summaries-take-center-stage-risks-grow-with-electro

Figures

Adoption of Coding Practice Over Time
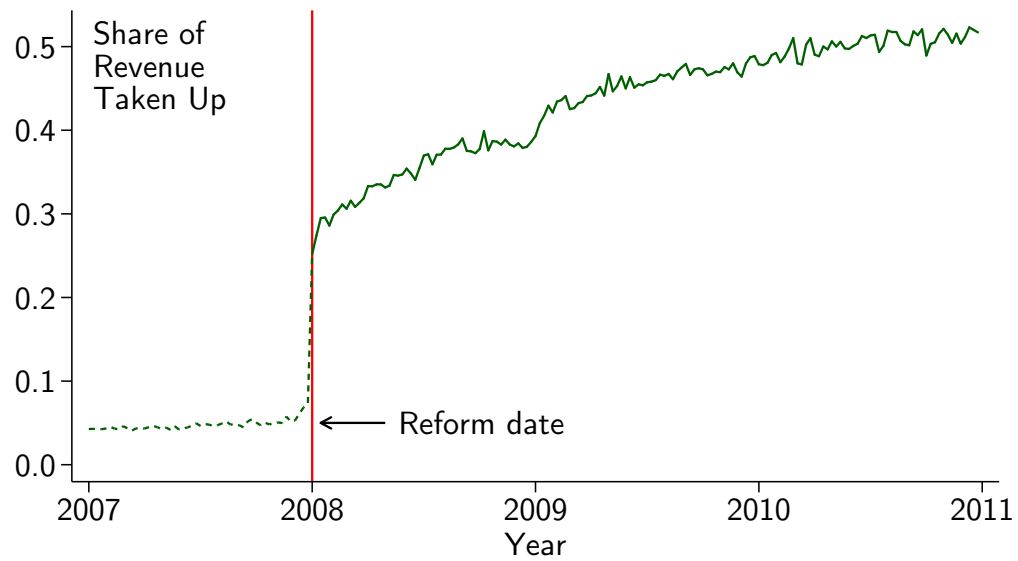
Figure plots the share of revenue available for detailed coding of HF that was captured by hospitals over time. Dotted line shows revenue that would have been captured in 2007 if hospitals had been paid per 2008 rules. See Appendix Section A.1.2 for more details.

Figure 1

Gain in Revenue by Type of Detailed HF Code
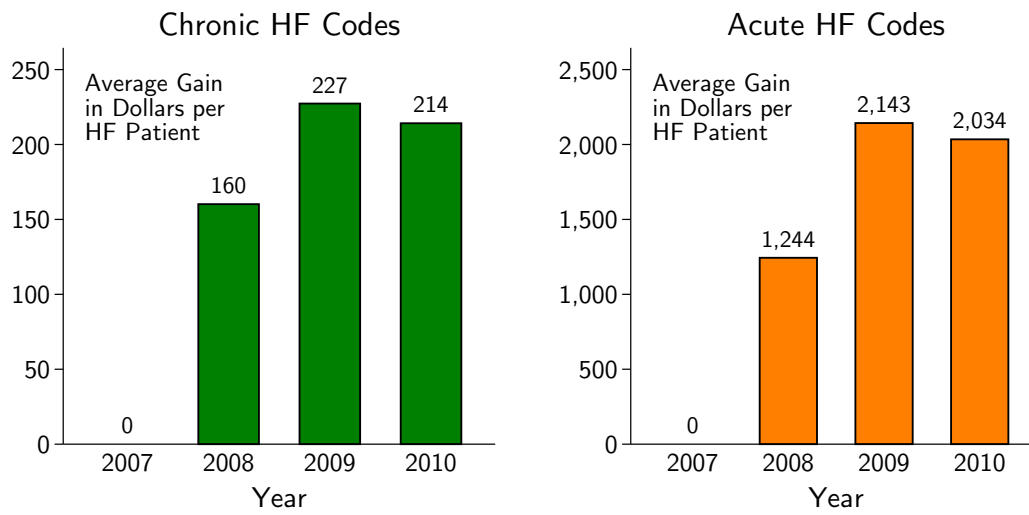
Figure plots the average per-HF patient gain in revenue going from always using vague codes for HF patients to always using chronic codes or acute codes. Prices in 2009 dollars.

Figure 2

Revenue at Stake per HF Patient across Hospitals
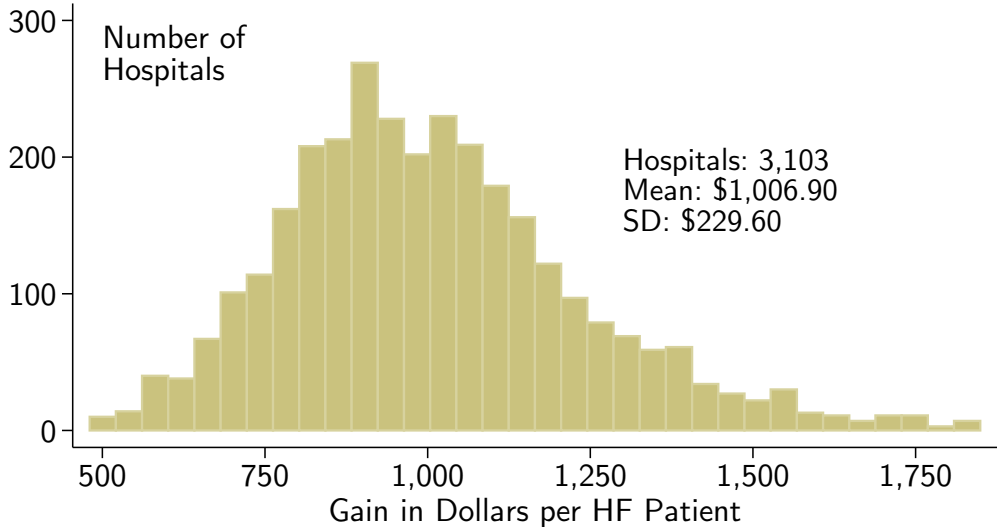
Revenue at stake is calculated using pre-reform (2007) patients processed under post-reform (2009) payment rules. The prediction process is described in the appendix. The 422 hospitals with <50 HF patients are suppressed and the upper and lower 1% in revenue at stake per HF patient are then removed.

Figure 3

Adoption of Coding Practice Across Hospitals Over Time

2007

Number of
Hospitals

Hospitals: 3,184
Mean: 0.125
SD: 0.101

2008

Hospitals: 3,143
Mean: 0.379
SD: 0.202

2009

Hospitals: 3,113
Mean: 0.470
SD: 0.222

2010

Hospitals: 3,081
Mean: 0.524
SD: 0.220

Share of HF Patients Receiving Detailed HF Code
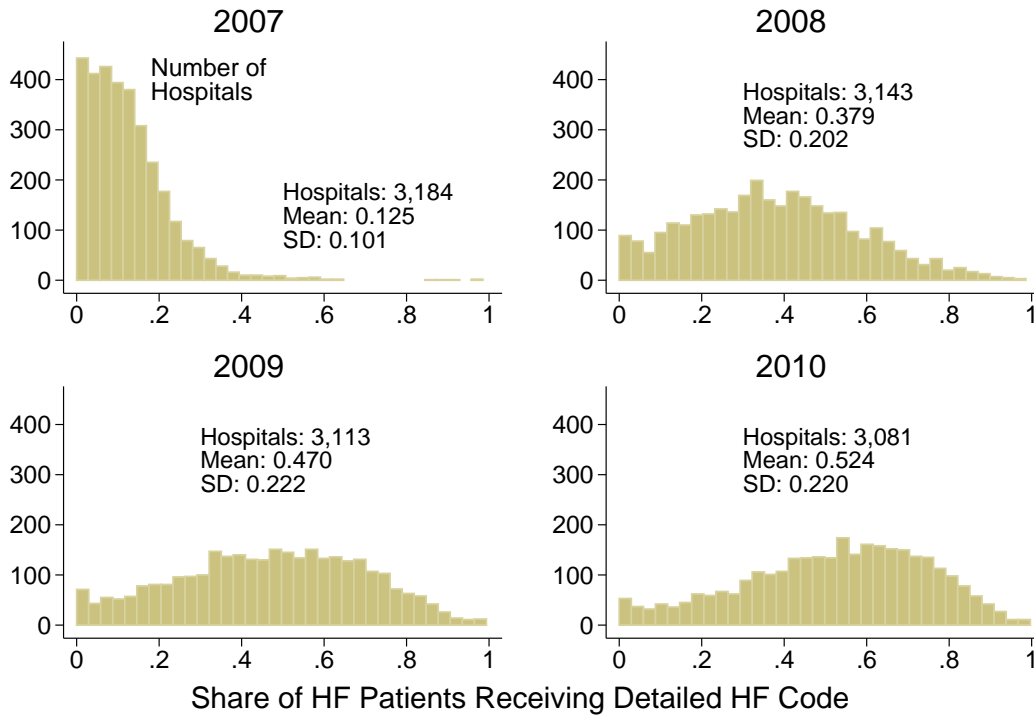
A hospital's adoption equals the share of its HF patients who received a detailed HF code in that year. Hospitals with fewer than 50 HF patients in the year excluded.

Figure 4

Tables

		Severity	
Code	Description	Before	After
Vague Codes			
428.0	Congestive HF, Unspecified	High	Low
428.9	HF, Other	High	Low
Specific Codes (Exhaustive Over Types of HF)			
428.20	HF, Systolic, Onset Unspecified	High	Medium
428.21	HF, Systolic, Acute	High	High
428.22	HF, Systolic, Chronic	High	Medium
428.23	HF, Systolic, Acute on Chronic	High	High
428.30	HF, Diastolic, Onset Unspecified	High	Medium
428.31	HF, Diastolic, Acute	High	High
428.32	HF, Diastolic, Chronic	High	Medium
428.33	HF, Diastolic, Acute on Chronic	High	High
428.40	HF, Combined, Onset Unspecified	High	Medium
428.41	HF, Combined, Acute	High	High
428.42	HF, Combined, Chronic	High	Medium
428.43	HF, Combined, Acute on Chronic	High	High

Table 1 - Vague and Specific HF Codes

Congestive HF (the description of code 428.0) is often used synonymously with HF.

Table 2 - Statistics about the First-Step Analysis Sample

	(1) Mean	(2) SD	(3) Min	(4) Max
Hospitals (N=2,831)				
HF Patients	533.73	504.59	1	3,980
Distinct Physicians	56.61	52.45	1	531
Mobile Physicians	19.02	21.02	1	169
Physicians (N=130,487)				
HF Patients	11.58	17.29	1	563
Distinct Hospitals	1.23	0.54	1	8
Mobile (>1 hospital)	0.184	0.388	0	1

The first-step analysis sample includes 1,510,988 HF patients. See text for more details.

Table 3 - Statistics about Physicians by Mobility Status

All values are means	(1) All	(2) Mobile	(3) Non-Mobile
Patient and Hospital Volume			
HF Patients	14.6	20.2	12.9
Share Given Specific Code	0.53	0.53	0.53
Distinct Hospitals	1.29	2.24	1
Mobile (>1 hospital)	0.24	1	0
Specialization			
Primary Care Physician	0.51	0.51	0.51
Medical Specialist	0.30	0.34	0.28
Surgeon	0.17	0.14	0.18
Unknown/Other	0.025	0.021	0.026
Demographics			
Female	0.19	0.15	0.20
Age	49.0	48.9	49.0
Training and Experience			
Years in Training*	5.94	6.51	5.76
Years Since Training*	15.9	15.4	16.0
Trained in US	0.69	0.59	0.72
Physicians	101,370	24,048	77,322

Mobile physicians are observed attending to HF patients at multiple hospitals in 2010; non-mobile physicians attend to patients at one hospital in that period. Data on specialization, demographics, training, and experience derived from AMA Masterfile. Excludes 29,117 "singleton" physicians who do not contribute to identification in the full econometric model and are omitted from the later physician analysis.

* Excludes physicians for whom years in/since training is unknown (3.5% in each column).

Table 4 - Hospital Summary Statistics

Patient Controls	(1) Mean	(2) SD
Heart Failure Coding and Physicians		
HF Patients	601.8	514.2
Share Given Specific Code	0.546	0.199
Distinct Physicians	62.97	54.16
Mobile Physicians	20.24	21.83
Hospital Characteristics		
Beds	287.9	235.0
Ownership		
Government	0.167	
Non-Profit	0.671	
For-Profit	0.161	
Location		
Rural Area	0.224	
Large Urban Area	0.422	
Other Urban Area	0.354	
Teaching Status		
Non-Teaching	0.623	
Major Teaching Hospital	0.101	
Minor Teaching Hospital	0.276	
Ex Ante $ at Stake / Patient	267.5	71.77
EMR and Hospital-Physician Integration		
EMR		
None	0.065	
Basic	0.502	
Advanced	0.434	
Hospital-Physician Integration		
None	0.305	
Contract	0.167	
Employment	0.351	
Unknown/Other	0.177	
Standards of Care (share of times standards used in 2006)		
for AMI Treatment	0.916	0.084
for Heart Failure Treatment	0.827	0.113
for Pneumonia Treatment	0.864	0.061
for High-Risk Surgeries	0.798	0.118
AMI Treatment (patients in 2000-2006)		
Adjusted 30-Day Survival	0.813	0.030

N=2,341 hospitals. See text for more details on the source and definitions of the characteristics. The standard deviations of specific coding for HF and AMI survival account for sampling variance.

Table 5 - Standard Deviation of Coding by Type of Hospital

Statistic	(1) Std Dev	(2) Std Dev	(3) Std Dev	(4) Std Dev	(5) N
All Hospitals	0.199	0.151	0.151	0.160	2,341
By Ownership					
Government	0.222	0.163	0.162	0.141	392
Non-Profit	0.191	0.147	0.147	0.167	1,571
For-Profit	0.192	0.143	0.143	0.143	378
By Location					
Rural	0.229	0.171	0.170	0.190	525
Large Urban	0.192	0.146	0.145	0.146	988
Other Urban	0.182	0.142	0.141	0.151	828
By Size					
Upper Tercile	0.174	0.137	0.137	0.129	780
Middle Tercile	0.184	0.141	0.141	0.143	775
Lower Tercile	0.227	0.168	0.167	0.196	786
By Teaching Status					
Non-Teaching	0.206	0.154	0.153	0.159	1,459
Major Teaching	0.183	0.146	0.146	0.129	237
Minor Teaching	0.182	0.141	0.141	0.168	645
By EMR Type					
None	0.184	0.143	0.143	0.135	151
Basic	0.207	0.155	0.154	0.151	1,175
Advanced	0.186	0.143	0.143	0.171	1,015
By Hospital-Physician Integration					
None	0.201	0.151	0.150	0.180	714
Contract	0.188	0.145	0.145	0.129	392
Employment	0.191	0.147	0.147	0.164	821
Patient Controls	None	Admission	Full	Full	
Physician Controls	None	None	None	FE	

Each row shows the standard deviation in coding score for a different partition of hospitals (hospital counts, which apply to columns 1-4, shown in column 5). Column 1 uses no controls to calculate the hospital effects. Column 2 adds controls for patient characteristics observable upon admission, and column 3 adds histories of chronic conditions. Column 4 adds physician fixed effects. All results are adjusted for sampling variation.

Table 6 - Association Between Hospital Characteristics and Coding

Outcome	(1) Score	(2) Score	(3) Score	(4) Score
Hospital Characteristics (C_h)				
ln(Beds)	0.018**	0.011*	0.011*	-0.009
	(0.008)	(0.006)	(0.006)	(0.011)
Non-Profit Ownership	0.037**	0.028**	0.028**	0.020
	(0.015)	(0.011)	(0.011)	(0.013)
For-Profit Ownership	0.018	0.016	0.015	0.031*
	(0.017)	(0.012)	(0.012)	(0.017)
Located in Large Urban Area	0.002	-0.001	-0.002	0.030
	(0.016)	(0.012)	(0.012)	(0.020)
Located in Other Urban Area	0.016	0.008	0.008	0.031*
	(0.014)	(0.011)	(0.011)	(0.016)
Major Teaching Hospital	0.024	0.028**	0.028**	0.045**
	(0.018)	(0.014)	(0.014)	(0.020)
Minor Teaching Hospital	-0.003	0.002	0.002	0.003
	(0.010)	(0.008)	(0.008)	(0.014)
Ex Ante \$ at Stake per Patient	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
EMR and Hospital-Physician Integration (I_h)				
Basic EMR	-0.038*	-0.034**	-0.034**	-0.012
	(0.020)	(0.016)	(0.016)	(0.022)
Advanced EMR	-0.010	-0.010	-0.009	0.003
	(0.020)	(0.016)	(0.016)	(0.023)
Contract Affiliation	0.001	0.004	0.004	0.021
	(0.013)	(0.010)	(0.010)	(0.014)
Employment Affiliation	0.024**	0.018**	0.018**	0.024*
	(0.012)	(0.009)	(0.009)	(0.013)
Unknown/Other Affiliation	-0.018	-0.011	-0.011	0.015
	(0.014)	(0.011)	(0.011)	(0.015)
Standards of Care and Clinical Performance (Z_h)				
Standards of Care Z-Score	0.026***	0.019***	0.019***	0.017***
	(0.006)	(0.004)	(0.004)	(0.006)
AMI Survival Z-Score	0.030***	0.025***	0.025***	0.030***
	(0.007)	(0.005)	(0.005)	(0.008)
Observations	2,341	2,341	2,341	2,341
R^2	0.107	0.109	0.107	0.032
Basic EMR = Advanced EMR (p-val)	0.002	0.000	0.000	0.178
Patient Controls	None	Admission	Full	Full
Physician Controls	None	None	None	FE

This table presents the results of regressing hospital coding scores on hospital characteristics. Column 1 uses no controls to calculate the hospital scores, column 2 adds controls for patient characteristics observable upon admission, and column 3 adds histories of chronic conditions. Column 4 adds physician fixed effects. Standard errors clustered at the market level in parentheses.

*** significant at 1% level; ** significant at 5% level; * significant at 10% level

Table 7 - Association Between Hospital Characteristics and Coding Over Time				
	(1)	(2)	(3)	(4)
Time Horizon of Patient Sample	2008Q1	2008	2010	2010Q4
Hospital Characteristics (C_h)				
ln(Beds)	0.008*	0.003	0.011*	0.012**
	(0.005)	(0.005)	(0.006)	(0.006)
Non-Profit Ownership	0.009	0.012	0.028**	0.025**
	(0.009)	(0.010)	(0.011)	(0.011)
For-Profit Ownership	-0.009	-0.012	0.015	0.018
	(0.011)	(0.012)	(0.012)	(0.013)
Located in Large Urban Area	-0.015	-0.010	-0.002	-0.005
	(0.010)	(0.010)	(0.012)	(0.013)
Located in Other Urban Area	0.002	0.005	0.008	0.005
	(0.010)	(0.011)	(0.011)	(0.011)
Major Teaching Hospital	0.011	0.017	0.028**	0.024
	(0.011)	(0.011)	(0.014)	(0.015)
Minor Teaching Hospital	0.003	0.005	0.002	0.002
	(0.007)	(0.007)	(0.008)	(0.008)
Ex Ante $ at Stake per Patient	-0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
EMR and Hospital-Physician Integration (I_h)				
Basic EMR	-0.001	-0.004	-0.034**	-0.035**
	(0.011)	(0.010)	(0.016)	(0.017)
Advanced EMR	0.013	0.012	-0.009	-0.013
	(0.011)	(0.010)	(0.016)	(0.016)
Contract Affiliation	0.000	0.001	0.004	0.002
	(0.009)	(0.008)	(0.010)	(0.010)
Employment Affiliation	0.007	0.007	0.018**	0.014*
	(0.008)	(0.008)	(0.009)	(0.008)
Unknown/Other Affiliation	-0.011	-0.004	-0.011	-0.015
	(0.009)	(0.009)	(0.011)	(0.011)
Standards of Care and Clinical Performance (Z_h)				
Standards of Care Z-Score	0.017***	0.018***	0.018***	0.019***
	(0.004)	(0.004)	(0.004)	(0.004)
AMI Survival Z-Score	0.013**	0.017***	0.025***	0.026***
	(0.005)	(0.005)	(0.005)	(0.005)
Observations	2,371	2,372	2,341	2,338
R^2	0.051	0.063	0.107	0.094
Basic EMR = Advanced EMR (p-val)	0.029	0.013	0.000	0.001
Patient Controls	Full	Full	Full	Full
Physician Controls	None	None	None	None

This table presents the results of regressing hospital coding scores on hospital characteristics. Each column estimates coding scores from a different time period: 2008Q1, 2008 (full year), 2010 (full year), and 2010Q4. Scores are estimated with the full set of patient controls but without physician fixed effects (replicating the specification of Table 5, column 3). Standard errors clustered at the market level in parentheses. *** significant at 1% level; ** significant at 5% level; * significant at 10% level

Table 8 - Physician Coding Dispersion and Correlates

	(1)	(2)	(3)	(4)
A. Standard Deviation of Coding				
Across Physicians	0.224	0.181	0.180	0.155
B. Regression of Physician Coding Score on Physician Characteristics				
Volume and Mobility Status				
ln(HF Patients)	0.046***	0.033***	0.032***	0.028***
	(0.002)	(0.002)	(0.002)	(0.002)
Mobile Physician	-0.025***	-0.023***	-0.023***	-0.013**
	(0.005)	(0.004)	(0.004)	(0.005)
Specialization				
Medical Specialist	0.046***	0.023***	0.022***	0.003
	(0.005)	(0.004)	(0.004)	(0.004)
Surgeon	-0.103***	-0.053***	-0.051***	-0.072***
	(0.005)	(0.005)	(0.005)	(0.005)
Unknown/Other	-0.013**	-0.011*	-0.011*	-0.020***
	(0.006)	(0.006)	(0.006)	(0.006)
Demographics				
Female	0.022***	0.018***	0.018***	0.011***
	(0.003)	(0.002)	(0.002)	(0.003)
Age	-0.002***	-0.002***	-0.002***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Training and Experience				
Years in Training	0.001***	0.001***	0.001***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Years in Training Unknown	-0.018**	-0.014**	-0.014**	-0.014**
	(0.009)	(0.007)	(0.007)	(0.007)
Trained in US	0.030***	0.025***	0.025***	0.015***
	(0.006)	(0.004)	(0.004)	(0.004)
Observations	101,370	101,370	101,370	101,370
R^2	0.074	0.042	0.041	0.031
Patient Controls	None	Admission	Full	Full
Hospital Controls	None	None	None	FE

This table first presents the standard deviation in coding scores across physicians (adjusted for sampling variation). Next, it presents the results of regressing physician coding scores on physician characteristics. Column 1 uses no controls to calculate the physician scores, column 2 adds controls for patient characteristics observable upon admission, and column 3 adds histories of chronic conditions. Column 4 adds hospital fixed effects. All models exclude 29,117 "singleton" physicians who do not contribute to identification of the hospital effects in the full econometric model of column 4. Standard errors clustered at the market level in parentheses.

*** significant at 1% level; ** significant at 5% level; * significant at 10% level

Appendix To:

Adoption and Learning Across Hospitals: The Case of a Revenue-Generating Practice

Adam Sacarny

April 2018

A Appendix

A.1 Revenue at Stake

To determine how a hospital would have been paid had it coded HF differently, I use a computer program called a grouper that translates an inpatient claim into its Medicare payment diagnosis-related group (DRG). I use the DRGGroupers.net Perl grouper software. For each patient i with a HF diagnosis, I process her claim as-is, then reprocess it replacing her secondary HF codes with a low-severity/non-CC code (428.0 – congestive HF, unspecified), medium-severity/CC code (428.22 – HF, systolic, chronic), and high-severity/MCC code (428.21 – HF, systolic, acute) using the Medicare DRG rules in year t^*. The result is a set of DRG weights $\left(w_i^{asis,t^*}, w_i^{noncc,t^*}, w_i^{cc,t^*}, w_i^{mcc,t^*}\right)$ – a measure of the expected cost of treatment for patients in each DRG that is uniform across hospitals. These weights are then used in the calculations for revenue at stake in a given year from higher intensity HF coding and to produce the *ex ante* revenue at stake for hospitals, described in the following sections.

A.1.1 Contemporaneous

To calculate the revenue at stake from HF coding in a given year t (as shown in Figure 2), I start with the set of all patients with HF in the grand sample in year t, P_t. I let C_t be the average conversion factor from DRG weights to dollars in year t (calculated by taking, for all patients in the MEDPAR file with FFS Medicare Part A & B coverage in year t, the average ratio of the "drgprice" variable to the DRG weight). The economy-wide potential gain per patient from chronic HF codes (expressed in constant 2009 dollars) is calculated as:

$$gainpp_t^{cc} = C_{2009} \times \frac{\sum_{i \in P_t} \left(w_i^{cc,t} - w_i^{noncc,t}\right)}{|P_t|}$$

The potential gain from acute HF codes is calculated as:

$$gainpp_t^{mcc} = C_{2009} \times \frac{\sum_{i \in P_t} \left(w_i^{mcc,t} - w_i^{noncc,t}\right)}{|P_t|}$$

These gains are visualized in Figure 2 for years $t = 2007 \ldots 2010$.

A.1.2 Predictor of Revenue at Stake

The revenue at stake from the reform for a particular patient depends on whether she was diagnosed with chronic or acute HF. I therefore construct a predictor of the acuity of the patient's HF. This predictor uses HF patients at hospitals that were relatively detailed coders in 2010 – hospitals that gave at least 85% of their HF patients a detailed code. The sample includes 90,653 patients and 171 hospitals. I regress whether the patient was coded as having high-severity HF on well-measured patient attributes: indicators for age, race, sex, month of admission, admission through the emergency department, 19 chronic conditions, and the 25 major diagnostic categories classifying the underlying cause of admission (the chronic conditions are listed in Appendix Section A.4).

I use the coefficients from this regression to fit the probability that a patient would have received a high-severity HF code under full adoption of the coding practice, $\hat{p}_i^{mcc}$, constraining the fitted value to be between 0 and 1. For patients who were already coded as getting a high severity code, I set $\hat{p}_i^{mcc} = 1$; patients already coded with a medium-severity code get $\hat{p}_i^{mcc} = 0$. Patients who do not receive a high-severity code are assumed to receive a medium-severity code i.e. $\hat{p}_i^{cc} = 1 - \hat{p}_i^{mcc}$. I then re-price these patients under the pricing rules of year t^*. Their expected DRG weight under full coding according to the payment rules of year t^* is defined as:

$$\hat{w}_i^{t^*} = \hat{p}_i^{mcc} w_i^{mcc,t^*} + \hat{p}_i^{cc} w_i^{cc,t^*}$$

The expected gain to using the detailed codes under the payment rules of year t^* equals the expected DRG weight under full coding less the DRG weight with no detailed codes:

$$\hat{gain}_i^{t^*} = \hat{w}_i^{t^*} - w_i^{nocc,t^*}$$

The *ex ante* per-patient gain from full HF coding for hospital h, depicted in Figure 3 and used in the analysis regressions, equals the rise per HF patient in DRG payments when the hospital's 2007 patients are processed under 2009 rules (expressed in 2009 dollars for consistency with the rest of the paper). Let $P_{h,t}$ be the HF patients at hospital h in year t for whom their chronic conditions are observed:

$$exantepp_h = C_{2009} \frac{\sum_{i \in P_{h,2007}} \hat{gain}_i^{2009}}{|P_{h,2007}|}$$

The depiction in Figure A3 follows the same formula but divides by the total number of patients, not just those with HF. To improve precision and reduce the leverage of outliers, when this predictor is used in the main regressions and displayed in the figures, hospitals with fewer than 50 HF patients in 2007 as well as those with an outlying top or bottom 1% of revenue on the table per patient were culled from this measure.

Figure 1 displays hospitals' capture of the HF revenue over time. The plot is at the weekly level and shows the fraction of revenue at stake that was captured according to the contemporaneous payment rules. It uses the aforementioned prediction algorithm to impute the probability that each patient has medium or high severity HF. Let weeks be indexed by k and let $t(k)$ be the year of week k; let P_k be all patients with HF in week k with chronic conditions observed. Since the figure also plots the revenue that would have been captured in 2007 if 2008 payment rules were in effect, let $\tilde{t}(k) = \max(t(k), 2008)$. Define the realized gain from specific coding for the patient according to rules of year t^* as:

$$gain_i^{t^*} = w_i^{asis,t^*} - w_i^{nocc,t^*}$$

Then each point in the figure is defined as:

$$capture_k = \frac{\sum_{i \in P_k} gain_i^{\tilde{t}(k)}}{\sum_{i \in P_k} \hat{gain}_i^{\tilde{t}(k)}}$$

A.2 Quality and Performance Measures

A.2.1 Standards of Care

I construct a composite measure of hospital utilization of standards of care by adding together standardized measures of AMI, HF, pneumonia, and surgery standards of care in 2006.

The AMI measure includes 8 processes (aspirin at arrival, aspirin at discharge, ACE inhibitors, smoking cessation advice, β-blockers at discharge, β-blockers at arrival, thrombolytics at arrival, and PCI at arrival). The heart failure measure includes 4 processes (discharge instructions, evaluation of left ventricular systolic function, ACE inhibitors, and smoking cessation advice). The pneumonia measure includes 7 processes (oxygenation assessment, pneumococcal vaccine, blood culture before antibiotics, smoking cessation advice, timely antibiotics, appropriate antibiotics, and influenza vaccine), and the surgery measure includes 3 measures (preventative antibiotics, appropriate antibiotics, and antibiotics stopped quickly).

For each of the 4 groups of scores, I calculate an overall score by summing together the numerators from all the component measures and dividing it by the sum of the denominators. I standardize this measure, then add together the four standardized measures and standardize the result, yielding one composite Z-score of process of care use.

A.2.2 Adjusted AMI Survival

I construct adjusted AMI survival by starting with a sample of all AMI episodes in FFS Medicare in fiscal years 2000-2006. This sample is generated as described in Chandra et al. (2013) and is a subset of the analysis sample used in that paper. I restrict the analysis to hospitals that treated at least 25 AMI patients during that time frame. I then regress an indicator for a patient's 30-day survival on age-race-sex interactions, logged inputs (real resources), 25 risk-adjusters, and hospital fixed effects.[15] The hospital fixed effects are then extracted and their standard errors estimated under a homoscedasticity assumption; they are then Empirical Bayes adjusted to account for measurement error when they are used in the analysis regressions (note that here I use constant weights across hospitals so that all facilities receive the same weight, whereas the procedure in the previous paper

---

[15]I use 8 additional risk-adjusters beyond those of Chandra et al. (2013) but constructed in the same way (i.e. on the basis of prior hospitalizations): heart failure, myocardial infarction, unstable angina, chronic atherosclerosis, respiratory failure, hypertensive heart disease, valvular heart disease, and arrhythmia.

uses optimal weights).

One difference between this study and Chandra et al. (2013) is that the latter used log-survival days censored at 1 year as its outcome measure, whereas I use an indicator for 30-day survival. In practice, these measures yield similar results in the main regressions when they are standardized because the two measures have a correlation coefficient of 0.916.

A.3   Coding at the Hospital System and Geographic Region Levels

Variations in coding can also be studied at the level of the hospital system and geographic region. The main text focuses on hospitals instead of systems and regions because the key economic questions concern provider behavior, and these alternative levels aggregate over providers. However, given the large literature on health care variations (which includes studies of coding, c.f. Song et al., 2010 and Finkelstein et al., 2017), and the potential role for hospital systems to drive diffusion, statistics at these levels may be of interest. I match hospitals to hospital systems using American Hospital Association survey data, to Dartmouth Hospital Referral Regions (HRRs) using ZIP code, and to Metropolitan Statistical Areas (MSAs) using the CMS Provider of Services file.

First, to get a sense of the explanatory power of these levels, I regress hospital coding scores on system/region fixed effects and report the R^2 adjusted for degrees of freedom and regressand (coding score) measurement error. The adjustment process is described at the end of this section.

The upper section of Table A4 presents the results. Each row estimates the hospital score with different first-step controls. Of the 2,341 hospitals in the analysis sample, 1,521 were in a hospital system, and there were 321 distinct systems. All hospitals had an HRR (since these regions partition the entire U.S.) and 1,705 were in an MSA. Both hospital system and geographic fixed effects left the majority of hospital coding variations unexplained, though both levels had nontrivial explanatory power: systems and geographies explained one-sixth to one-fourth of variation in coding across hospitals when the physician component was not removed in the first step and one-tenth to one-fifth of variation when the physician component was removed.

The lower section of Table A4 estimates the dispersion in system and region effects directly. I estimate equation 1 replacing the hospital and physician fixed effects with (respectively) system, HRR, and MSA fixed effects. As shown previously for hospitals and physicians, variations attenuate with the addition of patient characteristics observable upon admission, but do not further attenuate with additional controls for chronic conditions. There are meaningful variations across systems and regions, though the magnitudes are smaller than variations at the hospital and physician level. Accounting for patient characteristics, the standard deviation in coding across hospital systems is 12.1 percentage points. The standard deviation is 8.3 percentage points across HRRs and 9.9 percentage points across MSAs.

Then, in Table A5, I show the relationship between HRR coding scores and Medicare spending per enrollee from the 2010 Dartmouth Atlas by regressing the former on the latter. Spending is adjusted for demographics and prices (though not coding intensity – the price adjustment removes geographic factors, using, for example, nationally standardized prices for DRGs). Perhaps surprisingly, columns 1-3 of the table show that higher-spending regions are less likely to code heart failure in a detailed fashion; columns 4-6 show no association between inpatient spending (hospital and SNF) and coding.

However, breaking down spending into its components and looking at conditional associations, columns 7-9 find that inpatient spending (hospital and SNF) is positively associated with coding, while physician spending is negatively associated with it. That is, holding fixed the other components of spending, areas with higher hospital payments per enrollee tended to code more, while areas with higher physician payments per enrollee tended to code less.

A.3.1 Adjustment of R^2

To compute the R^2, I initially regress the hospital coding score on a set of level fixed effects:

$$\hat{\alpha}_h = \alpha_{l(h)} + \xi_h + \epsilon_h$$

Where h indexes hospitals and l indexes the level (system or region). $\hat{\alpha}_h$ is the estimate of the hospital coding score and α_l is the level effect. The regression error term has two components. ξ_h is measurement error in the regressand ($\hat{\alpha}_h = \alpha_h + \xi_h$). ϵ_h is the remaining error – the true regression error term that would have occurred with a precisely measured regressand. Because the overall error term includes measurement error, R^2 will understate the fraction of variation in underlying hospital coding that is explained by the fixed effects:

$$R^2 = 1 - \frac{\text{Var}\left(\xi_h + \epsilon_h\right)}{\text{Var}\left(\hat{\alpha}_h\right)} = 1 - \frac{\text{Var}\left(\xi_h\right) + \text{Var}\left(\epsilon_h\right)}{\text{Var}\left(\alpha_h\right) + \text{Var}\left(\xi_h\right)} < 1 - \frac{\text{Var}\left(\epsilon_h\right)}{\text{Var}\left(\alpha_h\right)}$$

I adjust the R^2 for regressand measurement error by estimating $\text{Var}\left(\xi_h\right)$ as the average squared standard error of the first-step hospital fixed effects. I estimate $\text{Var}\left(\xi_h + \epsilon_h\right)$ as the mean squared error of the regression and $\text{Var}\left(\hat{\alpha}_h\right)$ as the sample variance of the hospital coding scores. These estimates use a degrees of freedom adjustment and so already will yield a degrees of freedom adjusted

R^2 (i.e. what is typically called an adjusted R^2). Plugging in the estimates I arrive at my R^2 adjusted for measurement error and degrees of freedom:

$$\hat{R}^2_* = 1 - \frac{\hat{\text{Var}}\left(\xi_h + \epsilon_h\right) - \hat{\text{Var}}\left(\xi_h\right)}{\hat{\text{Var}}\left(\hat{\alpha}_h\right) - \hat{\text{Var}}\left(\xi_h\right)}$$

Note that the denominator is the formula used to estimate the underlying variance of the hospital effects.

A.4 First-Step Specifications

I use four specifications to estimate the hospital fixed effects using the analysis sample described in Section 4.1 of the main text. The regressions are of the form of equation 1. The first three of these specifications use no physician fixed effects while the final specification adds physician fixed effects. The regressions are run in stata with the commands `felsdvregdm` (Mihaly et al., 2010) and `reghdfe` (Correia, 2016). Standard errors are derived from `felsdvregdm` and a modified version of the command `fese` (Nichols, 2008).

The first specification uses no controls at all. The second specification uses only controls that were observable from the patient's admission and not historical data. These controls are: age-race-sex interactions (age in 5 year categories starting at 65 and with age 90+ treated as one category, race as white/nonwhite, sex as female/not female), month of year indicators, an indicator for being admitted through the emergency department, and indicators for 179 categories of the primary diagnosis code. The 179 categories are constructed from the HCUP Clinical Classifications Software ICD-9 diagnosis code multi-level categories. The aim is to include an indicator for each commonly used category of codes and roll up uncommon categories that are clinically similar. Starting with the most finely grained level (level 4), categories comprising at least 0.1% of the population were included as indicators. Categories comprising less than 0.1% were replaced with their level 3 codes. Then, looking at the level 3 codes, those comprising at least 0.1% were included as indicators, and the rest were replaced with their level 2 codes, and so on for levels 2 and 1.

The third specification adds to these controls additional indicators for the patient's history of chronic conditions. These indicators are based on the Medicare Chronic Conditions segment. This file reports whether patients had received a diagnosis for the conditions in Medicare claims during a reference period ranging from 1-3 years. An indicator is provided for each condition at midyear and at the end of the year. To identify preexisting conditions only, I use the most recent report of chronic conditions that occurred before the patient's admission to the hospital. I include indicators for 19 chronic conditions: acute myocardial infarction, atrial fibrillation, cataract, chronic kidney disease, COPD, HF, diabetes, glaucoma, hip fracture, ischemic heart disease, depression, osteoporosis, rheumatoid arthritis or osteoarthritis, stroke or transient ischemic attack, breast cancer, colorectal cancer, prostate cancer, lung cancer, and endometrial cancer.

The fourth specification amends the third specification to add physician fixed effects.

A.5 Procedure for Bias-Corrected Dispersion of High Dimensional Fixed Effects

Here I describe a feasible approach to estimating the underlying variance of hospital and physician fixed effects. The original formulas for this approach were laid out in Andrews et al. (2008), but involved the construction and inversion of a large matrix. Gaure (2014) then developed a version that bypassed the direct creation of this matrix. My procedure is nearly identical to Gaure (2014), except that where he estimates the observation-level (i.e. patient-level) underlying variance of the effects, the procedure here estimates the hospital-level and physician-level variance of the two sets of effects, respectively.

I assume outcomes are an additively separable function of patient-level covariates and hospital and physician effects:

$$Y = X\beta + D\theta + F\psi + \epsilon \tag{A1}$$

Where Y is an $N \times 1$ vector of outcomes, X is an $N \times K_\beta$ matrix of patient-level covariates, D is an $N \times K_\theta$ matrix of physician effects, F is an $N \times K_\psi$ matrix of hospital effects, and ϵ is an $N \times 1$ vector of disturbances. β is a vector of coefficients on the patient level covariates, θ are the physician fixed effects, and ψ are the hospital fixed effects.

I assume uncorrelated and homoscedastic disturbances:

$$\mathbb{E}\left[\epsilon\epsilon'\right] = \sigma_\epsilon^2 \times I_N$$

Let $M_{X,D}$ be the orthogonal projection matrix with respect to X and D. Premultiplying A1 by $M_{X,D}$ yields:

$$M_{X,D}Y = M_{X,D}F\psi + M_{X,D}\epsilon$$

Then standard OLS regression yields estimates $\hat{\psi}$ for ψ:

$$\hat{\psi} = \left(F'M_{X,D}F\right)^{-1} F'M_{X,D}Y$$

Plugging in A1 to replace Y with the right-hand side, I express $\hat{\psi}$ as a sum of a true component and a sampling noise component:

$$\hat{\psi} = \psi + \left(F'M_{X,D}F\right)^{-1}F'M_{X,D}\epsilon$$

Now define the true and estimated variance of the hospital effects:

$$\sigma_\psi^2 = \frac{\psi'M_1\psi}{K_\psi}, \ \tilde{\sigma}_\psi^2 = \frac{\hat{\psi}'M_1\hat{\psi}}{K_\psi}$$

Where M_1 is the demeaning matrix i.e. $I - 1\left(1'1\right)^{-1}1$. The expected value of $\tilde{\sigma}_\psi^2$ has a true component and bias term due to sampling variance:

$$\mathbb{E}\left[\tilde{\sigma}_\psi^2\right] = \sigma_\psi^2 + \delta_\psi$$

$$\delta_\psi = \mathbb{E}\left[\frac{\epsilon'M_{X,D}F\left(F'M_{X,D}F\right)^{-1}M_1\left(F'M_{X,D}F\right)^{-1}F'M_{X,D}\epsilon}{K_\psi}\right]$$

Relying on the exchange of the expectation and trace operators as well as the invariance of the trace to cyclic permutations, the bias term simplifies to:

$$\delta_\psi = \mathrm{tr}\left(\mathbb{E}\left[\epsilon\epsilon'\right]\frac{M_{X,D}F\left(F'M_{X,D}F\right)^{-1}M_1\left(F'M_{X,D}F\right)^{-1}F'M_{X,D}}{K_\psi}\right)$$

$$= \frac{\sigma_\epsilon^2}{K_\psi} \times \mathrm{tr}\left(M_1\left(F'M_{X,D}F\right)^{-1}\right) \tag{A2}$$

The trace term requires the inversion of a potentially large matrix. To avoid directly computing this matrix, I utilize an approach described in Gaure (2014). I note the following equality for $\mathbb{E}\left[x\right] = 0$ and $\mathrm{Var}\left(x\right) = 1$:

$$\mathrm{tr}\left(A\right) = \mathbb{E}\left[x'Ax\right] \tag{A3}$$

So I can compute the bias term with a stochastic approximation, i.e.:

$$\mathrm{tr}\left(M_1\left(F'M_{X,D}F\right)^{-1}\right) = \mathbb{E}\left[x'M_1\left(F'M_{X,D}F\right)^{-1}x\right]$$

A13

Per Gaure (2014), the variance-minimizing $x'Ax$ for symmetric A is from an x that is m independent draws from a discrete uniform distribution over the sign function i.e. $\{-1, 1\}$. I define x in this way. Now define v:

$$v = \left(F'M_{X,D}F\right)^{-1} x$$

Computing v from x would require inverting a large matrix, but I can also write v as the vector that solves the following linear system:

$$F'M_{X,D}Fv = x \tag{A4}$$

Equation A4 can be solved using a linear solver, though this would require computing $F'M_{X,D}F$, which is memory intensive as N and K_ψ grow. Thus for large N and K_ψ I mimic Gaure, 2014 in using a conjugate gradient algorithm. The advantage of this method is that $F'M_{X,D}F$ need never be directly calculated. The vector Fv merely needs to be residualized with respect to X and D, which is readily done by regression fixed effects methods.

Then plugging in x and v:

$$\delta_\psi = \frac{\sigma_\epsilon^2}{K_\psi} \times \mathbb{E}\left[x'M_1 v\right]$$

Let $\hat{\delta}_\psi$ be a feasible estimate of this bias term. One such feasible estimate is to replace σ_ϵ^2 in equation A2 with $\hat{\sigma}_\epsilon^2$:

$$\hat{\delta}_\psi = \frac{\hat{\sigma}_\epsilon^2}{K_\psi} \times \mathrm{tr}\left(M_1 \left(F'M_{X,D}F\right)^{-1}\right) \tag{A5}$$

However, since computing the trace term is computationally intensive, I define the stochastic approximation to the feasible estimate:

$$\hat{\delta}_\psi^s = \frac{\hat{\sigma}_\epsilon^2}{K_\psi} \times \hat{\mathbb{E}}\left[x'M_1 v\right] \tag{A6}$$

Then the feasible estimate of underlying variance is the variance of the estimated fixed effects less the bias estimate:

$$\hat{\sigma}_\psi^2 = \tilde{\sigma}_\psi^2 - \hat{\delta}_\psi^s$$

This estimate can be computed up to arbitrary levels of precision by repeatedly drawing new x vectors, re-running the algorithm to estimate $\hat{\delta}_\psi^s$, and taking the average of the results.

By identical argument, a feasible estimate of the underlying variance of the physician effects can be constructed as:

$$\hat{\sigma}_\theta^2 = \tilde{\sigma}_\theta^2 - \hat{\delta}_\theta$$

$$\tilde{\sigma}_\theta^2 = \frac{\hat{\theta}' M_1 \hat{\theta}}{K_\theta}$$

$$\hat{\delta}_\theta = \frac{\hat{\sigma}_\epsilon^2}{K_\theta} \times \text{tr}\left(M_1 \left(D' M_{X,F} D \right)^{-1} \right)$$

$$\hat{\delta}_\theta^s = \frac{\hat{\sigma}_\epsilon^2}{K_\theta} \times \hat{\mathbb{E}}\left[w' M_1 y \right]$$

Where w is an iid vector drawn from the sign distribution like x, and y solves $D' M_{X,F} D y = w$.
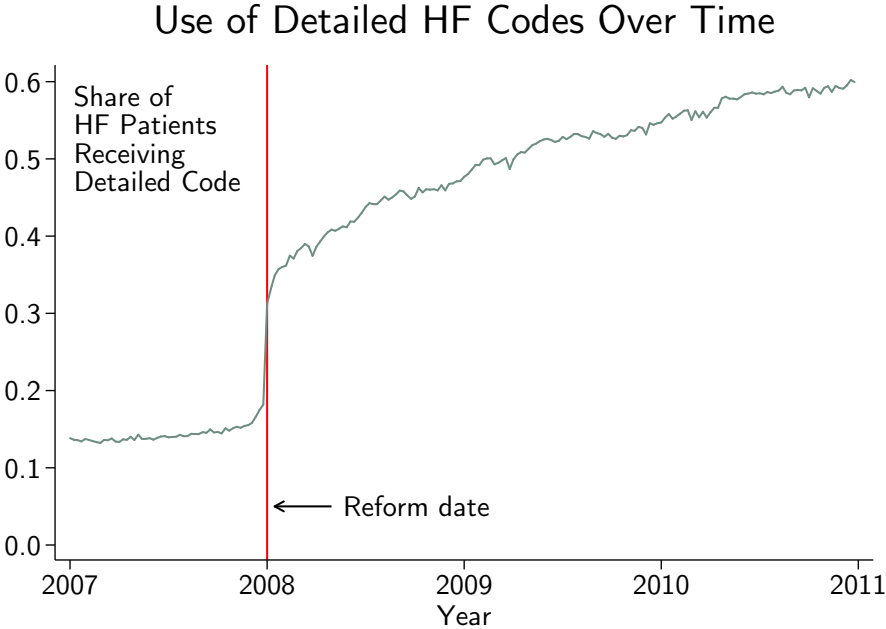
Appendix Figures

Use of Detailed HF Codes Over Time

Share of
HF Patients
Receiving
Detailed Code

0.6
0.5
0.4
0.3
0.2
0.1
0.0

← Reform date

2007   2008   2009   2010   2011
Year

Figure A1

HF Coding and Heart Testing Following Reform

Share of
Revenue
Taken Up
(Coding)

Share of
Patients
(Testing)

Detailed Coding
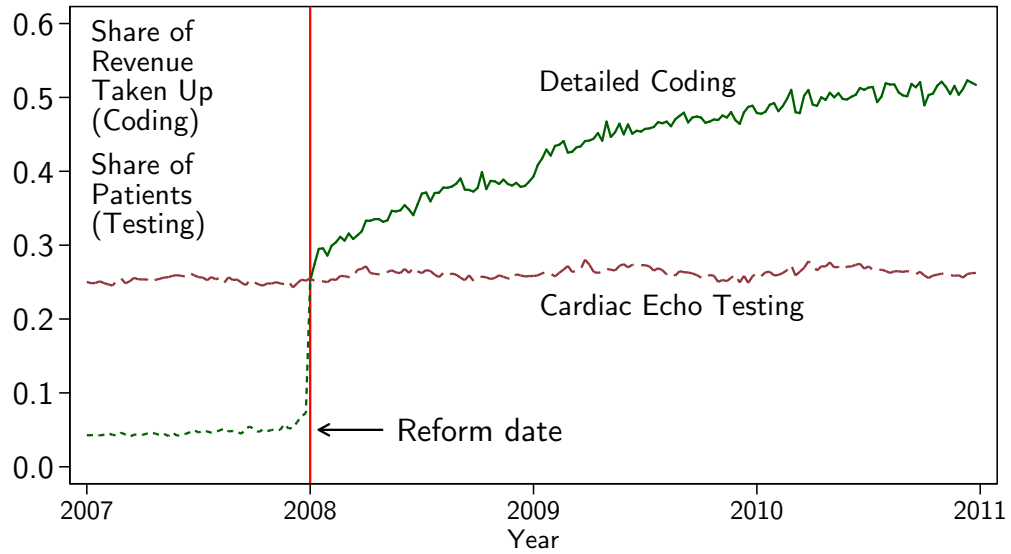
Cardiac Echo Testing

← Reform date

Year

Figure plots the weekly share of revenue available for detailed coding of HF that was captured by hospitals alongside the weekly share of all patients who received a cardiac echo, a heart test. The dotted line shows revenue that would have been captured in 2007 if hospitals had been paid per 2008 rules.

Figure A2

Revenue at Stake per Patient across Hospitals
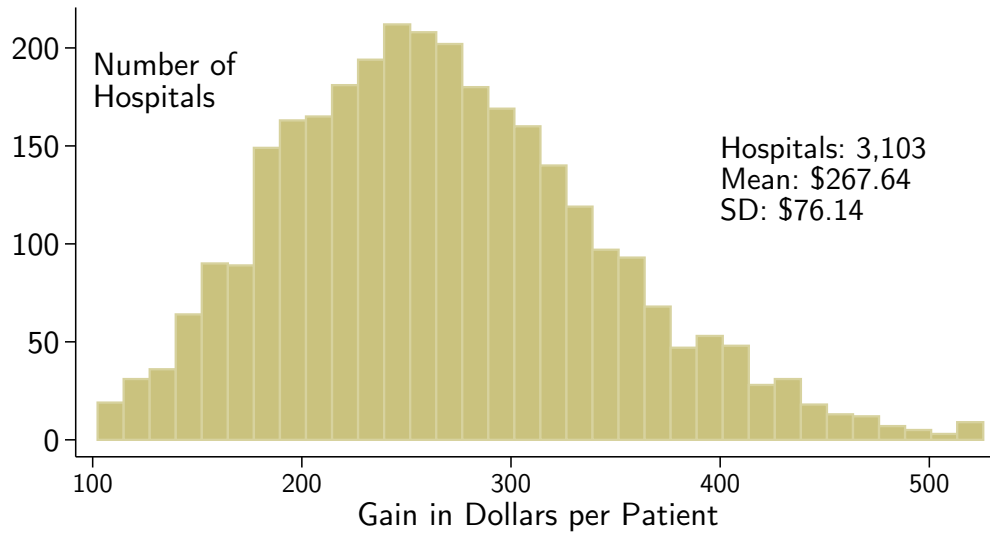
Number of Hospitals

Hospitals: 3,103
Mean: $267.64
SD: $76.14

Gain in Dollars per Patient

Revenue at stake is calculated using pre-reform (2007) patients processed under post-reform (2009) payment rules. The prediction process is described in the appendix. The 422 hospitals with <50 HF patients are suppressed and the upper and lower 1% in revenue at stake per patient are then removed.

Figure A3

Appendix Tables

Table A1 - Other Codes for HF

		Severity		Counted as
Code	Description	Before	After	Specific
428.1	Left HF	High	Medium	Yes
398.91	Rheumatic HF	High	Medium	Yes
402.01	Malignant HHD w/ HF	High	Medium	Yes
402.11	Benign HHD w/ HF	High	Low	No
402.91	Unspecified HHD w/ HF	High	Low	No
404.01	Malignant HHCKD, CKD stage 1-4/unspec w/ HF	High	Medium	Yes
404.11	Benign HHCKD, CKD stage 1-4/unspec w/ HF	High	Medium	Yes
404.91	Unspecified HHCKD, CKD stage 1-4/unspec w/ HF	High	Medium	Yes
404.03	Malignant HHCKD, CKD stage 5/ESRD w/ HF	High	Medium	Yes
404.13	Benign HHCKD, CKD stage 5/ESRD w/ HF	High	Medium	Yes
404.93	Unspecified HHCKD, CKD stage 5/ESRD w/ HF	High	Medium	Yes

This table lists ICD-9 codes besides those of Table 1 that indicate heart failure. These codes can be used alongside the codes listed in Table 1. Codes that raise patients to medium or higher severity after the reform are counted as specific codes. Definitions: HHD - hypertensive heart disease, HHCKD - hypertensive heart and chronic kidney disease, CKD - chronic kidney disease, ESRD - end stage renal disease.

Table A2 - Relationship Between Coding and Testing

Outcome	(1)	(2)	(3)	(4)	(5)	(6)
	Detailed HF Coding			Echocardiogram		
Post ×	0.46***	0.46***	0.45***	0.067***	0.059***	0.059***
Ex Ante HF Rate	(0.024)	(0.023)	(0.022)	(0.014)	(0.013)	(0.013)
Observations	47,355,656	47,355,656	47,355,656	9,460,759	9,460,759	9,460,759
R^2	0.076	0.082	0.117	0.087	0.098	0.101
Patient Controls	None	A/R/S	Full	None	A/R/S	Full

These regressions are at the patient level and are based on the "grand sample" described in the text. The sample is expanded to include patients with and without HF, and is extended to include all such patients from 2005-2010. Post is an indicator for year 2008 and later. Ex ante HF rate is the fraction of patients in the patient's major diagnostic category (MDC) in 2003-2004 with HF. The outcome in columns 1-3 is an indicator for whether the patient received a detailed HF code and in columns 4-6 it is an indicator for whether the patient received an echocardiogram (these columns use the 20% of observations for which echocardiograms are observed). All regressions include year and MDC fixed effects. Columns 1 and 4 include no patient controls. Columns 2 and 5 control for age, race, and sex interactions. Columns 3 and 6 further add controls for histories of chronic conditions. Standard errors clustered at the MDC level in parentheses.

*** significant at 1% level; ** significant at 5% level; * significant at 10% level

Table A3 - Coding Dispersion by Study Sample

Sample \ Metric	(1) N	(2) Mean	(3) Raw SD	(4) Adj SD
Grand sample (all hospitals)	3,414	0.494	0.244	0.230
Grand sample (hospitals ≥50 HF patients)	3,081	0.524	0.220	0.218
Analysis sample (step 1)	2,831	0.526	0.220	0.212
Analysis sample (step 2)	2,341	0.546	0.201	0.199

This table presents the average and standard deviation of HF coding across hospitals in 2010. Column 4 adjusts the standard deviation for measurement error (see Section 3.1.2). The grand sample refers to all HF patients and is described in Sections 2.3 and 4.1; I present statistics including all hospitals and dropping hospitals with fewer than 50 HF patients in 2010. The step 1 analysis sample is described in Section 4.1. The step 2 analysis sample is the subset of these hospitals for which all characteristics are observed, and is the focus of Sections 4.3-4.5.

Table A4 - Coding at the Hospital System and Region Levels

Level	(1) Hospital System	(2) Region (HRR)	(3) Region (MSA)
Share of Variation in Hospital Effects Explained by System/Region Fixed Effects			
No First-Step Controls	0.249	0.163	0.166
First-Step Admission Patient Controls	0.261	0.234	0.175
First-Step Full Patient Controls	0.260	0.233	0.176
First-Step Full Patient Controls and Physician FE	0.093	0.168	0.207
Observations (Number of Hospitals)	1,521	2,341	1,705
Fixed Effects (Number of Systems/Regions)	321	299	311
Standard Deviation of Coding Across Systems/Regions			
No Controls	0.167	0.109	0.129
Admission Patient Controls	0.122	0.084	0.100
Full Patient Controls	0.121	0.083	0.099
Number of Systems/Regions	341	303	314

This table analyzes variations in coding at the hospital system and geographic region levels. The upper section shows the share of variation in coding across hospitals that can be explained by regressing hospital coding on system (column 1) and region (columns 2 and 3) fixed effects. Shares are adjusted for regressand measurement error. The set of hospitals is restricted to the 2,341 in the previous dispersion and regression analyses.

The lower section estimates the dispersion in coding across systems/regions by running equation 1 with system or region fixed effects in place of hospital and physician fixed effects. Dispersion is adjusted for measurement error. The sample is all patients treated at the 2,831 "step one" analysis sample hospitals – the same sample previously used to estimate hospital coding scores.

Hospital systems are derived from AHA data. Regions are measured using hospital referral regions (HRRs) and metropolitan statistical areas (MSAs) based on the location of the hospital.

Table A5 - Association Between Region Characteristics and Coding

Outcome	(1) Score	(2) Score	(3) Score	(4) Score	(5) Score	(6) Score	(7) Score	(8) Score	(9) Score
Total spending per enrollee	-0.015*** (0.005)	-0.013*** (0.004)	-0.013*** (0.004)						
Hospital and SNF spending per enrollee				-0.006 (0.010)	-0.007 (0.008)	-0.007 (0.008)	0.046*** (0.013)	0.032*** (0.010)	0.031*** (0.010)
Physician spending per enrollee							-0.082*** (0.016)	-0.063*** (0.012)	-0.065*** (0.012)
Outpatient facility spending per enrollee							-0.062** (0.030)	-0.046* (0.023)	-0.047** (0.023)
Home health spending per enrollee							-0.058** (0.026)	-0.037** (0.018)	-0.036** (0.018)
Hospice spending per enrollee							0.018 (0.053)	-0.007 (0.041)	-0.005 (0.041)
Durable medical equipment spending per enrollee							-0.220 (0.156)	-0.183 (0.123)	-0.168 (0.122)
Observations	303	303	303	303	303	303	303	303	303
R^2	0.030	0.036	0.036	-0.002	-0.000	-0.000	0.110	0.108	0.109
Patient Controls	None	Admission	Full	None	Admission	Full	None	Admission	Full
Physician Controls	None	None	None	None	None	None	None	None	None

This table presents the results of regressing Hospital Referral Region (HRR) coding scores on HRR characteristics, taken from Dartmouth Atlas data on Medicare spending in 2010 across HRRs. All spending is measured in $1,000s and adjusted for price, age, sex, and race. Columns 1, 4, and 7 use no controls to calculate the HRR scores; columns 2, 5, and 8 add controls for patient characteristics observable upon admission; and columns 3, 6, and 9 add histories of chronic conditions. Robust standard errors in parentheses.

*** significant at 1% level; ** significant at 5% level; * significant at 10% level