Technological Diffusion Across Hospitals: The Case of a Revenue-Generating Practice

Adam Sacarny*

July 26, 2016

Abstract

Productivity-raising technologies tend to diffuse slowly, particularly in the health care sector. To understand how incentives drive adoption, I study a technology that generates revenue for hospitals: the practice of submitting detailed documentation about patients. After a 2008 reform, hospitals were able to raise their total Medicare revenue over 2% by always specifying a patient's type of heart failure. I find that hospitals only captured around half of this revenue, indicating that large frictions impeded takeup. A major barrier is a principal-agent problem, since doctors supply the valuable information but are not paid for it. 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 also robustly correlated with the ability to generate survival for heart attack patients and the use of inexpensive survival-raising standards of care. These findings highlight the importance of agency conflicts in explaining variations in health care performance, and suggest that principal-agent problems may drive disparities in performance across firms more generally.

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

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

*Department of Health Policy and Management, Columbia University Mailman School of Public Health, ajs2102@columbia.edu. I am grateful to Amy Finkelstein, Michael Greenstone, Jon Gruber, and Paulo Somaini for their advice and guidance on this project. I thank Isaiah Andrews, Emily Baneman, David Chan, Manasi Deshpande, Amos Dodi, Kate Easterbrook, Ben Feigenberg, Eliza Forsythe, Paul Goldsmith-Pinkham, Tal Gross, Sally Hudson, Greg Leiserson, Conrad Miller, David Molitor, Dianna Ng, Iuliana Pascu, Maxim Pinkovskiy, Maria Polyakova, Miikka Rokkanen, Annalisa Scognamiglio, Brad Shapiro, Henry Swift, Melanie Wasserman, Nils Wernerfelt, and participants in the MIT Public Finance lunch for their comments and suggestions. I would also like to thank Jean Roth for her assistance with the Medicare data. I gratefully acknowledge support from the Robert Wood Johnson Foundation and funding from the National Institute on Aging grant T32-AG000186.

1 Introduction

Technology is usually believed to be a key driver of cross-country income disparities and economic growth. A classic finding of studies of technology is that new forms of production diffuse 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 Loecker, 2015; Bloom et al., 2012). Given the enormous productivity gains that result from many of these technologies, the nearly ubiquitous finding of delayed takeup is particularly vexing.

In this paper, I study a health care technology that raises revenue for the hospital: the detailed reporting of heart failure patients. A 2008 Medicare policy change created a financial incentive for hospitals to provide more detail about their patients in insurance reimbursement claims. Yet hospitals could only provide these details if they were documented by physicians. By tracking the diffusion of the reporting practice across hospitals, this study examines the role of financial incentives and agency conflicts in the adoption of new technologies.

These incentives are particularly important as insurers, spurred by mandates in the Affordable Care Act, have sought to use their purchasing power to raise the productivity of health care providers. In designing payment schemes like Medicare's Hospital Value-Based Purchasing Program, policymakers have focused on differences in the utilization of survival-raising processes of care, including checklists, hand-washing, and drugs like β-blockers. 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). These processes of care require the coordination of hospitals and physicians, creating agency conflicts like those in the reporting of heart failure. While improved heart failure billing is a revenue-raising but not survival-raising technology, it is a clear test case of how financial incentives drive diffusion in the presence of agency frictions.

Hospitals have the option of listing heart failure on a reimbursement claim with detailed codes that describe the type of heart failure, or they may submit a vague code that provides little additional information about the condition. A 2008 reform changed the pricing function of Medicare to begin

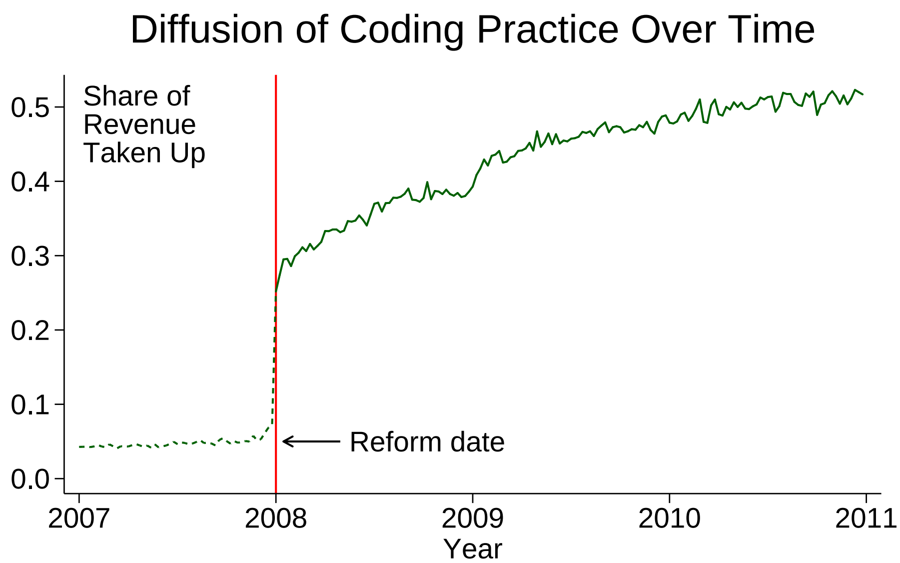


Figure plots the share of revenue available for detailed coding of HF that was captured by hospitals over time. Dotted line shows revenue that would have been captured in 2007 if hospitals had been paid per 2008 rules. See Appendix Section A.1.2 for more details.

Figure 1

providing additional payments for the detailed codes.[1] To capture this reward, hospitals needed to change how they reported their patients to Medicare. However, they could only make this change if doctors provided them with extra documentation about the heart failure to support it. 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.

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 (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

[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.

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) finds 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.

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. To explain the incomplete and varied adoption of the codes, I focus on frictions due to agency problems between a hospital and its doctors. Physicians are responsible for writing down the extra information about the heart failure, but Medicare does not pay physicians for the detailed codes or anything else that might be produced from the information.

The principal-agent problem that this reform invokes is a classic one in economics – in other settings, it has been suggested as a driver, for example, of the failure of high quality management practices to diffuse across firms (Gibbons and Henderson, 2012). It plays a particular role in the American health care system because hospitals and physicians are frequently paid on independent bases. Moreover, hospitals are legally restricted from formally sharing their Medicare payments with physicians as incentive pay. In spite of these restrictions, many policies to improve the quality of care have focused on the hospital's or physician's payment system alone.

The agency issues created by this reform arose from the bifurcated payment system. Hospitals – the principals – had large incentives to submit detailed codes about their patients, while physicians – the agents – had no direct incentive to provide the information. To resolve the principal-agent problem, hospitals would need to work with their doctors to better document their patients' conditions, then translate this documentation into the newly valuable specific codes.

To study the role of these agency problems, I consider adoption rates that control for the physician. 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 a new 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.

Subtracting the physician contribution removes dispersion in adoption due to hospitals having different kinds of doctors. I thus address the concern that doctors who work at some hospitals may be more willing to provide detailed documentation than doctors who work at other hospitals. I show that dispersion is, if anything, slightly increased when the hospital component is isolated: the standard deviation of the share of patients who received detailed documentation across all hospitals rises from 15 percentage points with rich patient controls to 16 percentage points with patient and physician controls. The presence of 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. This result raises the possibility that institution-level principal-agent problems underlie some of the productivity differences that have been found among seemingly similar enterprises.

I also consider the correlation between hospital adoption and hospital characteristics like size, ownership, location, and clinical performance. The signs of these relationships are not *ex ante* obvious, and they indicate which types of hospitals were most able to extract the codes from their doctors. The most powerful predictor of hospital adoption in this analysis is clinical quality, which I capture by two measures: adjusted heart attack survival (the survival rate of heart attack patients after accounting for spending on medical inputs and patient characteristics) and utilization of inexpensive, survival-raising processes of care (which includes administering aspirin after heart attacks and providing antibiotics before high-risk surgeries, among other evidence-based interventions). 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.

I contribute to the growing literature on productivity disparities and technological diffusion in three 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 that none might exist.

These disparities reflect differences in hospitals' basic ability to respond to incentives. Second, using decomposition techniques adapted from studies of the labor market, 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 doctors than others. Lastly, 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. Taken together, these findings hint that principal-agent problems may play a role in productivity dispersion more generally – inside and outside the health care sector.

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 presents results on dispersion in hospital takeup, then shows how takeup relates to hospital characteristics and measures of treatment performance. Section 4 provides a discussion of the results. Section 5 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 who suffer from HF (Linden and Adler-Milstein, 2008). HF is listed as a diagnosis on more than one-fifth of Medicare hospital stays. Limiting to hospital expenditures, the program spends more on diagnosing and treating patients with HF than on patients with heart attacks. HF spending also outstrips spending on patients with all forms of cancer combined (Massie and Shah, 1997).

Medicare's payment for heart failure is especially consequential for health expenditures and salient to hospital administrators, yet the classic economic literature on health care eschews studying HF in favor of less common conditions like 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 coding incentives.

My analysis focuses on the revenue generating practice of better documenting HF on hospital inpatient reimbursement claims to Medicare. The hospitals I study are paid through Medicare's Acute Inpatient Prospective Payment System (IPPS), a $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 was 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 25% of high-severity patients (or 20% of patients overall) in 2007.

Risk adjustment relies on detailed reporting of patients by providers, but according to the Centers for Medicare & Medicaid Services (CMS), the agency that administers Medicare, the overwhelmingly most common of the HF codes – 428.0, "congestive heart failure, unspecified" – was

[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 included only Non-CC and CC.

Table 1 - Vague and Specific HF Codes

Code	Description	Severity Before	Severity 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

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

vague. Moreover, 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 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. These codes and their severity classifications are listed in 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 or diastolic part of the cardiac cycle is affected and, optionally, whether the condition is acute or chronic. Submitting them is a process that requires effort from both physicians and hospital staff and coordination between the two. In this way it is similar to other technologies that have come into the focus of researchers and policymakers, including the use of β-blockers (an inexpensive class of drugs that have been shown to raise survival following a heart attack; 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 basic framework for analyzing takeup of the technology 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 the assumption of additive separability of the hospital and the doctor's effects on the coding probability, 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 extract specific HF 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 of providing a specific code.

The dispersion of the hospital types is of direct interest, and 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., 2013, 2016b). Dispersion in hospital types is therefore a form of productivity dispersion.

There are several potential drivers of this dispersion, one of which relates to agency issues. Hospitals were constrained from directly incentivizing their doctors to provide the additional documentation needed to submit a specific HF code. When a doctor moves from a low-type hospital to a high-type hospital, her HF patients become more likely to have a detailed code, regardless of the doctor's type. One perspective is that this difference is due to the high-type hospital better solving the principal-agent problem. The variation in hospital types can reflect variation in whether hospitals can bring their doctors' behaviors in line with the hospital's incentives. Dispersion may also be driven by differences across hospitals in the use of advanced electronic medical records which extract codes from the physician's documentation and variations in the quality of hospital health information staff, who must translate the documentation into codes.

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 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 two sets of characteristics:

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

The signs of the elements of ρ 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 and Z_h and drive takeup. I discuss these potential relationships and estimate this equation in Section 3.5.

2.3 Data

I study the impact of the IPPS reform on the diffusion of the revenue generating practice using a dataset of all inpatient hospitalizations for Medicare beneficiaries. My data is primarily drawn from the MEDPAR Research Information File (RIF), a 100% sample 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 MEDPAR files, and I source additional information about patients from the enrollment and chronic conditions files. 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 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 having a medium- or high-severity patient was so much higher than the low-severity payment, hospitals had a clear incentive to use detailed codes whenever possible. Before the reform, the gain from these detailed codes relative to the vague code was zero because they were 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

[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 cardiac cycle affected – and to raise the hospital's payment. See Table 1 and Table A1.

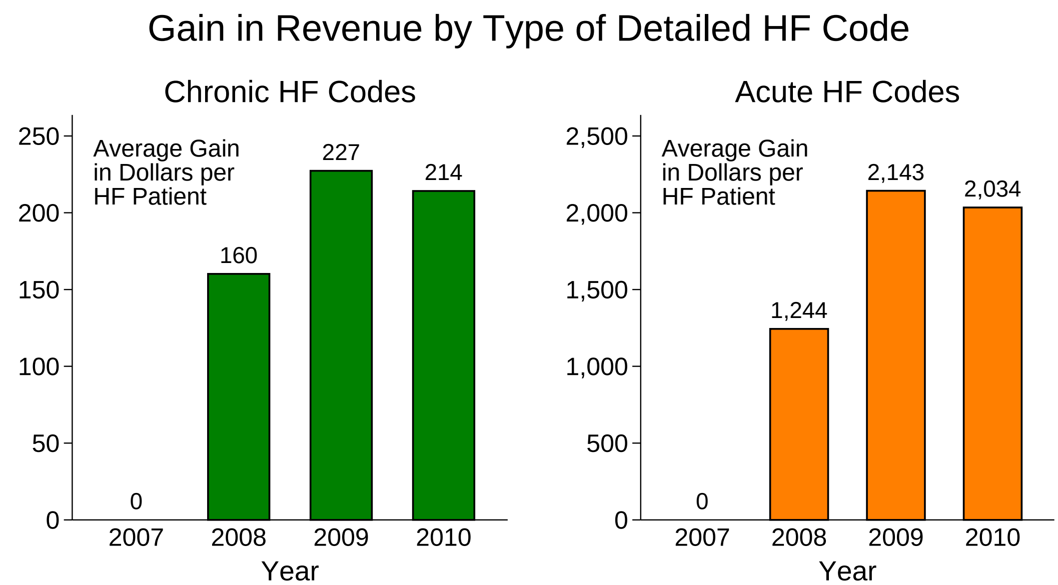


Figure plots the average per−HF patient gain in revenue going from always using vague codes for HF patients to always using chronic codes or acute codes. Prices in 2009 dollars.

Figure 2

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, so that the 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 the use of these codes is shown in Figure 3.

[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.

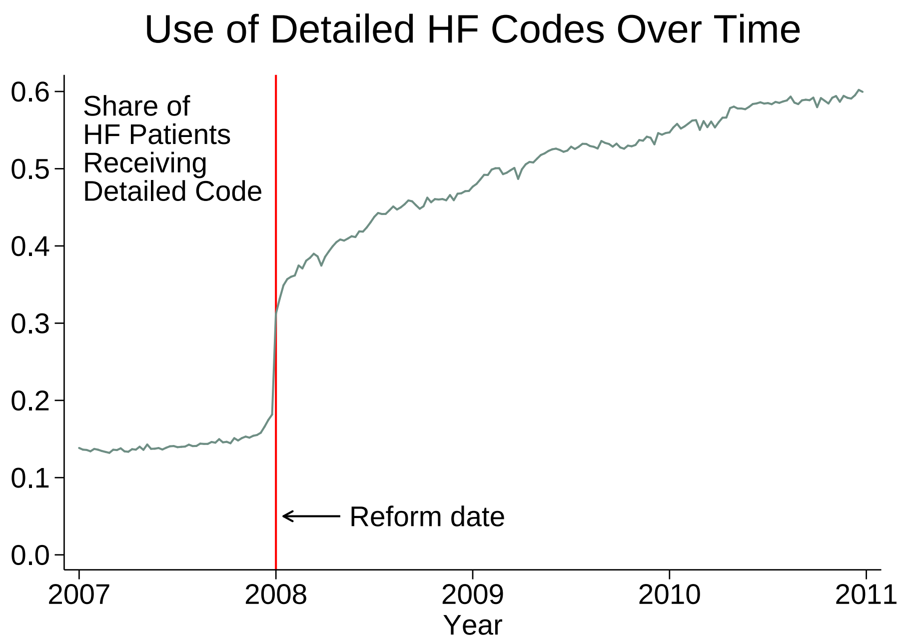


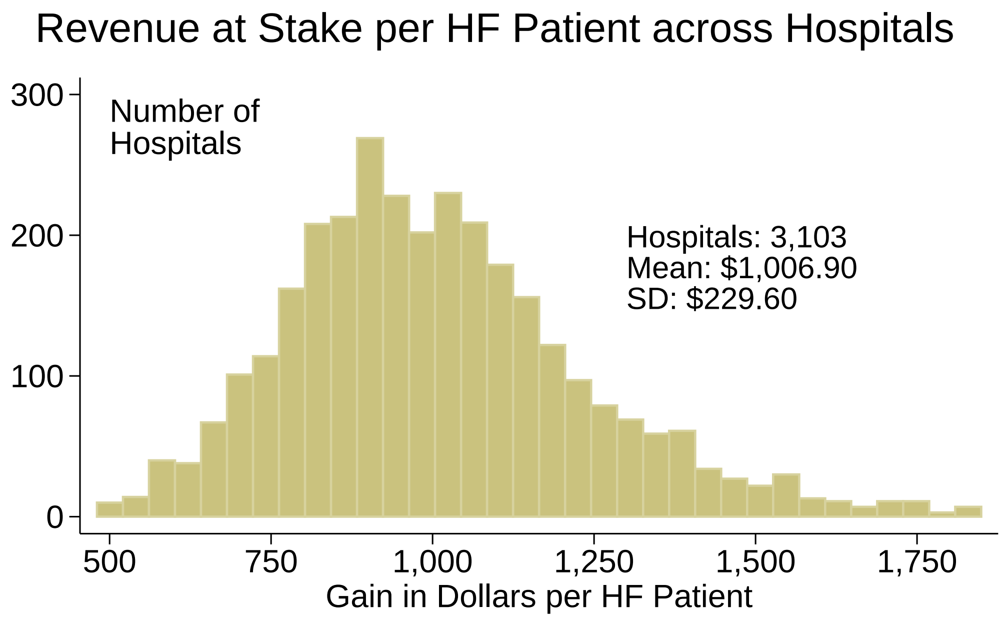
Figure 3

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, probabilistically fills in the detailed HF codes the patients would have received under full adoption of the coding technology, and determines the ensuing expected gain in payment from these codes by processing the patient under the new payment rules. Heart failure codes are predicted using the relationship between coding and patient characteristics in hospitals that were relatively specific coders in 2010.[6]

Figure 4 shows the high level of and variation in *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 (Figure A2 shows the dispersion in the gain when it is spread across all Medicare admissions). The standard deviation of the revenue at stake per HF patient was $230.

To provide a sense of scale, one can consider these amounts relative to hospital operating margins.

[6]This predictor is applied to all 2007 HF patients in the grand sample with data on chronic conditions. It is described in greater detail in Appendix Section A.1.2.



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 4

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 Costs of Takeup

Figure 1 shows that the large amount of revenue at stake for specific coding induced an almost instantaneous partial takeup of the coding. 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 the technology.

For a hospital to legally submit a detailed code, a doctor must state the details about the HF in

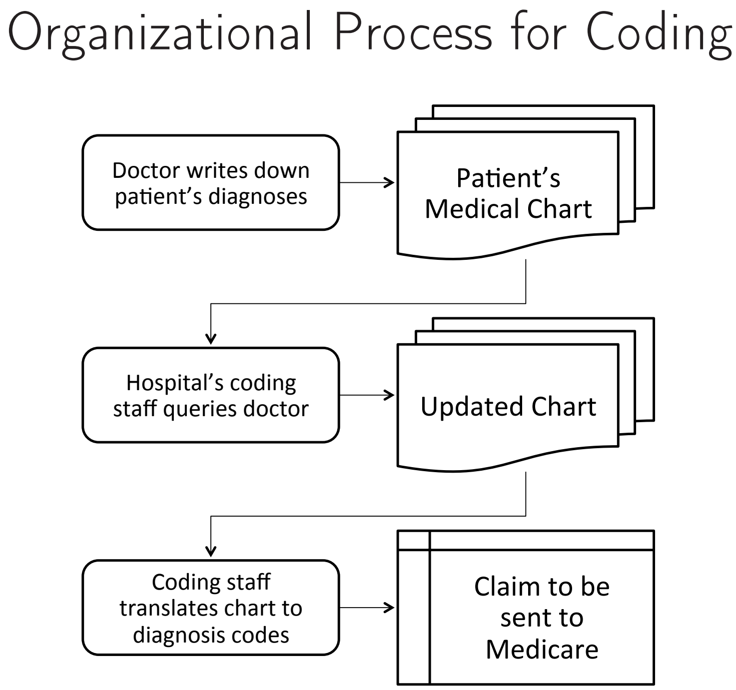


Figure 5

the patient's medical chart.[7] Figure 5 presents a flowchart of the organizational processes involved in the coding of patients. As the physician treats a patient, she writes 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) use when processing the chart (Youngstrom, 2013). The staff can review the chart and send it back to the doctor with a request for more information – this process is called querying. Then, the staff must convert the descriptions of diagnoses into the proper numeric diagnosis codes, which becomes 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

[7]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.

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).

One possibility is that taking up the reform requires medical testing of HF patients to confirm the details of their conditions. The gold standard for confirming whether there is systolic or diastolic dysfunction – the minimum amount of information needed to use a specific code – is an echocardiogram, a non-invasive diagnostic test. 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.

Contrary to this story, the official coding guidelines indicate that whatever were the costs of more detailed HF coding, they 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). Thus these codes require only suggestive evidence, not the certainty of an echocardiogram.

With enough information to diagnose and submit a vague HF code, it is almost always possible to provide enough additional documentation to legally submit 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. Figure A1 shows that the enormous increase in the capture of HF coding revenue was not matched by any perceptible change in heart testing as measured by the share of all patients receiving an echocardiogram, as identified by physician claims on the 20% sample of patients for which I observe them.

A key source of takeup frictions 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. For example, the American College of Physicians has expressed the view that:

> "The primary purpose of clinical documentation is to facilitate excellent care for patients. Whenever possible, documentation for other purposes should be generated as a byproduct of care delivery rather than requiring additional data entry unrelated to care delivery." (Kuhn et al., 2015; p. 10)

Taking up the revenue-generating practice required hospitals to pay a variety of fixed and variable costs that were unrelated to patient treatment but could influence physicians' documentation styles. Examples of these costs include training hospital staff to prompt doctors for more information when a patient's chart lacks details and purchasing health information technology that prompts staff to look for and query doctors about high-value 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 mortality and morbidity (Voss and Widmer, 1997; Stafford and Radley, 2003; Pittet et al., 1999).

2.6 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 from 2010. For 84% of these stays, I observe the patient's history of chronic conditions as well as the 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.[8] Hospital and physician

[8] I use the attending physician identifier from the Medicare Inpatient RIF. To ensure that only valid individual physicians are included, I drop physician identifiers that could not be found in the AMA Masterfile, a census of all physicians, which accounts for most of the stays for which the physician was not observed.

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

Table 2 - Statistics about the 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 analysis sample includes 1,510,988 HF patients. See text for more details.

types are only separately identified within a "mobility group" – the set of hospitals and physicians that can be connected, in graph theory terms, by physicians who work at multiple facilities (this concept is explained in greater detail in Section 3.2). I call the 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 analysis 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. Mobile doctors are crucial for my analyses because their behavior separately identifies the hospital and doctor types. In this sample, the average doctor sees 12 HF patients in a given year and works at 1.23 distinct hospitals. About 19% of doctors are mobile.

Table 3 provides additional information about the doctors by mobility status using data from the AMA Masterfile. The average mobile physician treats about twice as many patients as a non-mobile

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.

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	11.6	20.2	9.63
Distinct Hospitals	1.23	2.24	1
Mobile (>1 hospital)	0.18	1	0
Type of Physician			
Primary Physician	0.45	0.50	0.44
Medical Specialist	0.30	0.34	0.29
Surgeon	0.23	0.14	0.24
Demographics			
Female	0.19	0.15	0.20
Age	49.1	48.9	49.1
Training and Experience			
Years in Training	5.96	6.51	5.83
Years Since Training	16.0	15.4	16.1
Trained in US	0.71	0.59	0.73
Physicians	130,487	24,060	106,427

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 physician type, demographics, training, and experience derived from AMA Masterfile.

physician.[9] 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 practicing since completing training – than their non-mobile counterparts, and they are also more likely to have received their medical training outside the U.S.

3 Hospital Adoption

In this section, I present an analysis of the role that physicians played in the adoption of the revenue generating practice. I decompose the hospital's average coding into the component that is due to the

[9] 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

facility and the component that is due to its doctors. The notion of outcomes being due to a hospital and doctor component follows a commonly used econometric model of wages that decomposes them into firm and worker effects (see e.g. Abowd et al., 1999 and more recently Card et al., 2013, which study wages in France and Germany, respectively).

This section undertakes two key analyses. First, it considers the dispersion in the adoption of detailed HF coding among observably similar hospitals and 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. Equivalently, it tests whether the probability a HF patient treated by a particular doctor gets a specific code varies across hospitals.

Second, it explores the relationship between adoption and hospital characteristics like size, ownership, and clinical quality. The signs of these relationships are not *ex ante* obvious, but they speak to several important and open questions in health economics. Though these results are descriptive they are useful policy inputs: they can be interpreted as indications of which providers are most elastic to incentives for revenue generating practices.

3.1 Econometric Specification

The key analyses of this section describe the distribution of the adoption of the coding technology 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 doctor effects. In the second step, I analyze the distribution of the fixed effects by calculating their variance (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 $d\left(p\right)$ 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 $\hat{\alpha}_h$:

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

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 to code. The specifications with rich sets of patient observables account for 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 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, including its ability to resolve 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 Andrews et al., 2008 and Abowd et al., 2002).

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).[10]

Describing the Adopters The other key analysis of this section 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 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 ρ or θ.

3.2 Separate Identification of Hospital and Physician

The health care context is unique because it allows the separate identification of the contribution of the principal and the contribution of the agent to takeup – a decomposition that cannot be performed when agents are observed under just one principal. The key insight behind the decomposition in the heart failure setting is that physicians are frequently observed treating patients 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 physician fixed effects, when they are included in the first step, sweep out the component of the hospital's coding that is due to the behavior of its doctors. The hospital and physician effects can be separately identified within a mobility group – the set of doctors and hospitals that are said to be "connected" to each other. 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

[10] The Chandra et al. (2013) 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.

or hospital and includes all other doctors and hospitals that are connected to her or it (called a maximal connected subgraph). There are 3,414 hospitals in the grand sample that report at least one HF patient in 2010. At 3,381 hospitals there is at least 1 HF patient for whom his history of chronic conditions and attending physician is observed. These patients number 1.6 million and are treated by 136,067 distinct physicians. The largest mobility group – the analysis sample – is a subset of this sample containing 2,831 hospitals, 1.5 million patients, and 130,387 physicians.

The econometric model of the first step follows from identification assumptions described at length in Abowd et al. (1999) and Card et al. (2013). In this context, the key assumption 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 only an approximation, 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 the explanatory power of the model minimally, suggesting that additivity is not a restrictive assumption in this context.[11]

One implication of the conditional expectation equation is that patients do not select hospitals or doctors on the basis of unobserved costs of coding. If this were the case, for example, the fixed effect of a hospital with unobservably more costly to code patients would be estimated with negative bias. I test this assumption by including increasingly rich sets of patient characteristics as controls. The coefficients with practically significant magnitudes in the regressions of adoption on hospital covariates tend to attenuate by at most one-third due to the inclusion of rich patient characteristics observable in the patient's hospital billing claim, but they are not further reduced by including controls for patient histories of chronic illnesses (the controls are described in section 3.4). These coefficients remain statistically significant even though they attenuate. Likewise, the

[11] 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.

standard deviation of adoption is reduced by about one-fourth from the patient controls, and again the reduction is entirely due to characteristics in the billing claim and not from the further addition of patient histories of chronic illness.

It is perhaps unsurprising that patient characteristics influence the hospital's use of the codes. The fact that adding patient illness histories as an additional set of controls does not further affect dispersion in adoption suggests that the key factors are attributes of the patient's admission. The identifying econometric assumption is that unobserved characteristics are not playing a role in coding, and the information observable about the admission is detailed in the claims data that I use. The great majority of disparities in adoption across hospitals cannot be attributed to anything observable about the patient, and I present all results in this study under three patient-level specifications to show that they persist in all of the approaches.

A related 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 11 – 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.

3.3 Hospital Characteristics

Table 4 shows summary statistics for the cross section of hospitals that I include in the dispersion and characteristics of adopters analyses. This cross section consists of the 2,386 hospitals in the analysis sample for which I observed complete information on all baseline characteristics, standards of care, and treatment performance.

The rows of the table comprise the key hospital characteristics and performance measures that are used in the analyses. Hospital size (beds) and ownership are taken from the Medicare Provider of Services file. Ownership is split between non-profit (about two-thirds of hospitals), for-profit

Table 4 - Hospital Summary Statistics

Patient Controls	(1) Mean	(2) SD
Heart Failure Coding		
HF Patients	597.6	512.9
Share Given Specific Code	0.546	0.199
Hospital Characteristics		
Beds	286.5	232.0
Ownership		
Non-Profit	0.668	0.471
For-Profit	0.164	0.370
Government	0.168	0.374
Location		
Large Urban Area	0.422	0.494
Other Urban Area	0.353	0.478
Rural Area	0.225	0.418
Teaching Status		
Major Teaching Hospital	0.101	0.302
Minor Teaching Hospital	0.274	0.446
Non-Teaching	0.625	0.484
Ex Ante $ at Stake / Patient	267.6	71.88
Standards of Care (share of times standards used in 2006)		
for Heart Attack Treatment	0.916	0.085
for Heart Failure Treatment	0.826	0.113
for Pneumonia Treatment	0.864	0.061
for High-Risk Surgeries	0.797	0.119
Heart Attack Treatment (patients in 2000-2006)		
Adjusted 30-Day Survival	0.813	0.030

N=2,386 hospitals. See text for more details on the source and definitions of the characteristics. The standard deviations of specific coding for HF and heart attack survival account for sampling variance.

(one-sixth), and government-run (one-sixth). 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 23% of the hospitals in the sample are rural, reflecting that many hospitals that site in rural areas are classified as critical-access facilities, which were exempt from this reform and excluded from my analyses. Teaching hospitals are defined as those with residents; major teaching facilities are the 10% with a resident-to-bed ratio of at least 0.25; minor teaching facilities are the 27% 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. These calculations are described in full detail in Appendix Section A.1.2.

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 heart attack, heart failure, pneumonia, and high-risk surgery patients in 2006. 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 heart attacks and has been the subject of several studies of technology diffusion (see e.g. Skinner and Staiger, 2007, 2015).

Adjusted heart attack 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 heart attack 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

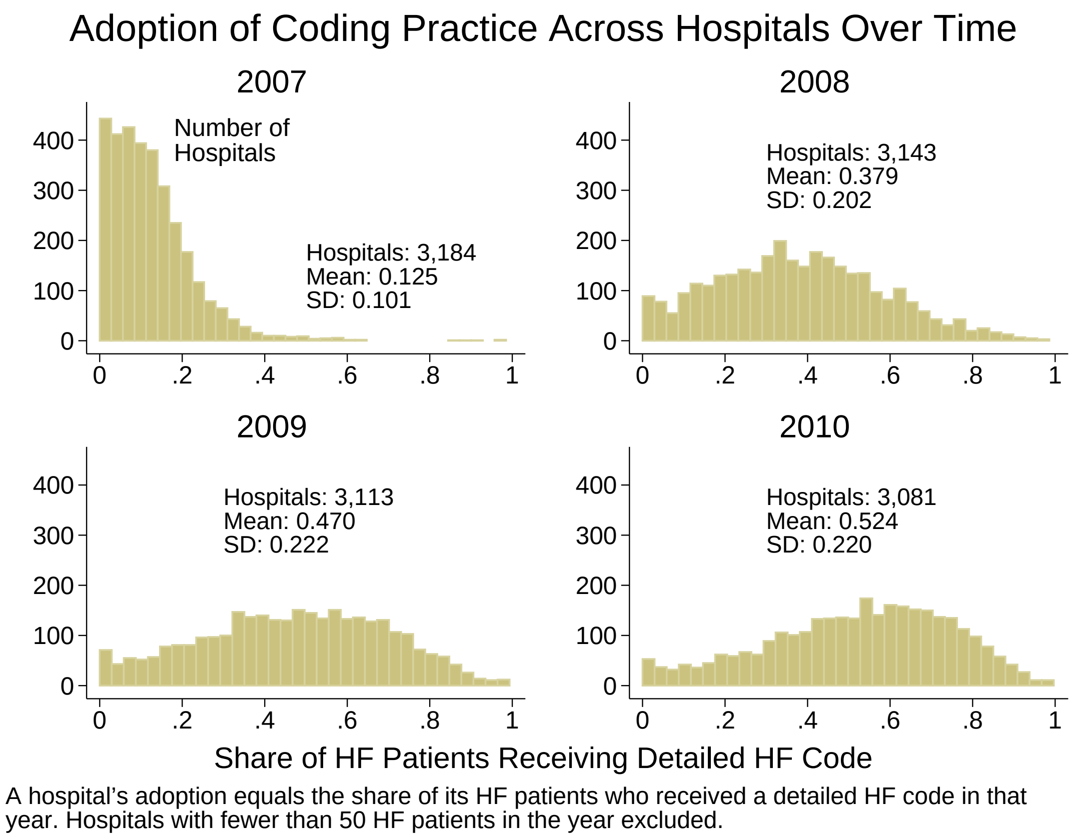
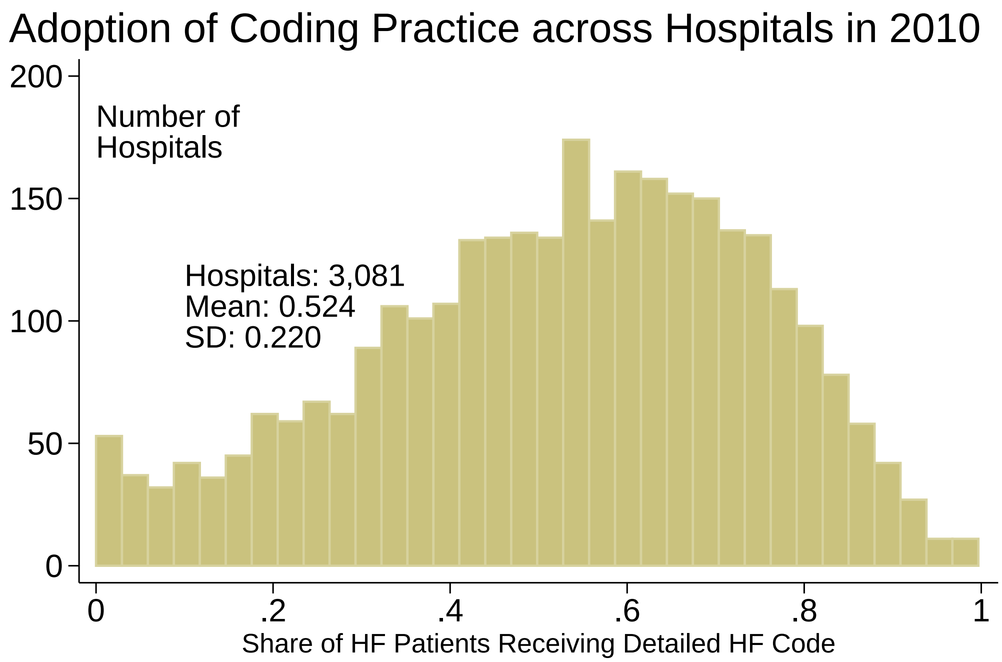


Figure 6

greater probability that the patient survives 30 days. This rate is adjusted to account for measurement error using an Empirical Bayes shrinkage procedure described in more detail in Appendix C of Chandra et al., 2016a. 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.

3.4 Dispersion

I find dispersion in adoption with and without rich patient and physician controls. To provide a sense of the time series of adoption, Figure 6 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 moved rapidly after the reform. By 2010, the median hospital used specific codes 55% of the time. Figure 7 shows the full distribution of $\hat{\alpha}_h^{simple}$ in 2010, the analysis sample year. There was great variation in takeup across hospitals even in the third year following the reform. In



A hospital's adoption equals the share of its 2010 HF patients who received a detailed HF code. Hospitals with fewer than 50 HF patients in 2010 are excluded.

Figure 7

particular, there was a substantial mass of hospitals using detailed codes less than 20% of the time, and a nontrivial number of hospitals that almost never used them.[12]

Table 5 shows the standard deviation of adoption among homogeneous categories of hospitals. I divide the space of hospitals on the basis of characteristics that have been the focus of literature on hospital quality. The table includes three sets of patient controls in the first step, which is where the hospital effects are extracted. In the left three columns, each patient control specification is presented without first-step physician effects; in these results, the hospital effects include the component of coding that is due to the physicians. The right three columns add first-step physician effects, which subtracts the physician component.

The first patient control approach, presented in columns 1 and 4, includes no patient-level controls. The second, presented in columns 2 and 5, includes observables about the patient's hospital admission found in the hospital's billing claim: age, race, and sex interactions; whether

[12] These figures plot the $\hat{\alpha}_h^{simple}$ with no adjustment for measurement error, but they exclude hospitals with fewer than 50 HF patients to limit the scope for measurement error to drive dispersion. To construct the $\hat{\alpha}_h^{simple}$, I use all patients in the grand sample from 2007 to 2010 from hospitals with at least 50 HF patients in that year. Table A2 shows that the standard deviation of coding across hospitals is similar among the sample definitions.

Table 5 - Standard Deviation of Coding by Type of Hospital

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Patient Controls	None	Admission	Full	None	Admission	Full	Number of
Physician Controls		None			Physician Fixed Effects		Hospitals
Ownership							
Non-Profit	0.192	0.148	0.147	0.198	0.169	0.168	1,595
For-Profit	0.192	0.143	0.143	0.168	0.147	0.147	391
Government	0.221	0.162	0.161	0.176	0.148	0.148	400
Location							
Large Urban	0.192	0.145	0.145	0.176	0.147	0.146	1,007
Other Urban	0.184	0.143	0.143	0.183	0.157	0.156	842
Rural	0.229	0.170	0.170	0.213	0.191	0.192	537
Size							
Upper Tercile	0.174	0.137	0.137	0.149	0.131	0.130	793
Middle Tercile	0.185	0.141	0.141	0.177	0.145	0.143	794
Lower Tercile	0.228	0.169	0.168	0.231	0.202	0.201	799
Teaching Status							
Major Teaching	0.182	0.145	0.144	0.149	0.129	0.128	242
Minor Teaching	0.182	0.141	0.141	0.204	0.169	0.168	653
Non-Teaching	0.207	0.154	0.154	0.189	0.164	0.163	1,491
All Hospitals	0.199	0.152	0.151	0.190	0.163	0.162	2,386

Numbers in brackets count hospitals used to calculate the standard deviation. All results are adjusted for sampling variation. Columns 1 and 4 include no patient controls when calculating the hospital's coding score. Columns 2 and 5 control for patient characteristics observable upon admission. Columns 3 and 6 further add controls for histories of chronic conditions. Columns 4-6 control for physician fixed effects when calculating the hospital's coding score. See text for definitions of hospital characteristics.

the patient was admitted through the emergency department; and finely grained categories for the patient's primary diagnosis. The third patient approach, shown in columns 3 and 6, augments the second to also include indicators for a broad set of chronic conditions. The controls are described in greater detail in Appendix Section A.3. To improve comparability across analyses in this section, I limit to hospitals for whom all the characteristics and treatment performance metrics are observed.

The result of Table 5 is that across categories of hospitals, dispersion shrinks by about one-quarter moving from column 1 to column 2, as rich controls about the patient's hospital admission are added in the first step. Additionally controlling for patient illness histories in column 3 has little effect. The further addition of physician effects in column 6 does not systematically reduce variations, and it even raises dispersion slightly in the full cross-section of hospitals.

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 all patient controls included, the standard deviation falls to 0.15 (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 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 6). 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 raises dispersion for 9 groups and reduces it for 3 groups. Non-profit hospitals, for example, have a standard deviation in coding rates of 19 percentage points without any first-step controls, which falls to 15 with patient controls and rises to 17 with patient and physician controls. Likewise, the standard deviation of coding rates among major teaching hospitals falls from 18 percentage points with no controls to 14 with patient controls, and falls to 13 when physician controls are further added. These patterns are replicated in the other groups of hospitals: 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. Adding the physician controls removes the average physician component. When dispersion in adoption rises when these controls are added, it indicates that the average physician component was negatively correlated with the hospital component – evidence of negative assortative matching.

3.5 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 baseline hospital characteristics and clinical performance measures. 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.

3.5.1 Potential Roles of Hospital Characteristics and Clinical Performance

Size (Number of Beds) A long line of research has documented a relationship between hospital size and quality in many areas, though with an unclear causal link – this is called the volume-outcomes hypothesis, and Epstein (2002) provides a critical review. This relationship may be the result of increasing returns to scale or learning by doing, markets allocating patients to high quality facilities (which just happen to be large), or other factors correlated with hospital size that drive patient volume (Chandra et al., 2016a; Johnson, 2011).

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 that would more effectively prompt physicians for documentation and enable coding staff to transform documentation into the high-value codes. 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 The relationship between hospital ownership and coding straddles two broad strands of literature: one that investigates differences in the quality of care by ownership, and another that looks at ownership and the responsiveness to billing incentives. With respect to quality of care, there is no consensus on whether non-profit or for-profit hospitals are superior (McClellan and Staiger, 2000; Sloan, 2000). For-profit hospitals historically lagged public and non-profit facilities in the use of standards of care like β-blockers; more recent work looking at for-profit conversions finds little effect of moving from non-profit to for-profit ownership on clinical quality (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). A key difference between this setting and some earlier work (particularly Silverman and Skinner, 2004) is that it focused on upcoding, or the exaggeration of patient severity to raise payments. In contrast, achieving a high HF coding rate does not require a hospital to risk the fraud allegations that upcoding can bring. In theory, a hospital can provide a detailed HF code for all its HF patients with detailed documentation but no upcoding. Upcoding, in this context, would be submitting a detailed code when no supporting documentation existed.

Location Rsearch 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. At least some of this difference can be explained by rural hospitals being smaller, though research finds disparities both unconditional (MEDPAC, 2012; Baldwin et al., 2010) and conditional on other hospital characteristics including bed count (Goldman and Dudley, 2008).

Teaching Status Historically, teaching hospitals have been found to have better outcomes and higher quality processes of care than non-teaching hospitals (see Ayanian and Weissman, 2002 for a review). A more recent analysis that controls for hospital size, ownership, and other attributes shows that teaching intensity remains associated with greater patient survival rates, though the relationship between teaching intensity and process of care use was only monotonically positive for one of three conditions studied (Mueller et al., 2013). 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, the hospitals with the most revenue at stake could be safety net facilities – and these facilities may differ from the others in managerial quality and physician control.

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 pushing them to treat patients well.

3.5.2 Results

Table 6 displays the key estimates of the role of hospital characteristics and clinical performance in explaining takeup of the coding technology. 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 and are described in full detail in Appendix Section A.3. 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 patient controls, but no physician controls. Column 1, which includes no patient-level adjustments, shows how the raw probability a HF patient at the hospital is billed with a detailed code depends on hospital characteristics. However, these relationships could depend on some hospitals having patients that are harder or easier (or more worthwhile) to code. To address this concern, the next two columns add patient-level risk adjusters. Column 2 shows how hospital characteristics are correlated with the probability that the hospital uses a specific code for a HF patient, first adjusting these probabilities to remove the effect of age, race, sex, source of admission, and main diagnosis. Column 3 adds adjustments for patients' chronic conditions.

There is a robust relationship between coding and hospital size, non-profit status, teaching status, use of standards of care, and heart attack survival. Hospitals that are 10% larger give 0.19 percentage points more of their HF patients a specific code. Adding patient controls when estimating hospital adoption reduces this effect to 0.12 percentage points – 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. Likewise, non-profit hospitals give 3.8 percentage points more of their patients a specific code than government-run facilities, though adding patient controls reduces the difference to 2.9 percentage points. There is no significant difference between the takeup rates of for-profit and government-run hospitals, nor do urban hospitals statistically significantly outperform their rural peers. Major teaching hospitals are significantly more likely to provide

Table 6 - Describing the Distribution of Coding with Hospital Characteristics and Productivity

	(1)	(2)	(3)	(4)	(5)	(6)
Patient Controls	None	Admission	Full	None	Admission	Full
Physician Controls		None			Physician Fixed Effects	
Hospital Characteristics (C_h)						
ln(Beds)	0.019**	0.012**	0.012**	-0.001	-0.008	-0.008
	(0.008)	(0.006)	(0.006)	(0.012)	(0.011)	(0.010)
Non-Profit	0.038**	0.029**	0.029**	0.023	0.023	0.021
Ownership	(0.016)	(0.012)	(0.012)	(0.016)	(0.014)	(0.014)
For-Profit	0.009	0.009	0.008	0.029	0.027	0.026
Ownership	(0.017)	(0.012)	(0.012)	(0.020)	(0.017)	(0.017)
Located in Large	0.003	0.000	-0.001	0.054**	0.035*	0.034*
Urban Area	(0.016)	(0.012)	(0.012)	(0.023)	(0.020)	(0.020)
Located in Other	0.019	0.011	0.010	0.060***	0.035**	0.034**
Urban Area	(0.014)	(0.011)	(0.011)	(0.019)	(0.017)	(0.017)
Major Teaching	0.033*	0.035**	0.035**	0.026	0.055***	0.054***
Hospital	(0.018)	(0.014)	(0.014)	(0.022)	(0.019)	(0.019)
Minor Teaching	0.001	0.005	0.005	-0.008	0.006	0.005
Hospital	(0.010)	(0.008)	(0.008)	(0.015)	(0.013)	(0.013)
Ex Ante $ at Stake	0.000	0.000	0.000	0.000	0.000	0.000
per Patient	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Standards of Care and Clinical Performance (Z_h)						
Standards of Care	0.030***	0.022***	0.022***	0.023***	0.019***	0.019***
Composite Z-Score	(0.005)	(0.004)	(0.004)	(0.007)	(0.006)	(0.006)
Heart Attack Adj 30-	0.029***	0.024***	0.024***	0.030***	0.029***	0.029***
Day Survival Z-Score	(0.006)	(0.005)	(0.005)	(0.008)	(0.007)	(0.007)
Observations	2,386	2,386	2,386	2,386	2,386	2,386
R^2 (adjusted)	0.095	0.096	0.094	0.035	0.034	0.033

Standard errors clustered at the market level in parentheses. Columns 1 and 4 include no patient controls when calculating the hospital's coding score. Columns 2 and 5 control for patient characteristics observable upon admission. Columns 3 and 6 further add controls for histories of chronic conditions. Columns 4-6 control for physician fixed effects when calculating the hospital's coding score. See text for definitions of hospital characteristics and performance measures.
*** significant at 1% level; ** significant at 5% level; * significant at 10% level

detailed codes than non-teaching facilities, a difference of 3.5 percentage points with patient controls.

Finally, with the full set of patient controls, for each standard deviation rise in the use of standards of care , about 2.2 percentage points more HF patients tend to get a specific code. The effect for each standard deviation rise in heart attack survival is 2.4 percentage points. In other words, hospitals that appear to be higher quality in their treatment are also more likely to use these high-revenue billing codes.

With Physician Controls Columns 4 to 6 repeat the results of columns 1 to 3 with first-step physician controls, changing the interpretation of the coefficients. In these columns, a positive (negative) relationship between a hospital characteristic and coding indicates that the facility was able to extract more (less) detailed coding out of its physicians – the hospital effect on the left-hand side of these regressions conditions on the physicians that treated the patients. I focus on the coefficients of column 6, which adjust for the full set of patient characteristics of column 3 as well as the physician when estimating the hospital component of adoption.

The gradient between hospital size and extraction of detailed HF codes is positive and significant without first-step physician controls, but it is eliminated and its point estimate even becomes negative when the physician component of adoption is swept away. 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 2.1 and 2.6 percentage points, respectively, more likely to extract specific codes from doctors than their government-run counterparts, though these coefficients were imprecisely measured. Compared to the same differential calculated unconditional on physicians – the result in column 3 – the value for non-profit hospitals is reduced by about one-quarter and no longer significant. Since removing the physician component of adoption reduces the coding advantage of non-profit facilities, my results imply that the physicians who work at non-profit hospitals are more likely to provide the detailed documentation wherever they practice. On the other hand, the value for for-profit facilities expands to a practically significant level, 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. The net result is that it is only unconditional on physicians that government and for-profit hospitals are about equally likely to use the detailed billing codes – the finding in columns 1 to 3.

Hospitals in large urban and other urban areas – areas of high and intermediate population density, respectively – extract specific codes from their doctors for 3.4 percentage points more of their patients than hospitals in rural areas. This relationship does not exist without the physician controls, which indicates that urban hospitals, like for-profit hospitals, are effective at extracting the codes but are held back by their physicians. For major teaching facilities, the gradient expands from 3.5 to 5.4 percentage points with the removal of the physician component, suggesting that physicians also hold back these hospitals but the facility effect is so big that they outperform their non-teaching peers even unconditional on the physician.

Finally, the use of detailed HF codes is correlated with both heart attack survival and the use of consensus standards of care: even if all hospitals had the same kinds of patients and doctors, hospitals with one standard deviation greater use of standards of care or one standard deviation greater treatment performance would use specific codes for 1.9 and 2.9 percentage points (respectively) more of their patients. This 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 able to extract the codes from their physicians than their lower treatment quality peers.

4 Discussion

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. Conformance with detailed coding across hospitals 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 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 exceptional on the dimension of technology adoption – this finding is nearly universal in cases of new technology.

The hallmark features of a new technology 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, a growing literature has found persistent dispersion in productivity within narrowly defined industries (Fox and Smeets, 2011; Syverson, 2011); this literature is now expanding to include the health care sector (Chandra et al., 2016b). I have shown that adoption of the HF coding technology across hospitals follows the established pattern.

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 adapt 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 context, 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, most physicians are paid for each procedure they perform. Though hospitals might want to encourage detailed coding by paying doctors for it, doing so runs afoul of federal laws that prohibit directly incentivizing physicians by basing their payments on the hospital's payment, a practice commonly known as gainsharing (see HHS, 1999). In a sense, the principal-agent problem in patient documentation is written into the law.

Some hospitals may be very detailed coders because their doctors are likely to provide specific documentation wherever they practice. 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 physician 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 than other hospitals.

Since hospitals but not physicians were paid for the HF documentation, I have argued that the hospital component of adoption is related to whether the hospital was able to solve a principal-agent problem. This component is robustly correlated with heart attack survival and the use of consensus standards of care when treating patients. Thus hospitals that use and achieve high clinical quality are able to extract more specific documentation from a fixed set of physicians than other hospitals. The correlation between these two measures suggests that agency problems could play a role in the adoption of a variety of technologies in the facility. Another view of this correlation is that revenue and treatment performance are positively related.

The dispersion that I find in the hospital component of adoption, which removes the physician and patient components, is about four-fifths the raw level of dispersion. This residual dispersion has a standard deviation of 16 percentage points. One point of comparison is the standard deviation of the consensus standards of care scores, which measure adherence to evidence-based treatment guidelines. The measures of coding of HF and standards of care are both hospital-level shares. To the extent that there are substantial disparities across hospitals in their adherence to these 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.

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 policy elastic to financial incentives for other processes of care. For the policy elasticity, 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, and higher treatment performance hospitals are more policy elastic.

One reason to incentivize the use of evidence-based inexpensive medical technologies 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 it is an open question whether these policies will have their intended effect of raising quality; according to these findings, policy elastic providers tend to be getting better results from treatment and 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.

5 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 improving the documentation of patients in their claims. Yet this improvement in documentation was incomplete and uneven, a characteristic feature of the adoption of new technologies. 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 hospitals that had high treatment performance and followed consensus standards of care were better able to extract detailed documentation from their physicians. I argue that this is consistent with these hospitals solving principal-agent problems.

My results have important policy implications as public and private insurers seek to directly raise hospital productivity by reforming health care payment systems. Principal-agent problems owing to a bifurcated system that pays doctors and hospitals on separate bases may be major impediments to further productivity-raising technology adoption. For example, when Medicare opts to pay hospitals to use evidence-based clinical practices like giving heart attack 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 to reform the physician payment system. Provisions of the Affordable Care Act that require this system to incentivize standards of care, much as Medicare is already doing for hospital payments, are one way forward. By bringing these incentives to both hospitals and doctors, these provisions may substantially improve the effectiveness of value-based payment reforms.

My results do not preclude other frictions from driving the disparities in adoption. Some hospitals may have outdated billing software that misses opportunities to use high-value diagnosis codes even when physicians document them. Poorly trained hospital billing staff could similarly fail to capture the revenue from these codes. 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

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.

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.

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.

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.

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.

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.

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, 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.

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.

HHS. 1999. "Gainsharing Arrangements and CMPs for Hospital Payments to Physicians to Reduce or Limit Services to Beneficiaries." July, URL: https://oig.hhs.gov/fraud/docs/alertsandbulletins/gainsh.htm.

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.

Johnson, Erin. 2011. "Ability, Learning, and the Career Path of Cardiac Specialists." URL: http://www.mit.edu/~erinmj/files/Draft_042011.pdf, Mimeo.

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.

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

Massie, Barry M., and Nihir B. Shah. 1997. "Evolving Trends in the Epidemiologic Factors of Heart Failure: Rationale for Preventive Strategies and Comprehensive Disease Management." *American Heart Journal*, 133(6): 703–12.

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

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.

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.

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. 2000. "Not-for-Profit Ownership and Hospital Behavior." In *Handbook of Health Economics.* eds. by Anthony J. Culyer, and Joseph P. Newhouse, Volume 1: Elsevier, 1141–1174.

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

Appendix To:

Technological Diffusion Across Hospitals: The Case of a Revenue-Generating Practice

Adam Sacarny

July 2016

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 the figure 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.3).

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 4 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 A2 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 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 heart attack, heart failure, pneumonia, and surgery standards of care in 2006.

The heart attack 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 Heart Attack Survival

I construct adjusted heart attack survival by starting with a sample of all heart attack 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 heart attack 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.[13] 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.

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 First-Step Specifications

I use six specifications to estimate the hospital fixed effects using the analysis sample described in Section 2.6 of the main text. The regressions are of the form of equation 1. Three of these

[13] 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.

specifications use no physician fixed effects while three of them include physician fixed effects. Within each three, one specification uses no patient controls. The next 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.

Appendix Figures

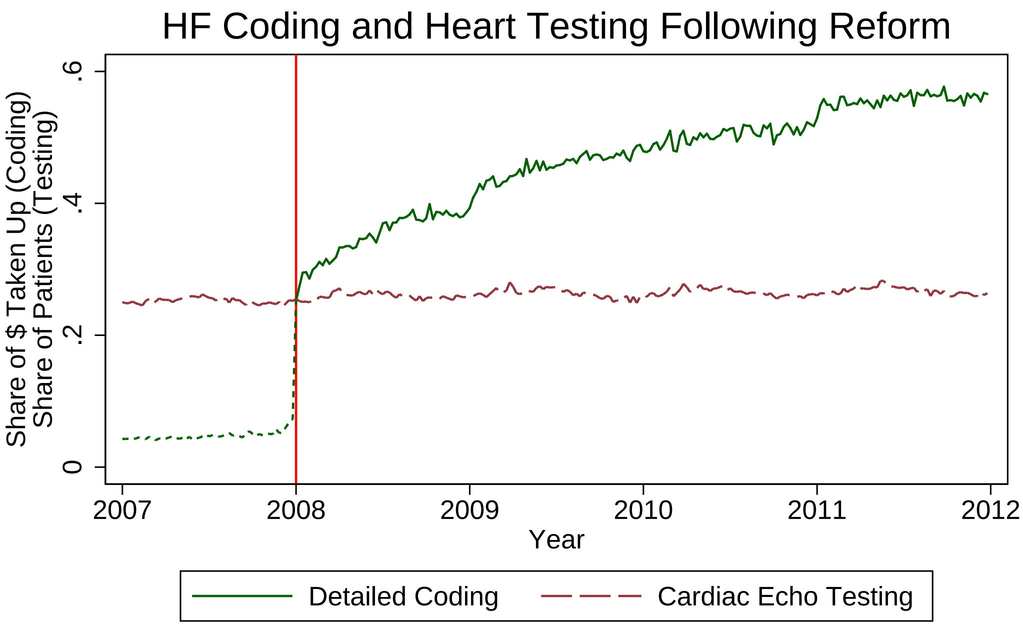
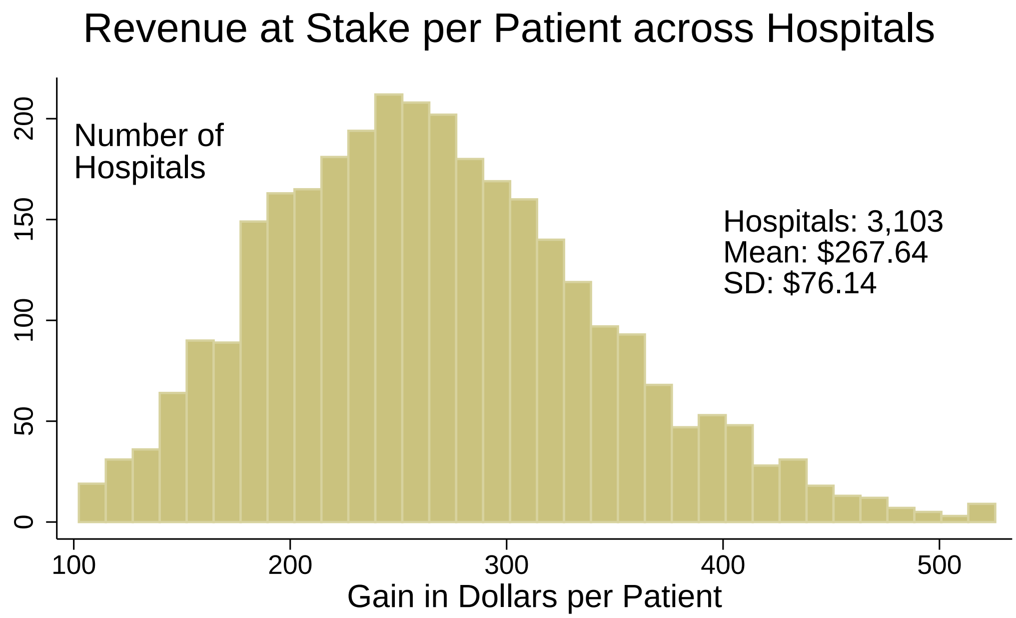


Figure plots the weekly share of revenue available for detailed coding of HF that was captured by hospitals alongside the weekly share of all patients who received a cardiac echo, a heart test. The dotted line shows revenue that would have been captured in 2007 if hospitals had been paid per 2008 rules.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

Figure A1



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 A2

Appendix Tables

Table A1 - Other Codes for HF

Code	Description	Severity Before	Severity After	Counted as 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

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 - 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)	2,386	0.546	0.201	0.199

Table presents 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 Section 2.3; 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 2.6. The step 2 analysis sample is the subset of these hospitals for which all characteristics are observed, and is the focus of Sections 3.4 and 3.5.