Essays in Health Economics and Productivity

by

Adam Jon Sacarny

B.A. Economics, Columbia University (2007)

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2014

Signature redacted

Author ....

Department of Economics
May 15, 2014

Signature redacted

Certified by ....

Amy Finkelstein
Ford Professor of Economics
Thesis Supervisor

Signature redacted

Certified by ....

Michael Greenstone
3M Professor of Economics
Thesis Supervisor

Signature redacted

Accepted by ....

Michael Greenstone
3M Professor of Economics
Chairman, Departmental Committee on Graduate Studies

Essays in Health Economics and Productivity

by

Adam Jon Sacarny

Submitted to the Department of Economics
on May 15, 2014, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

The first chapter studies how incentives drive adoption by looking at 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. The key barrier to takeup 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. My results suggest that agency conflicts may drive disparities in health care performance more generally.

The second chapter (co-authored with Amitabh Chandra, Amy Finkelstein, and Chad Syverson) challenges the conventional wisdom in health economics that large differences in average productivity across hospitals are the result of idiosyncratic, institutional features of the healthcare sector which dull the role of market forces. Strikingly, we find that productivity dispersion in heart attack treatment across hospitals is, if anything, smaller than in narrowly defined manufacturing industries such as ready-mixed concrete. We also find evidence against the conventional wisdom that the healthcare sector does not operate like an industry subject to standard market forces. In particular, we find that hospitals that are more productive at treating heart attacks have higher market shares at a point in time and are more likely to expand over time. These facts suggest that the healthcare sector may have more in common with "traditional" sectors than is often assumed.

The third chapter explores whether hospitals change their treatment decisions when they are are paid more for certain treatment approaches. I exploit a Medicare reform that altered payment rates depending on whether patients were relatively healthy or sick. Looking at three treatment approaches for lung cancer patients, I demonstrate economically significant own-price elasticities and right-signed cross-price elasticities – though these estimates sometimes lack statistical power and should be interpreted with caution due to concerns about endogeneity. These findings indicate that payment reforms, including movements toward capitation and away from fee-for-service, may have large effects on the intensity of care that patients receive in the hospital.

Thesis Supervisor: Amy Finkelstein
Title: Ford Professor of Economics

Thesis Supervisor: Michael Greenstone
Title: 3M Professor of Economics

Contents

Acknowledgements

I am deeply thankful to have been a part of the MIT community during the last five years. As a student here I have crossed paths with countless brilliant, caring, and thoughtful people. These wonderful individuals turned into friends, colleagues, and mentors. I am lucky to have had their support.

Many of my classmates rave about their advisors, but I am pretty sure my advisors were the best. Amy Finkelstein took even my worst research ideas seriously – then, to my horror, she pushed me to make them better. Amy demanded that I do great work even when I thought I wasn't capable of it. The process sometimes wasn't pretty, but when my results blew up, she always managed to talk me down. I am thankful that she managed to provide this support even while slowly and deliberately loosening the bolts on the training wheels. In future research I look forward to better understanding the sources of Amy's productivity.

I vividly recall the terror I felt during my first lunch seminar as Michael Greenstone began to ask me a question (and I still remember his point because it was so crucial). That fear subsided over the years, but only a little bit, which kept our meetings interesting. Michael was direct in his criticism because he wanted my research to be awesome, and I know that because he said the word 'awesome' all the time. When he said it about my own work, I knew I had accomplished something.

Jon Gruber met with me throughout my time as a student, and he always managed to make me more excited about my research and career. His advice about my ideas was as invaluable as his guidance on the job market. Many other faculty took an interest in my work and for this I am grateful. I was particularly lucky that Jon Skinner volunteered himself as a mentor. Jon took several panicked phone calls from me while I was on the job market. By the end of the calls I always felt more confident about my papers and my career path. Paulo Somaini was willing to discuss my research with no time limit whatsoever; our meetings usually ended in a structural model. Heidi Williams had the same generosity with her time; our conversations usually ended with a great potential dataset or identification strategy. I am indebted to Bob Gibbons for introducing me to new perspectives on organizations and productivity. Amitabh Chandra was a terrific co-author and his perspective on health economics never ceased to blow my mind. Ray Fisman gave me my first research assistance job and has kept in touch ever since – I am deeply lucky that Ray needed 20 hours per week of work in the summer of 2005 and felt inclined to give that job to me.

John Arditi, Cherisse Haakonsen, and Deb Jamiol provided extraordinary help as I navigated the job market. They never complained about my countless last-minute requests, though they certainly had the right to. As far as I am concerned, Gary King and Beata Shuster keep the world revolving on its axis; at the bare minimum they keep the department running in the most sincerely friendly way possible. I benefitted immensely from the Health and Aging pre-doctoral training program at the NBER, and I am thankful for the terrific work that Janet Stein and Emily Weissner perform administering it.

My classmates were awesome and I am grateful to have met every last one of them. I am especially thankful for the support I received from our weekly thesis group. Ben Feigenberg, Eliza Forsythe, Conrad Miller, Miikka Rokkanen, Annalisa Scognamiglio, and Henry Swift all listened to my rambling and in the process helped me turn my research ideas into real papers. Isaiah Andrews, Anil Jain, Maria Polyakova, and Brad Shapiro all gave lots of advice and help even when I was down on my work.

I spent much of my first few years at MIT bothering Jason Abaluck, Greg Leiserson, David Molitor, Iuliana Pascu, Mike Powell, Sarah Venables, and Xiao Yu Wang, often in the grad lab. When they escaped to their own offices, I figured out where they sat and bothered them there too. They were kind to put up with me even though they had job market papers to work on. Arun Chandrasekhar, Jenn Peck, and Maxim Pinkovskiy were old friends who received me with open arms when I got to MIT – I also bothered them while they tried to work on their job market papers and they were equally cool about putting up with me.

Even though Manasi Deshpande, Kate Easterbrook, Sally Hudson, Sarah Moshary, Hoai-Luu Nguyen, Melanie Wasserman, and Nils Wernerfelt came to MIT a year too late, they all provided help and support far beyond their years. Gustav Pascu and Star Shapiro both provided help and support beyond their cat and dog years, respectively. Leah Berkenwald, Ethan Feuer, and Paul Minnice absorbed the instant messaging energy that my classmates could not.

I have been a part of the MIT community for almost five years but my parents have loved me unceasingly for almost thirty (my brother Ian comes in a close third at almost twenty-seven). At every turn in my academic and personal life they insisted that I could and would succeed, a fact that was too often lost on me – until I doubted myself, when it became the most important thing in the world. They provided all of this support while limiting their demands that I call my grandparents to one request per conversation. Ian provided the same support except he took calls until 2am and was more adept at responding to text messages.

These acknowledgements omit too many family members and friends around Boston, New York, and spread throughout the country. I am extraordinarily lucky to have such wonderful people – those both listed and unlisted here – in my life. This work would not have been possible without their love and care.

List of Tables

List of Figures

Chapter 1

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

1.1 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, 2013b; Bloom et al., 2012b). 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

*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, 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 all aspects of the Medicare data. I gratefully acknowledge funding from the National Institute on Aging grant T32-AG000186.

hospitals, this study examines the role of financial incentives and agency conflicts in the adoption of new technologies.

These incentives are particularly important in the wake of the Affordable Care Act, which mandates that public insurers use their purchasing power to raise the productivity of health care providers.[1] In designing new payment schemes, 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, 2009a; Baicker and Chandra, 2004b; 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[2] changed the pricing function of Medicare to begin providing additional payments for the detailed codes. 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 affect the pay of physicians.

Figure 1-1 shows that the change in incentives triggered a rapid but incomplete response by hospitals: in just the weeks following the reform, hospitals started capturing 30% of the revenue made available; by the end of 2011 they were capturing about 55%. 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 recent work that has demonstrated large differences in productivity across seemingly similar firms (Fox and Smeets, 2011a; Doms and Bartelsman, 2000; Syverson, 2011b), 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, 54% 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 84% of its patients.

[1]The most prominent example of a policy resulting from that provision of the act is Medicare's Value-Based Purchasing Program, which rewards hospitals that adopt evidence-based standards of care and perform well in surveys of patient satisfaction.

[2]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.

These findings suggest that hospitals were aware of the financial incentive to use the detailed codes, but that this awareness was tempered by significant frictions. I 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, with few exceptions, hospitals are not allowed to give a share of their Medicare payments to physicians as incentive pay. In spite of these restrictions, new policies to improve the quality of care have focused on the hospital's payment 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 physician effects. 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 novel application of a labor economics technique that has been frequently used in the context of workers and firms (see e.g. Abowd et al., 1999; Card et al., 2013) but has rarely been applied in studies of health.

Sweeping out the physician contribution removes dispersion in adoption due to hospitals having different kinds of doctors. This procedure addresses the concern that doctors who work at some hospitals may be more willing to provide the details 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 0.15 percentage points with rich patient controls to 0.16 percentage points with patient and physician controls.[3] 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 – with physician effects removed – and hospital characteristics like size, ownership, location, and productivity. The signs of these relationships are not *ex*

[3]When there is negative assortative matching between hospitals and physicians, dispersion in adoption can rise when the physician component is removed.

ante obvious, and they indicate which types of hospitals were most able to extract the codes from their doctors. The most robust finding of this analysis is that adoption was greater among hospitals that were higher quality by two measures: heart attack treatment productivity (the survival rate of heart attack patients after adjusting 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.

In an additional exercise, I look at correlates of hospital-level adoption but do not remove the component of takeup that is due to the hospital's physicians. These results indicate which types of hospitals are the most "policy-elastic" with respect to financial incentives. Adoption is strongly correlated with hospital size, ownership, and productivity. Large and non-profit facilities were more likely to adopt, as were facilities that complied with consensus survival-raising standards of care. Hospital responses were also positively correlated with heart attack treatment productivity. This result 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 novel 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 uniquely plausible that none might exist. These disparities reflect differences in hospitals' basic ability to respond to incentives. Second, using decomposition techniques that are normally associated with labor economics, I show that variations in adoption are largely driven by the ability of some hospitals to extract more high-revenue codes from their doctors than others – disparities persist when the physician component of adoption is removed. 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 1.2 discusses the heart failure billing reform, the data I use to study it, and provides a simple analytical framework. Section 1.3 presents results on dispersion in hospital takeup, then shows how takeup relates to hospital characteristics and measures of treatment productivity. Section 1.4 provides a discussion of the results. Section 1.5 concludes.

1.2 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). 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 most economic literature on health care eschews studying HF in favor of less common conditions like heart attacks. The literature has focused on these conditions because they are thought to be sensitive to treatment quality and are well observed in most administrative data. Since this paper concerns how hospitals learn to improve their billing practices, not the effect of treatment on health, issues like endogenous selection of patients on unobserved determinants of survival are not the principal potential confounders. Rather, the great deal of revenue at stake for heart failure reimbursement makes it a condition that is well suited for this study's aim of understanding how hospitals respond to coding incentives.

My analyses focus 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 $111 billion program that pays for most Medicare beneficiaries who are admitted as inpatients to most hospitals in the United States (MEDPAC, 2012a). As part of a 2008 overhaul of the IPPS – the most significant change to the program since its inception – the relative payment for vaguely documented and specifically documented HF was changed. This element of the reform made the documentation valuable and provided the financial incentive for the spread of the technology.

1.2.1 Payment Reform for 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 3 levels of severity (low, medium, or high) where there had been 2 (low or high), shuffling the severity level of the many heart failure codes in the process.

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; Dafny, 2005a studies how hospitals exaggerate their reporting of patient severity due to incentives; Song et al., 2010 studies how reporting varies across regions). 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 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 were found to predict treatment cost and, being specifically identified illnesses, were medically consistent with the agency's definitions of medium and high severity. 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 1.1.

The detailed codes were exhaustive over the types of heart failure, so with the right documentation, a hospital could continue to raise nearly any HF patient 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 coordination between physicians and hospital staff. In this way it is similar to other technologies that have come into the focus of researchers and policymakers recently, including the use of β-blockers (an inexpensive class of drugs that have been shown to raise survival following a heart attack) in health care and the implementation of best managerial practices in firms.

For a hospital to legally submit a detailed code, a doctor must state the details about the HF in the patient's medical chart.[4] Figure 1-2 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 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.

[4]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 on payers like Medicare. 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.

1.2.2 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 specific code in the year before the reform.

Following the reform, the gain was always weakly positive and could be as high as tens of thousands of dollars; the exact amount depended on the patient's main diagnosis and whether the patient had other medium- or high-severity conditions. For patients with other medium-severity conditions, hospitals could gain revenue if they could find documentation of a high-severity form of HF. For patients with other high-severity conditions, finding evidence of high-severity HF would not change Medicare payments. However, 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.

In 2009, 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).[5] 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. The evolution of the gain to specific coding is shown in Figure 1-3 and the corresponding takeup in the use of these codes is shown in Figure 1-4.

For each hospital, the gain to taking up the revenue-raising technology – the money put on the table by 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 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]

Figures 1-5 and 1-6 show 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 (or, spreading this gain

[5]These averages include the patients for whom the detailed codes do not raise payments because, for example, they already had another medium- or high-severity condition. 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). The gains to specific HF codes were calculating by reprocessing all HF patients, replacing the observed HF codes with only vague, only chronic, and only acute HF codes.

[6]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, whether admitted through the emergency department, 19 chronic conditions, and 25 major diagnostic categories classifying the underlying cause of admission). I use this regression to fit the probability that patients in 2007 would have received a medium- or high-severity HF code, then re-price these patients under the 2009 post-reform pricing rules. The result of this procedure is an *ex ante* expected gain to using the detailed codes, which I aggregate to the hospital level.

across all admissions, \$268 per patient) in 2009 by giving all of its HF patients specific HF codes rather than vague ones. The standard deviation of the revenue at stake per HF patient was \$230 (the standard deviation of the gain spread over all patients was \$76). To provide a sense of scale, one can consider these amounts relative to hospital operating margins. The 2010 Medicare inpatient margin, which equals hospitals' aggregate inpatient Medicare revenues less costs, divided by revenues, was -1.7% (MEDPAC, 2012a). 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.

1.2.3 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}:

$$code_{ph} = \alpha_h + \alpha_d + X_p\beta + \epsilon_{ph} \tag{1.1}$$

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 wide literature has documented persistent productivity differentials in the manufacturing sector (see Syverson, 2011b for a review), and work is ongoing to develop documentation of similar facts in the service and health care sectors (Fox and Smeets, 2011a; Chandra et al., 2013). In this framework, 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. Dispersion in hospital types is therefore a form of productivity dispersion.

What might drive this dispersion? Recall that 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.

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 (defined as the number of beds in the facility), ownership (non-profit, for-profit, or government-run), location (whether the hospital is in a large urban area, other urban area, or rural area), teaching status (whether it has residents), and *ex-ante* per-patient revenue put at stake by the reform. The second set, called Z_h, includes measures of the hospital's productivity – the amount of survival the hospital can generate for a fixed amount of inputs.

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

$$\alpha_h = \gamma + C_h\rho + Z_h\theta + \eta_h \tag{1.2}$$

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 1.3.

1.2.4 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 fee-for-service system. This file is essentially a copy of all the reimbursement claims that hospitals sent Medicare. For 92% of these stays, I can identify 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.[7] Since physicians are paid for each procedure they perform, for these stays I can also identify echocardiograms and other heart tests.

[7] 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 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.

I use these data to construct an analysis sample of hospitals' claims to Medicare for their HF patients. Starting with all patients in 2010, I eliminate those who lacked full Medicare coverage at any point during their hospital stay, were covered by a private plan, or were under age 65. 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 the approximately 3% of 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 2% of stays at Maryland hospitals (which are exempt from the IPPS). I then limit the sample to heart failure patients, which are identified as those with a principal or secondary diagnosis ICD-9 code of 428.x, 398.91, 402.x1, 404.x1, or 404.x3.

This analysis sample is described in Table 1.2 and includes 1.9 million HF patients. There are about 3,400 distinct hospitals in the full analysis sample, which is the set of claims that is used to estimate hospital types when physician effects are not being swept out. Most hospitals see a large number of HF patients in a given year: the average treats 553 of them, and even the 10th percentile includes 52 of them.

By using the 92% of patients for whom I observe the physician, I am able to identify hospital effects after controlling for the doctor. Hospital and physician 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 1.3.2). I call the mobility sample the set of patient claims that occur within the largest mobility group of hospitals and physicians. There are about 2,900 hospitals and 135,000 doctors in the sample. The average mobility sample hospital sees 582 HF patients in 2010 and its HF patients are treated by 58 distinct doctors. At the average hospital, 20 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 1.3 provides additional information about the doctors by mobility status. The average mobile physician treats about twice as many patients as a non-mobile physician. Information on physician specialty, demographics, training, and experience comes from the AMA Masterfile; specialties are grouped according to the Dartmouth Atlas definitions.[8] Mobile physicians are 11pp less likely to be surgeons and are correspondingly more likely to be primary physicians like internists or medical specialists like cardiologists. Mobile physicians also have about 8 months more training – but about 8 months less experience practicing since completing training – than their non-mobile counterparts.

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.

[8] See Table 2 of the document found at http://www.dartmouthatlas.org/downloads/methods/research_methods.pdf

1.2.5 Costs of Takeup

Figure 1-1 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 2011. The finding of incomplete takeup raises the question of what costs must be incurred by the hospital to adopt the technology.

One possibility is that taking up the reform requires medical testing of HF patients to confirm the details of their conditions. For example, 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 true, one could interpret the reform as encouraging higher intensity medicine – and the costs and principal-agent frictions as the refusal of doctors and hospital staff to order tests that they had not already thought necessary.

Contrary to this story, both the time series evidence and the official coding guidelines show that whatever were the costs of more detailed HF coding, they were not realized through changes in real medical treatment. Figure 1-7 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. This finding is sensible considering that 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 the official 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.

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 many 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. For example, a doctor's first-order concern may be documenting aspects of the patient that are crucial for proper post-acute care, making documentation that matters solely for the hospital's billing a secondary issue.

Taking up the revenue-generating technology 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).

1.3 Hospital Adoption

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."

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

The adoption of the coding technology was incomplete at the national level, but the national time series masks enormous heterogeneity at the level of the hospital. In this section, I construct and present measures of adoption of detailed coding across hospitals. These measures have wide dispersion, with some hospitals almost never using specific codes and other hospitals almost always using them. A perhaps natural view is that 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, and persistent differences in productivity have been found in nearly every sector in which they have been studied.

I present a novel 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 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 shows that dispersion in the adoption of detailed HF coding persists even among observably similar hospitals. The dispersion result is robust to sweeping out the physician component of coding – even if hospitals had the same doctors, there would still be coding disparities. Equivalently, the probability a HF patient treated by a particular doctor gets a specific code is significantly greater at some hospitals relative to others.

Second, it explores the relationship between adoption and hospital characteristics like size, ownership, and 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, not causal, they are useful policy inputs: they can be interpreted as indications of which providers are most elastic to incentives for revenue generating technologies.

1.3.1 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 1.1. 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 productivity (to see which facilities are most likely to adopt).

1.3.1.1 First Step: Estimating Hospital Fixed Effects

In the first step, I run the regression given in equation 1.1. I consider versions of this regression with patient controls of varying degrees of richness, and run these regressions both with and without physician fixed effects. I then extract estimates of the hospital fixed effects $\hat{\alpha}_h$. These estimates equal the share of HF patients at the hospital who received a specific code ($\overline{code_h}$) less the contribution of the hospital's average patient ($\overline{X}_h\hat{\beta}$) and the patient-weighted average physician effect ($\frac{1}{N_h}\sum_{p\in P_h}\hat{\alpha}_{d(p)}$, where N_h is the number of HF patients at the hospital, P_h indexes the patients, and $d(p)$ indicates the doctor that attended to patient p):

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

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

$$\hat{\alpha}_h^{simple} = \overline{code}_h \tag{1.3}$$

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 an endogenous component of the hospital's 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 principal-agent problems.

1.3.1.2 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 other coefficients in the model are at the patient level, and are estimated from millions of observations. These coefficients are estimated quite precisely, reducing the role for this noise.

When the specification includes physician fixed effects, the imprecision of the hospital effect reflects imprecision in the estimates of the physician effects. 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.

Estimates of the variance of α_h must account for measurement error in order to avoid overstating dispersion. To produce these estimates, I adopt part of the Empirical Bayes procedure described in Appendix C of Chandra et al. (2013). 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).

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 1.2. 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 productivity measures that are found on the right-hand side of the key regressions, and it does not bias the estimates of ρ or θ.

1.3.2 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 (called a node in graph theory). In the graph, a doctor and hospital have a line (called 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 (called a path) going from one to the other in the graph. A mobility group starts with a doctor or hospital and includes all other doctors and hospitals that are connected to her or it. In the graph of doctors and hospitals, a mobility group is called a maximal connected subgraph. Among the 3,414 hospitals in the analysis sample, the largest mobility group contains 2,868 hospitals.

The econometric model of the first step follows from certain identification assumptions. 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. α_{hd}), 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.[9]

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 patients would be estimated with negative bias. I test this assumption by including increasingly rich sets of patient characteristics as controls. The key results on the characteristics of the adopters and the dispersion in adoption are somewhat sensitive to these controls. Specifically, the significant coefficients 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 1.3.4). These coefficients remain highly significant even though they attenuate. Likewise, the 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.

It is perhaps unsurprising that patient characteristics influence the hospital's use of the codes. The fact

[9] 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.

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 quite detailed in the claims data that I use. While the great majority of disparities in adoption across hospitals cannot be attributed to anything observable about the patient, I present all results in this study under three patient-level specifications to be clear about this potential source of endogeneity.

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. For example, a hospital might demand less specificity in HF coding from physicians who were friendly with its owners. These physicians would have negative match effects with that hospital. If they tended to practice at the hospital, the negative match effects would load onto the hospital effect, biasing it downward. The role of match-specific synergies is also bounded by the match effects model described in footnote 9 – 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.

1.3.3 Hospital Characteristics

Table 1.4 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 2,411 hospitals, and includes any facility with a heart failure coding score and which had complete information on all baseline characteristics, standards of care, and productivity.

The rows of the table comprise the key hospital characteristics and productivity measures that are used in the analyses. Hospital size (beds) and ownership are taken from the Medicare Provider of Services file. Ownership may be non-profit (about two-thirds of hospitals), for-profit (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. The hospitals in this sample are found in all three areas, though the number of rural hospitals is reduced because many were classified as critical-access facilities, which were exempt from this reform. Teaching hospitals, which comprise just over one-third of facilities, are defined as those with a resident-to-bed ratio greater than zero.

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. Since most patients were coded vaguely in 2007, this variable is constructed by filling in each 2007 patient's specific HF code using the relationship between well-observed patient characteristics and specific HF codes at hospitals that were excellent at coding in 2010 (see footnote 6 for more information). To improve precision and reduce the leverage of outliers, 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.

Heart attack treatment productivity is constructed using the sample and methods of Chandra et al. (2013). A hospital's treatment productivity is the average log-survival 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. Raising hospital productivity by 10% means that, at the same level of inputs, the hospital is able to produce 10% more survival-days for its patients. Productivity is adjusted to account for measurement error using an Empirical Bayes shrinkage procedure described in more detail in Appendix C of Chandra et al., 2013.

The standards of care measures were collected by CMS under its Hospital Compare program. 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.[10] These standards of care are inexpensive, evidence-based treatments that have been shown to improve patient outcomes. 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 that has been the subject of several economic studies (see e.g. Skinner and Staiger, 2009a, 2007b).

1.3.4 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 1-8 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 2003 to 2010. Takeup across hospitals moved rapidly after the reform. By 2010, the median hospital used specific codes 55% of the time. Figure 1-9 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 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. 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. All results shown in tables make adjustments for excess dispersion due to sampling variance, however.

[10]The processes of care included in the measures were chosen by CMS based on medical evidence. The heart attack measure includes prescription rates of β-blockers and aspirin for appropriate patients as well as 5 other processes of care. The heart failure measure includes an evaluation of left ventricular systolic function (a key input to determining the part of the cardiac cycle that is weakened) and 3 other processes of care. The pneumonia measure includes prompt prescription of antibiotics and 6 other processes of care, and the surgery measure includes antibiotics and 2 other processes of care.

Table 1.5 shows the standard deviation of adoption among observably similar hospitals. I divide the space of hospitals into 7 mutually exclusive and exhaustive groups 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 sweeps out the physician component.

The first patient control specification, presented in columns (1) and (4), includes no patient-level controls at all. 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 the patient was admitted through the emergency department; and 179 categories for the patient's primary diagnosis. The third set of patient controls, shown in columns (3) and (6), augments the second set to also include indicators for whether the patient had any of 19 chronic conditions.

The result of Table 1.5 is that dispersion shrinks somewhat as rich controls about the patient's hospital admission are added in the first step, though controlling for patient illness histories has little effect. Moreover, the addition of physician effects does not systematically reduce these 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, meaning that a hospital with one standard deviation greater adoption gives 20pp 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. 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. 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 16pp 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 may raise or reduce dispersion within these groups. Large, urban, non-profit teaching hospitals, for example, have a standard deviation in coding rates of 0.17 without any first-step controls, 0.13 with patient controls, and 0.12 with patient and physician controls. Likewise, the standard deviation of coding rates among non-urban non-profit teaching hospitals falls from 0.17 with no controls to 0.13 with patient controls, but rises to 0.19 when physician controls are further added. These patterns are replicated in the other groups of hospitals: dispersion declines by 5-6pp with the inclusion of patient characteristics, but may decline (up to 1pp) or rise (up to 3pp) with the additional inclusion of physician effects.

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.

1.3.5 Describing the Adopters

In this subsection I first present the *ex ante* relationships one might expect between hospital characteristics and productivity based on theory and prior literature. I then show how these correlations are borne out in my data, with the caveat that these results are descriptive, not causal.

1.3.5.1 Potential Roles of Hospital Characteristics and Productivity

Size (Number of Beds) Larger hospitals may be more likely to adopt detailed HF coding if there are fixed costs of adoption – fixed costs are smaller, and more worthwhile to incur, on a per-patient basis when the hospital is larger. However, since size may be confounded with other factors that bear on coding, this explanation is only suggestive. In particular, a long line of research has documented a strong relationship between hospital size and quality in many areas, though with an unclear causal link (this is usually called the volume-outcomes hypothesis; see Epstein, 2002a for a critical review).

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), though for-profit hospitals have lagged public and non-profit facilities in the use of standards of care like β-blockers (Sloan et al., 2003). 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, 2005a; Silverman and Skinner, 2004). A key difference between this setting and the earlier work is that the prior literature 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.

Location Whether rural hospitals should be more effective at adopting the revenue-raising technology than urban hospitals is unclear *ex ante*, though evidence on outcomes and processes along the dimension of hospital location may be suggestive. There is substantial research indicating that health care outcomes and quality of care are lower in rural hospitals relative to their urban counterparts. At least some of this difference can be explained by rural hospitals being smaller. Hospitals in the farthest outlying rural areas appear to be the main driver of rural hospital underperformance (MEDPAC, 2012b; Baldwin et al., 2010).

Teaching Status Teaching hospitals have been found to have better outcomes and higher quality processes of care than non-teaching hospitals in observational studies (see Ayanian and Weissman, 2002 for a review). These studies do not necessarily control for hospital size, ownership, and other attributes. Still, 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.

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. However, 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. Even after controlling for a host of observables about the hospitals, unobserved characteristics may still exert an effect on adoption along this gradient.

Treatment Productivity and Quality Whether high treatment productivity hospitals are more likely to adopt the coding technology is not obvious. High productivity hospitals may have high quality managers who effectively work with physicians to incorporate consensus standards of care. These managers may use the same techniques to extract more detailed descriptions from their physicians. The managers could also use their treatment productivity-raising techniques to ensure that coding staff does not miss revenue-making opportunities.

On the other hand, a negative correlation between treatment and revenue productivity is also plausible. To the extent that productivity depends on managerial quality, the relationship between revenue productivity and treatment productivity 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.

1.3.5.2 Results

Table 1.6 displays the key estimates of the role of hospital characteristics and productivity in explaining takeup of the coding technology. The columns of this table show the results when different sets of first-step controls are included, repeating the sets of controls used in the dispersion analysis.

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, use of standards of care, and heart attack treatment productivity. Hospitals that are 10% larger give 0.21pp more of their HF patients a specific code. Adding patient controls when estimating hospital adoption reduces this effect to 0.15pp – 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 3.3pp more of their patients a specific code than for-profits and government-run facilities, though adding patient controls reduces the difference to 2.7pp. There is no significant difference between the takeup rates of for-profit and government-run hospitals. Finally, with the full set of patient controls, for each standard deviation rise in heart attack treatment productivity or in the use of standards of care (the composite measure is the sum of the heart attack, heart failure, pneumonia, and surgeries measures), about 2.4pp more HF patients tend to get a specific code. In other words, hospitals that appear to be higher quality and more productive in their treatment are also more likely to use these high-revenue billing codes. Hospital location, teaching status, and revenue at stake are not robustly correlated with takeup.

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 relationship between a hospital characteristic and coding indicates that the facility was able to extract more 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. In this section I focus on the coefficients of column (6), which adjust for the full set of patient characteristics as well as the physician when estimating the hospital component of adoption.

The inclusion of physician effects makes the hospital effects noisier, adding left-hand side measurement error to the regressions. This measurement error comes from sampling variance, so it does not bias the coefficients reported in columns (4) to (6), but it lowers the precision of the regression coefficients.

The use of detailed HF codes is clearly correlated with both heart attack treatment productivity 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 productivity would use specific codes for 2-3pp more of their patients. This gradient was also observed unconditional on the doctors (in columns 1-3), and 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.

The other significant relationship that exists with first-step patient and physician controls is that between hospital location and coding. Hospitals in large urban and other urban areas – areas of high and intermediate population density, respectively – extract specific codes from their doctors for 3-4pp more of their patients than hospitals in rural areas. This relationship does not exist without the physician controls, which indicates that urban hospitals 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 unconditional on physicians, urban and rural hospitals are about equally likely to use the detailed billing codes – the finding in columns (1)-(3).

Non-profit and for-profit hospitals were 1.7pp 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) – this value is reduced and no longer significant for the non-profit hospitals. Since removing the physician component of adoption reduces the coding advantage of non-profit facilities, it appears that the physicians who work at non-profit hospitals are more likely to provide the detailed documentation wherever they practice.

The gradient between hospital size and extraction of detailed HF codes is positive and significant without first-step physician controls, but it is eliminated when the physician component of adoption is swept away. Similar to the results for non-profit hospitals, this finding suggests that larger hospitals outperform smaller hospitals in column (3) because they utilize physicians that always provide more documentation wherever they treat patients.

1.4 Discussion

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 Griliches (1957), and it has also been found in health care, for example in the use of β-blockers and other therapies (see e.g.

Bradley et al., 2005 and Peterson et al., 2008). Likewise, a growing literature is finding persistent dispersion in productivity within narrowly defined industries (Fox and Smeets, 2011a; Syverson, 2011b); this literature is now expanding to include the health care sector (Chandra et al., 2013). I have shown that adoption of the HF coding technology across hospitals follows the established pattern.

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 an indicator for whether the hospital was able to solve a principal-agent problem. This component is robustly correlated with the use of consensus standards of care when treating patients. Thus hospitals that use treatment productivity-raising techniques 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 productivity and treatment productivity 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 0.16 percentage points, but it is not immediately clear whether this magnitude is small or large. 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, so it is reasonable to compare their variances. To the extent that there are substantial disparities across hospitals in their adherence to these standards, the disparities in coding also appear to be nontrivial. According to Table 1.4, the four standards of care scores have standard deviations ranging from 0.07pp to 0.14pp. The dispersion in the hospital component of HF coding adoption falls just 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, higher treatment quality, and more treatment-productive 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., 2003b), 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 will reduce 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 more productive in treatment and more likely to follow consensus standards of care already. Lower quality providers – i.e. those that are less productive or 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.

1.5 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 technology 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 productivity 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 technology adoption. For example, when Medicare opts to pay hospitals to use β-blockers, 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 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 could substantially improve the effectiveness of value-based payment reforms.

Figures

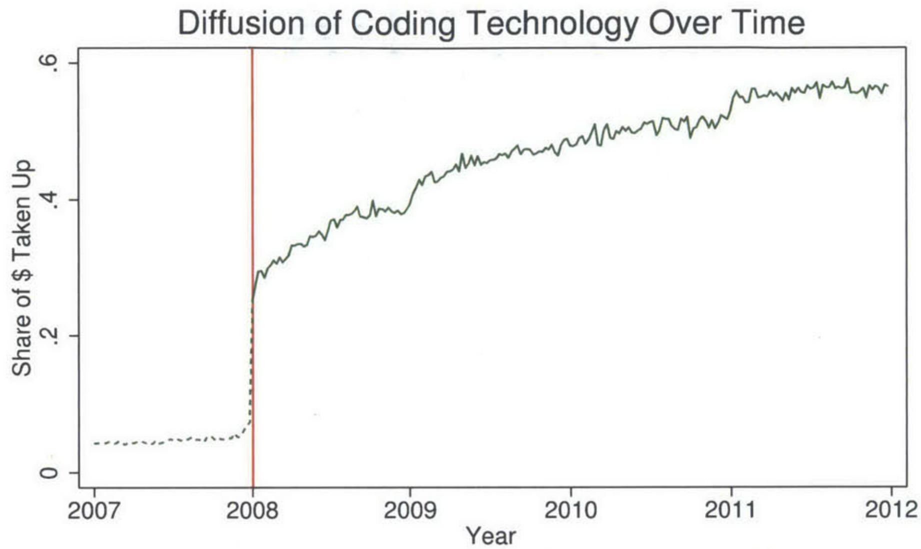


Figure plots the share of revenue available for detailed coding of HF that was captured by hospitals over time. Dotted line shows revenue that would have been captured in 2007 if hospitals had been paid per 2008 rules. The series is at the weekly level and the red line denotes the reform date.

Figure 1-1

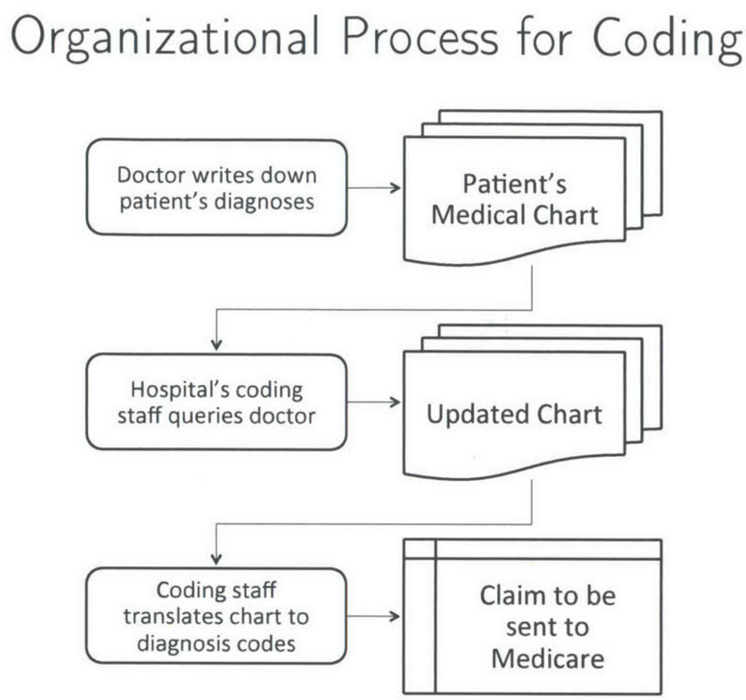


Figure 1-2

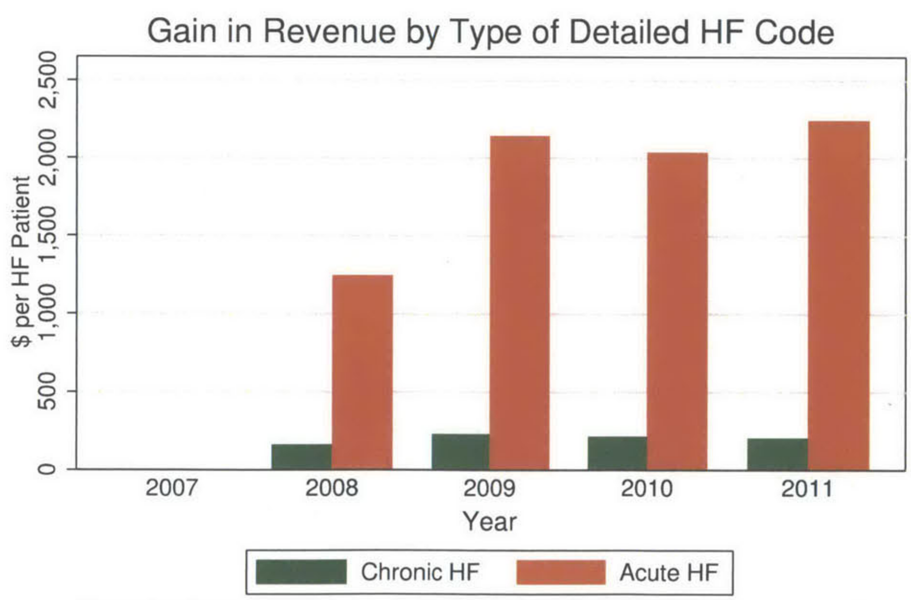


Figure plots the average per–HF patient gain in revenue going from always using vague codes for HF patients to always using chronic codes or acute codes.

Figure 1-3

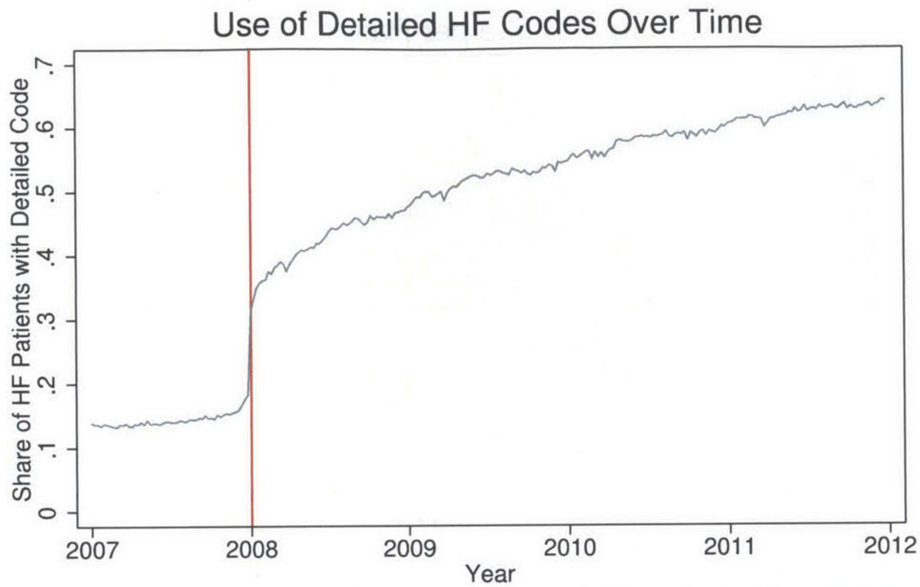
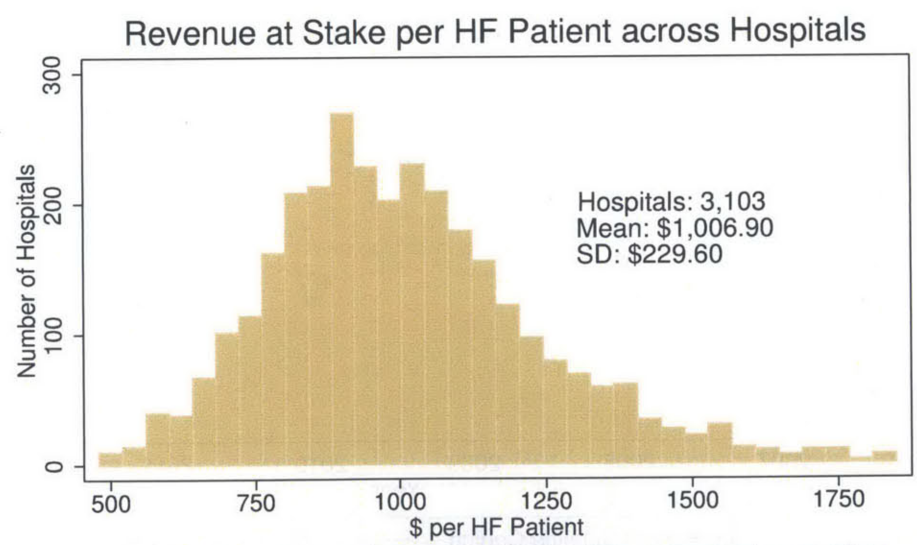


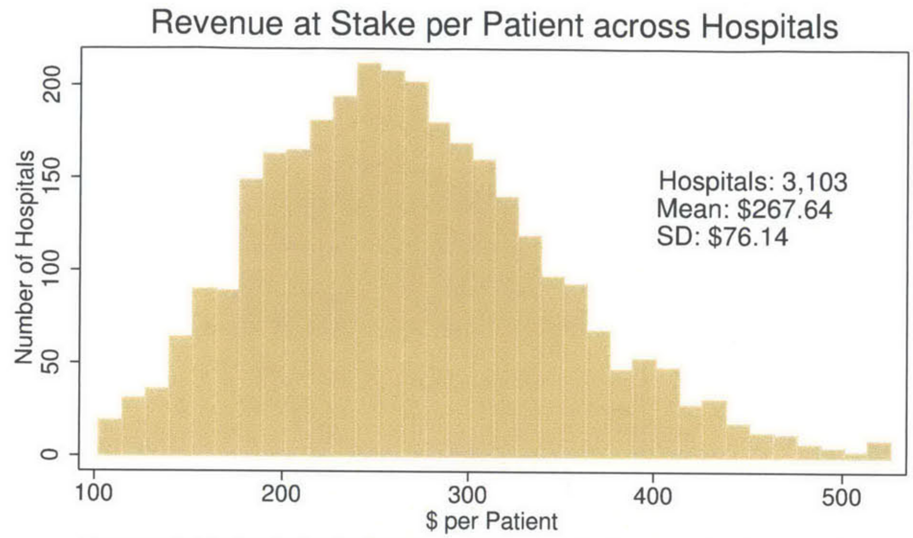
Figure plots the share of HF patients who received a detailed HF code over time. The series is at the weekly level and the red line denotes the reform date.

Figure 1-4



Revenue at stake is calculated using pre–reform (2007) patients processed under post–reform (2009) payment rules. The amount at stake equals the per–HF patient revenue with all HF patients given detailed codes less that revenue with all HF patients given vague codes. The 422 hospitals with <50 HF patients are suppressed. The outlying upper and lower 1% of hospitals were also suppressed.

Figure 1-5



Revenue at stake is calculated using pre–reform (2007) patients processed under post–reform (2009) payment rules. The amount at stake equals the per–patient revenue with all HF patients given detailed codes less that revenue with all HF patients given vague codes. The 422 hospitals with <50 HF patients are suppressed. The outlying upper and lower 1% of hospitals were also suppressed.

Figure 1-6

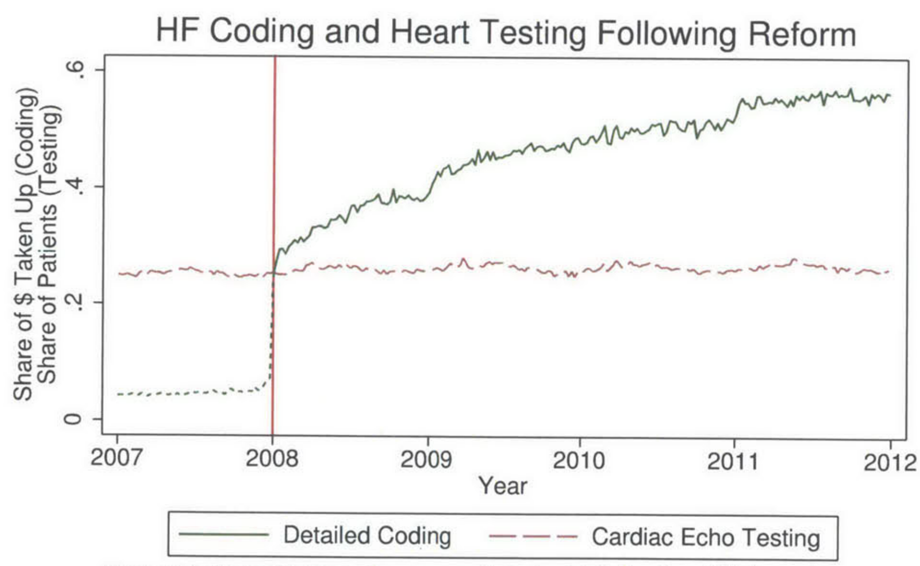
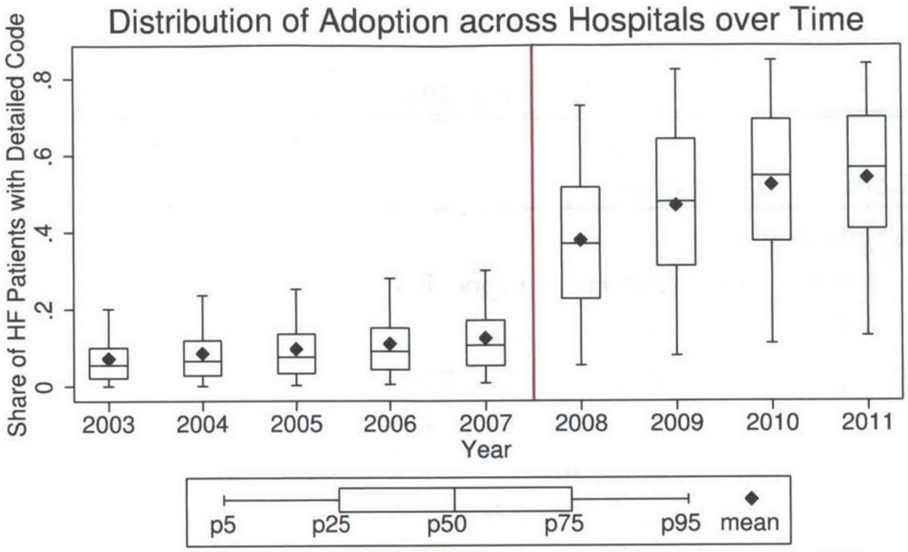


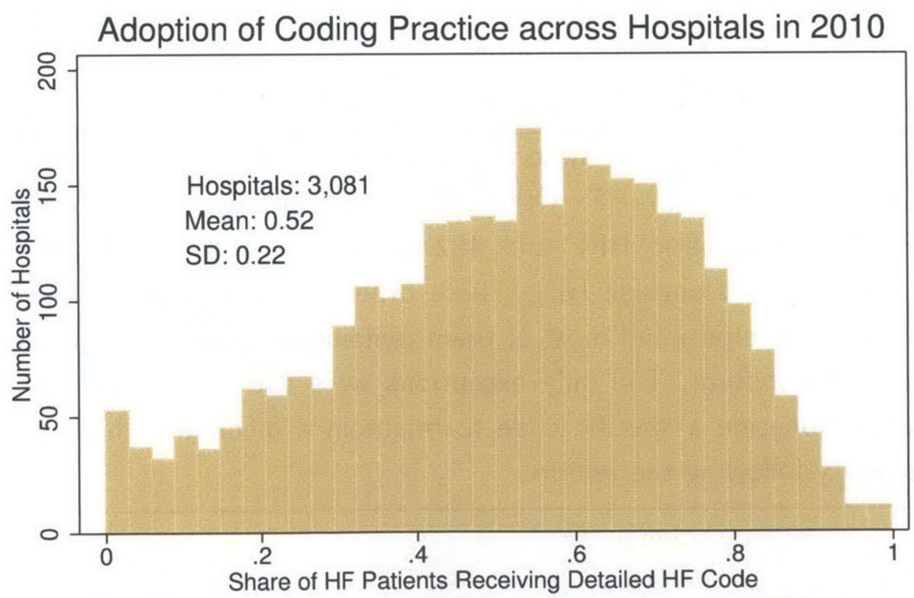
Figure plots the weekly share of revenue available for detailed coding of HF that was captured by hospitals alongside the weekly share of all patients who received a cardiac echo, a heart test. The dotted line shows revenue that would have been captured in 2007 if hospitals had been paid per 2008 rules.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

Figure 1-7



Box and whiskers show the distribution of adoption across hospitals in each year. A hospital's adoption equals the share of its HF patients in that year who received a specific HF code. Hospital-years with fewer than 50 HF patients are excluded. Red line separates pre- and post-reform years.

Figure 1-8



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 (N=441) are excluded.

Figure 1-9

Tables

Table 1.1 - 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. Other HF codes include 428.1 (Left HF); 398.91 (Rheumatic HF); and 402.x1, 404.x1, and 404.x3 (forms of hypertension alongside HF). These other codes were all high-severity before the reform and most of them remained medium- or high-severity after the reform. The only exceptions were two codes that could be used alongside a specific code to maintain a medium or high level of severity following the reform.

Table 1.2 - Statistics about the Full Analysis Sample and Mobility Sample

	(1) Mean	(2) SD	(3) Min	(4) Max
Hospitals (N=3,414; *N=2,868 in mobility sample*)				
HF Patients	552.7	578.5	1	5,435
HF Patients (mobility sample)	*581.7*	*553.1*	*1*	*4,607*
Distinct Physicians	*57.9*	*54.1*	*1*	*561*
Mobile Physicians	*19.8*	*21.9*	*1*	*173*
Physicians (N=134,502)				
HF Patients	*12.4*	*18.8*	*1*	*644*
Distinct Hospitals	*1.23*	*0.55*	*1*	*8*
Mobile (>1 hospital)	*0.188*	*0.391*	*0*	*1*

Italicized rows refer to the mobility sample: the subset of the analysis sample in which I observe the physician and can separately identify the hospital and physician effects. See text for more details. Full sample includes 1.9M HF patients. Mobility sample includes 1.7M HF patients.

Table 1.3 - Statistics about Physicians by Mobility Status

All values are means	(1) All	(2) Mobile	(3) Non-Mobile
Patient and Hospital Volume			
HF Patients	12.4	21.6	10.3
Distinct Hospitals	1.23	2.25	1
Mobile (>1 hospital)	0.19	1	0
Type of Physician			
Primary Physician	0.44	0.50	0.43
Medical Specialist	0.30	0.34	0.29
Surgeon	0.23	0.14	0.25
Demographics			
Female	0.19	0.15	0.20
Age	49.1	48.9	49.1
Training and Experience			
Years in Training	5.96	6.52	5.83
Years Since Training	16.0	15.4	16.1
Trained in US	0.71	0.59	0.74
Physicians	134,502	25,253	109,249

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.

Table 1.4 - Hospital Summary Statistics

Patient Controls	(1) N	(2) Mean	(3) SD
Heart Failure Coding			
HF Patients	2,411	709.4	606.3
Share Given Specific Code	2,411	0.546	0.201
Hospital Characteristics			
Beds	2,411	285.0	231.4
Ownership			
Non-Profit	2,411	0.667	0.471
For-Profit	2,411	0.164	0.371
Government	2,411	0.169	0.375
Location			
Large Urban Area	2,411	0.419	0.494
Other Urban Area	2,411	0.350	0.477
Rural Area	2,411	0.231	0.421
Teaching Hospital	2,411	0.371	0.483
Ex Ante $ at Stake / Patient	2,411	268.6	72.22
Standards of Care (share of times standards were used in 2006)			
for Heart Attack Treatment	2,411	0.916	0.085
for Heart Failure Treatment	2,411	0.826	0.113
for Pneumonia Treatment	2,411	0.864	0.061
for High-Risk Surgeries	2,411	0.797	0.119
Heart Attack Treatment (all heart attack patients 2000-2006)			
ln(Productivity)	2,411	0.919	0.171

A large urban area is an MSA with a population of at least 1 million; the remaining MSAs are considered other urban areas. A rural area is any location outside an MSA. A hospital's *ex ante* $ at stake per patient is the revenue put at stake by the reform per patient in the hospital (including non-HF patients). See text for more details on the standards of care and heart attack treatment measures. The standard deviation of heart attack treatment productivity is adjusted for sampling variance.

Table 1.5 - Standard Deviation of Coding by Type of Hospital

	(1)	(2)	(3)	(4)	(5)	(6)
Patient Controls	None	Admission	Full	None	Admission	Full
Physician Controls		None			Physician Fixed Effects	
Urban Non-Profit Hospitals						
Large and	0.168	0.135	0.134	0.145	0.116	0.123
Teaching	[281]	[281]	[281]	[281]	[281]	[281]
Other	0.187	0.142	0.141	0.193	0.158	0.170
	[407]	[407]	[405]	[407]	[407]	[405]
Non-Urban Non-Profit Hospitals						
Teaching	0.167	0.134	0.134	0.223	0.191	0.193
	[316]	[316]	[316]	[316]	[316]	[316]
Other	0.211	0.161	0.159	0.200	0.167	0.169
	[604]	[604]	[593]	[604]	[604]	[593]
For-Profit and Government-Run Hospitals						
Urban For-Profit	0.194	0.146	0.145	0.188	0.144	0.135
	[167]	[167]	[167]	[167]	[167]	[167]
Non-Urban For-	0.188	0.141	0.140	0.178	0.158	0.156
Profit	[229]	[229]	[224]	[229]	[229]	[224]
Government-Run	0.222	0.163	0.161	0.177	0.146	0.148
	[407]	[407]	[400]	[407]	[407]	[400]
All Hospitals	0.199	0.152	0.151	0.191	0.159	0.162
	[2,411]	[2,411]	[2,386]	[2,411]	[2,411]	[2,386]

Numbers in brackets count hospitals used to calculate the standard deviation. All results are adjusted for sampling variation. Urban hospitals are defined as those located in a "Large Urban" area. Large hospitals are defined as having at least 250 beds. Columns 1 and 4 include no patient controls when calculating the hospital's coding score. Columns 2 and 5 control for patient age, race, sex, admission through the emergency department, and principal diagnosis category. Columns 3 and 6 add controls for histories of chronic conditions. Columns 4-6 control for physician fixed effects when calculating the hospital's coding score.

Table 1.6 - 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)	0.0214***	0.0146**	0.0145**	0.00603	-0.000296	-0.00256
	(0.00775)	(0.00604)	(0.00599)	(0.0117)	(0.0100)	(0.0104)
Non-Profit	0.0328**	0.0250**	0.0269**	0.0194	0.0173	0.0172
Ownership	(0.0158)	(0.0118)	(0.0118)	(0.0160)	(0.0133)	(0.0138)
For-Profit	-0.000220	0.00119	0.00398	0.0176	0.0149	0.0173
Ownership	(0.0168)	(0.0124)	(0.0124)	(0.0198)	(0.0165)	(0.0172)
Located in Large	0.00830	0.00352	0.00178	0.0627***	0.0446**	0.0370*
Urban Area	(0.0156)	(0.0117)	(0.0117)	(0.0219)	(0.0187)	(0.0196)
Located in Other	0.0211	0.0125	0.0103	0.0632***	0.0406**	0.0332**
Urban Area	(0.0140)	(0.0105)	(0.0106)	(0.0190)	(0.0167)	(0.0168)
Teaching Hospital	0.00854	0.0118	0.0112	-0.00531	0.0111	0.0144
	(0.0100)	(0.00809)	(0.00804)	(0.0144)	(0.0123)	(0.0130)
Ex Ante $ at Stake	6.67e-05	6.48e-05	7.64e-05	9.34e-05	8.85e-05	0.000108
per Patient	(6.41e-05)	(4.82e-05)	(4.66e-05)	(8.90e-05)	(8.17e-05)	(8.64e-05)
Standards of Care and Productivity (Z_h)						
Standards of Care	0.0336***	0.0244***	0.0247***	0.0247***	0.0209***	0.0225***
Composite Z-Score	(0.00559)	(0.00428)	(0.00432)	(0.00691)	(0.00624)	(0.00632)
Heart Attack Treat	0.0288***	0.0240***	0.0237***	0.0309***	0.0305***	0.0272***
Productivity Z-Score	(0.00629)	(0.00479)	(0.00485)	(0.00863)	(0.00760)	(0.00776)
Observations	2,411	2,411	2,386	2,411	2,411	2,386
R^2 (adjusted)	0.091	0.091	0.092	0.037	0.036	0.030

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 age, race, sex, admission through the emergency department, and principal diagnosis category. Columns 3 and 6 add controls for histories of chronic conditions. Columns 4-6 control for physician fixed effects when calculating the hospital's coding score. The standards of care composite z-score is the sum of the four standards of care measures, normalized to mean 0 and standard deviation 1.

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

Chapter 2

Healthcare Exceptionalism? Productivity and Allocation in the U.S. Healthcare Sector*

2.1 Introduction

A central observation about the U.S. healthcare sector is the existence of substantial differences in productivity across regions and across hospitals. For example, annual Medicare spending per capita ranges from \$6,264 to \$15,571 across geographic areas (Skinner et al., 2011), yet health outcomes do not positively covary with these spending differentials (e.g. Fisher et al., 2003a,b; Baicker and Chandra, 2004b; Chandra et al., 2010; Skinner, 2011). Similar patterns have been documented across hospitals within geographic markets (e.g. Yasaitis et al., 2009). These facts have in turn generated substantial academic interest in understanding the root causes of the underlying productivity dispersion and what can increase productivity at under-performing hospitals (e.g. Skinner et al., 2006a; Chandra and Staiger, 2007; Skinner and Staiger, 2009a). Outside of academia, these "Dartmouth Atlas" facts have also attracted consider popular attention (see, for example, Gawande's 2009 *New Yorker* article) and were heavily cited by the Obama administration during the discussions leading up to the 2010 Affordable Care Act (e.g. Pear's 2009 *New York Times* article or Office of Management and Budget, 2009).

*This section is co-authored with Amitabh Chandra, Amy Finkelstein, and Chad Syverson. We are grateful to Daron Acemoglu, Nick Bloom, Iain Cockburn, Chris Conlon, Angus Deaton, Mark Duggan, Joe Doyle, Liran Einav, Matthew Gentzkow, Michael Greenstone, Jonathan Gruber, Ben Olken, Jonathan Skinner, Doug Staiger, Scott Stern, Heidi Williams, and numerous seminar participants for helpful comments and advice, and to Maurice Dalton and Nivedhitha Subramanian for expert research assistance. We gratefully acknowledge funding from the National Institute on Aging: P01 AG005842 and P01 AG019783 (Chandra), R01 AG032449 (Finkelstein) and T32-AG000186 (Sacarny)

The conventional wisdom in health economics is that the driving forces behind these large average productivity differences are various idiosyncratic, institutional features of the healthcare sector that effectively reduce competitive pressures on providers. Oft-cited culprits include uninformed consumers who lack knowledge of the quality and price differences across providers, generous health insurance that insulates consumers from the direct financial consequences of their healthcare consumption decisions, and public sector reimbursement that provides little incentive for productive efficiency by providers. These factors are widely believed to dull the basic disciplining force of demand-side competition that exists in most other sectors. Echoing and advancing this view, Cutler (2010) notes:

> "There are two fundamental barriers to organizational innovation in healthcare. The first is the lack of good information on quality. Within a market, it is difficult to tell which providers are high quality and which are low quality... Difficulty measuring quality also *makes expansion of high-quality firms more difficult* [emphasis added]... The second barrier is the stagnant compensation system of public insurance plans."

In a similar vein, Skinner (2011) states in his overview article on regional variations in healthcare:

> "[low productivity producers are] ... unlikely to be shaken out by normal competitive forces, given the patchwork of providers, consumers and third-party payers each of which faces inadequate incentives to improve quality or lower costs..."

This notion of "healthcare exceptionalism" has a long tradition in health economics. It dates back at least to the seminal article of Arrow (1963), which started the modern field of health economics by emphasizing key features of the health care industry that distinguish it from most other sectors and therefore warrant tailored study.

But when it comes to productivity dispersion, the ostensibly unique features of the healthcare sector stand alongside a large empirical literature outside of the health care sector that has documented extensively - almost without exception - enormous differences in average productivity across producers within narrowly defined industries (see Bartelsman and Doms, 2000; Syverson, 2011b; and references therein). For example, on average within narrow US manufacturing (4-digit SIC) industries, the 90th productivity percentile plant creates almost twice as much output as the 10th percentile plant, given the same inputs (Syverson, 2004a). This dispersion exists both within and across geographic markets (e.g. Syverson, 2004b,a).

We estimate that productivity dispersion across hospitals in treating heart attacks is about the same order of magnitude as productivity dispersion within narrowly defined manufacturing industries. Figure 2-1 (whose construction we describe in much more detail later in the paper) shows, for example, that productivity dispersion across hospitals for heart attack treatment is slightly lower than productivity dispersion across ready-mixed concrete plants. Ready-mixed concrete is, like healthcare, a spatially differentiated good in that it is produced and consumed locally, but one in which the product is less differentiated, insurance does not dampen price sensitivity, and prices aren't set administratively. More generally, looking across 450 different narrowly defined (4-digit SIC code) manufacturing industries in the US, average within-industry productivity dispersion in manufacturing is quite similar to our estimates across hospitals for heart attack treatment (Syverson, 2004a).

This finding is striking and, we believe, surprising. But, it admits multiple possible explanations. Productivity dispersion has been shown, both theoretically and empirically, to shrink with greater competition within and across industries (e.g. Syverson, 2004b,a; Martin, 2005; Balasubramanian and Sivadasan, 2009). However, we would not be comfortable drawing any direct inferences about the relative roles of competition in these two very different sectors from comparisons of their productivity dispersions.

Rather, these facts serve as a point of departure that motivates us to re-examine productivity and allocation within the healthcare sector using the analytical insights from the broader productivity literature. In particular, we draw on a long tradition of theoretical and empirical work in manufacturing examining whether higher productivity producers are systematically allocated greater market shares; in healthcare, the prevailing wisdom captured by the Cutler (2010) and Skinner (2011) quotations above is that these re-allocation forces are weak or non-existent.

Our findings suggest otherwise. Figures 2-2a and 2-2b (again discussed in more detail later in the paper) give a qualitative flavor of our results. They show that within a market-year, hospitals that have higher productivity for heart attack treatment tend to have greater market share (i.e., more heart attack patients) at a point in time (Figure 2-2a) and experience more growth in market share over time (Figure 2-2b). Quantitatively, we find that a 10 percent increase in hospital productivity is associated with about a 25 percent higher market share at a point in time and 4 percent more growth over the next 5 years.

A finding that the market allocates more market share to more productive firms at a point in time and over time is a robust characteristic of US manufacturing industries (Syverson, 2011b provides a recent review) but is noticeably absent from manufacturing in less competitive settings such as Central and Eastern European countries at the beginning of their transition to a market economy (Bartelsman et al., 2013), Chile prior to trade reforms (Pavcnik, 2002), or the US steel industry in the 1960s (Collard-Wexler and Loecker, 2013b). As a result, these allocation metrics are often interpreted as "signposts of competition." As in much of this previous work in manufacturing, we do not establish a causal link between competition and the signs of competition in the data. It could be that competitive market forces re-allocate market share to higher productivity hospitals, or it could be that higher productivity hospitals happen to have other features - such as beautiful lobbies or good managers - which separately increase demand. But whatever the driving force behind them, some force or forces in the healthcare sector lead it to evolve in a manner favorable to higher productivity producers. This finding puts US healthcare on a very different part of the map than, say, Romanian or Slovenian manufacturing in the early 1990s, where there appears to have been little (or even negative) correlation between a firm's productivity and its market share (Bartelsman et al., 2013). The results are particularly noteworthy given the context of heart attack treatments, where the acute nature of the condition might be expected to generate a smaller role for market forces in allocating patients to more productive hospitals than for less time-sensitive conditions such as cancer treatment, the management of chronic conditions, or elective procedures.

Taken together, our results suggest that healthcare may have more in common with "traditional" sectors than is commonly recognized in popular discussion and academic research. Continued efforts to understand productivity dispersion and uncover what may improve productivity in the US healthcare sector may therefore benefit from greater attention to the theoretical and empirical insights from the broader productivity literature. Naturally, the converse applies as well.

The rest of the paper proceeds as follows. Section 2.2 describes the analytical framework. Section 2.3 discusses our estimation of hospital productivity - the key empirical input to all our analyses. Section 2.4 presents our main results on the relationship between hospital productivity and market share. Section 2.5 discusses some questions of interpretation, including possible mechanisms behind the findings and various gauges of their magnitude. Section 2.6 shows that our main findings are robust to a variety of alternative specifications. A concluding section follows.

2.2 Analytical Approach: Static and Dynamic Allocation

Our primary empirical exercise examines the correlation between producer (i.e. hospital) productivity and market share at a point in time, and the correlation between producer productivity and growth in market share over time. These relationships have been analyzed in a variety of industries and countries as a proxy for the role of competition in these settings (e.g. Olley and Pakes, 1996; Pavcnik, 2002; Escribano and Guasch, 2005; Bartelsman et al., 2013; Collard-Wexler and Loecker, 2013b). Intuitively, competitive forces exert pressure on low productivity firms, causing them to either become more efficient, shrink, or exit.

Models of such reallocation mechanisms among heterogeneous-productivity producers have found applications in a number of fields, including industrial organization, trade, and macroeconomics.[1] While these models differ considerably in their specifics, they share a common intuition: greater competition - as reflected in greater consumer willingness or ability to substitute to alternate producers - makes it more difficult for higher-cost (lower-productivity) firms to earn positive profits, since demand is more responsive to their cost and price differentials relative to other firms in the industry. As substitutability increases, purchases are reallocated to more productive firms, raising the correlation between productivity and market share at a point in time (*"static allocation"*) and causing more productive firms to experience higher growth over time (*"dynamic allocation"*). Appendix A describes this archetypical mechanism slightly more formally.

[1] See, for example, Ericson and Pakes (1995); Melitz (2003); and Asplund and Nocke (2006).

For the static allocation analysis, we will use the following regression framework:

$$\ln(N_{h,t}) = \beta_0 + \beta_1 a_{h,t} + \gamma_{M,t} + \varepsilon_{h,t} \tag{2.1}$$

where $N_{h,t}$ is a measure of the market size of hospital h in year t, $\gamma_{M,t}$ are market-year fixed effects, and $a_{h,t}$ is our estimate of total factor productivity (which we refer to throughout as TFP) of hospital h in year t; we discuss in detail below how we estimate $a_{h,t}$. Thus β_1 reflects the static relationship between a hospital's TFP and its market share, within a hospital market-year. If the coefficient is positive, as has been found in many U.S. industries (e.g., Olley and Pakes, 1996; Hortaçsu and Syverson, 2007; Bartelsman et al., 2013), it indicates that higher productivity producers have a greater share of activity. If β_1 is zero or negative, as has been found for example in some former Soviet-bloc countries in the early 1990s (Bartelsman et al., 2013), in Chile prior to trade reforms (Pavcnik, 2002), and in the U.S. Steel industry circa 1960-70 (Collard-Wexler and Loecker, 2013b), it indicates that less productive industry producers are the same size or larger than their high productivity counterparts and suggests that forces beyond standard competition are driving the allocation of market activity.[2]

The static allocation analysis in equation 2.1 can reflect the market's ability to reallocate activity from less productive hospitals to more productive ones. But it shows the outcome of this process rather than the process itself. To measure the actual dynamics of the market's selection and reallocation mechanisms, we employ two additional metrics.

Our first dynamic allocation metric examines the relationship between hospital TFP and its probability of closing. We will estimate:

$$I\left[exit_{h,t+1}\right] = \beta_0 + \beta_1 a_{h,t} + \gamma_{M,t} + \varepsilon_{h,t} \tag{2.2}$$

where $I\left[exit_{h,t+1}\right]$ is an indicator equal to one if hospital h exits at time $t+1$, and the right hand side variables are defined as in equation 2.1. Thus β_1 reflects the relationship between a hospital's TFP and its probability of exit, controlling for any changes in aggregate exit probabilities across market-years. A negative relationship between TFP and hospital exit is one of the most robust findings in the productivity literature (See Bartelsman and Doms, 2000 and Syverson, 2011b for surveys). It is indicative of a Darwinian selection process at work: less productive producers find it more difficult to survive.

Our second dynamic measure is the relationship between hospital TFP and future hospital growth. We will

[2] A positive correlation between a hospital's productivity and the number of patients it treats is also consistent with increasing returns to scale, in which causality runs from scale to productivity rather than vice versa. This is a general issue for interpreting the static allocation measure in any industry. In the particular context of health care, the "volume-outcome" hypothesis conjectures that treating more patients improves provider performance. Not surprisingly, it has proven challenging to establish empirically whether an observed positive correlation between provider volume and outcomes is causal (see e.g. Epstein, 2002a for a discussion of the interpretation difficulties in this literature). Moreover, it is harder to understand why scale economies would predict our "dynamic allocation" finding that *current* productivity predicts increases in the number of *future* patients.

estimate:

$$\Delta_{h,t,t+1} = \beta_0 + \beta_1 a_{h,t} + \gamma_{M,t} + \varepsilon_{h,t} \tag{2.3}$$

where $\Delta_{h,t,t+1}$ is a measure of the hospital's growth rate (in terms of number of heart attack patients treated) between year t and $t+1$. A positive correlation between TFP and growth indicates that more productive hospitals see larger gains in patient traffic, and points to the operation of a selection and reallocation process. While not as robust as the negative TFP-exit relationship, there is widespread evidence in developed country manufacturing and retail that higher TFP producers experience growth in market shares (e.g. Scarpetta et al., 2002; Disney et al., 2003; and Foster et al., 2006).

Regression equations 2.1 through 2.3 form the heart of our empirical analysis. They describe the associations between a hospital's productivity and market share and indicate whether forces exist that are favorable to the expansion of higher productivity producers. Although motivated by models in which competitive forces create these re-allocation pressures, the correlations are naturally not direct evidence of the impact of competition. After presenting our results, we discuss possible interpretations in light of other forces that may mimic the effects of competition.

2.3 Estimation of the Hospital Production Function

The key empirical input for estimation of our analytical equations 2.1 through 2.3 is a measure of a producer's (i.e. hospital's) TFP. We estimate hospital TFP in the specific context of hospital treatment of heart attacks, analyzing the treatment and outcomes of about 3.5 million heart attack patients from 1993 through 2007. TFP is the amount of output a supplier can produce per unit input. In our setting, variation in TFP across hospitals reflects differences in patient survival (output) conditional on treatments (inputs) the patient receives. We describe the data and approach we use to estimate hospital TFP, and discuss key estimation challenges.

2.3.1 Setting: Heart Attack Treatments in US Hospitals

Heart attacks present an excellent setting for studying hospital productivity for a number of reasons. First, cardiovascular disease, of which heart attacks (acute myocardial infarctions, or AMIs) are the primary manifestation, is the leading cause of death in the United States. Second, the high post-AMI mortality (survival rates at one year are less than 70 percent in our Medicare population) provides an accurately measured outcome with a great deal of variation across hospitals. There is broad agreement that for AMIs, survival is the most important endpoint both clinically and in terms of patient preferences, and therefore a key measure of output, particularly in an elderly population.[3] Third, the emergency nature of heart attacks provides a setting in which the sorting of patients across providers is likely to be more limited than in many other healthcare settings, reducing empirical concerns arising from patients selecting into hospitals on the basis of their underlying health. At the same time, the reduced scope for sorting also makes the null hypothesis that higher productivity hospitals do not attract greater market share a particularly plausible one in this context. Finally, inputs are well measured and there exist rich data on the relevant health characteristics of the patients (called risk-adjusters) which can be used in the estimation. Not surprisingly, therefore, heart attacks have been the subject of considerable study in the medical and economics literature on the value of medical technology and the returns to medical spending (e.g. Cutler et al., 1998; Cutler and McClellan, 2001a; Skinner et al., 2006a; Chandra and Staiger, 2007).

2.3.2 The Hospital Production Function for AMI Patients

We posit a patient-level health production function of the following form:

$$y_p = A_{h,t} \left(\prod_k R_{p,k}^{\alpha_k} x_p \right)^{\mu} e^{\varepsilon_p} \tag{2.4}$$

where x is the number of post-AMI survival days of patient p treated at hospital h in year t, and x_p is a measure of hospital inputs used to treat this patient. All production functions relate outputs to inputs; our particular function uses patient survival days as a measure of output and a single (dollar-denominated) index of resources spent on the patient as inputs.[4] Because patients are inherently heterogeneous, survival may also depend on characteristics of the patient, which could potentially also be correlated with input choices. In addition, the marginal effect of inputs on survival may vary with patient characteristics. To capture both of these effects, we follow the literature and adjust inputs for a vector of observable patient-level risk factors, $R_{p,k}$, where k indexes the factors. The parameters α_k capture the influence of these risk factors on health. Thus the expression in the parentheses reflects risk-adjusted inputs on the patient. The parameter μ is the elasticity of survival days with respect to risk-adjusted inputs. Finally, the expression e^{ε_p} is a patient-level error term that accounts for random variations in health outcomes.

The key input into all of our analyses described in Section 2.2 is the logarithm of $A_{h,t}$, which we have previously called $a_{h,t}$. $A_{h,t}$ measures the (exponent of) total factor productivity (TFP) of hospital h in year t. It is common across all (risk-adjusted) patients in that hospital in that year.[5] Holding risk-adjusted

[3] Clinical trials for heart-attack therapies compare treatments by focusing on survival as the key outcome (see for example, Andersen et al., 2003), but this is not true for trials of treatments for more elective coronary conditions such as stable coronary disease where quality of life concerns make it more difficult to measure output. A review of over twenty-three trials for heart-attack treatments is provided by Keeley et al. (2003).

[4] This sort of single-input production function is unusual but convenient; one could reasonably interpret the single input as an index of the use of multiple inputs that go into producing health. In Appendix E we show the results are robust to the use of a multi-input production function instead.

[5] We allow hospital productivity to vary across years because it allows us to capture intertemporal variation in hospitals'

inputs constant, differences in $A_{h,t}$ across hospitals produce systematic differences in survival length. In other words, if it were possible to send a particular heart attack patient to two hospitals with different TFP levels, providing him the same level of inputs at both, the patient's expected survival would be greater in the higher TFP hospital than in the lower one.

The hospital production function model in 2.4 allows variation across providers in the marginal health product of inputs (i.e., $A_{h,t}\mu$ varies across hospital-years) but constrains them to have the same elasticity of output with respect to input (i.e., μ is common across hospitals). Our empirical specification therefore allows the "marginal return to inputs" curve to vary across hospitals, as suggested by Chandra and Staiger (2007) and Garber and Skinner (2008). Figure 2-3 provides a stylized illustration of our production function specification.

Taking logs, we have our main estimating equation for the hospital production function:

$$\ln(y_p) = \ln(A_{h,t}) + \mu \sum_k \alpha_k \ln(R_{p,k}) + \mu \ln(x_p) + \varepsilon_p \tag{2.5}$$

To estimate equation 2.5 we regress the log of patient survival days on a vector of risk factors $(R_{p,k})$, the inputs applied to each patient (x_p), and a set of hospital-year fixed effects. These hospital-year fixed effects are in turn our TFP estimates $(a_{h,t} \equiv \ln(A_{h,t}))$ which we then use as inputs to estimate our main analytical equations 2.1 through 2.3.

2.3.3 Data and Measurement of Key Variables

Our primary dataset consists of all Medicare Part A (i.e., inpatient hospital) claims for all heart attacks (AMIs) in individuals age 66 and over in the United States from 1993 through 2007. We limit the sample to AMIs in patients who have not had an admission for an AMI in the prior year. We have information on mortality through 2008, so we can observe at least one year of post-AMI survival for all patients. In order to have enough data to estimate annual hospital productivity, we follow standard practice (e.g. Skinner and Staiger, 2009a) and eliminate any hospital-year with fewer than 5 heart attack patients that year. This restriction eliminates less than 1 percent of patients, but about 10 percent of hospital-years and 6 percent of hospitals; naturally the dropped hospitals are disproportionately small.

Tables 2.1a and 2.1b present some basic summary statistics on our sample. Our final sample consists of about 3.5 million heart attacks in 55,540 hospital-years and 5,346 unique hospitals. The average hospital-year has about 65 patients, but the median hospital-year has only 39 patients. We follow the literature in defining a hospital market (M) for an AMI as a Hospital Referral Region (HRR, see e.g. Chandra and Staiger,

efficiencies, and because it is consistent with standard practice in the broader productivity literature outside the healthcare sector. As we discuss below, we find that hospital productivity is highly persistent across years within our sample.

2007).[6] Our sample includes 304 HRRs, and on average they have about 12 hospitals in them. The Medicare claims data also include information on patient demographics (age, race and sex) and detailed information on co-morbidities (i.e. admissions for other conditions) during the prior year. We use this information as a basis of our risk adjusters $R_{p,k}$.

Our baseline output (survival) measure (y_p) is the number of days that the patient survives after receiving initial treatment, up through the first year. Survival includes the first day of treatment itself, so y_p is bounded from below at 1 and above at 367 days. As shown in Table 2.1, average survival through 1 year, censoring anyone who survives more than 1 year at 367 days of survival, is 268 days; about two-thirds of our sample survives past one year. We show below that our core results are robust to alternative time horizons for measuring output (i.e. 30 day or 5 year survival windows).

Our baseline input measure defines hospital factor inputs for a patient as the (dollar-converted) sum of diagnostic-related group (or DRG) weights during the first 30 days following a heart attack. These DRG weights reflect the Centers for Medicare and Medicaid Services' (CMS's) assessment of the resources necessary to treat a patient as a function of the patient's comorbidities and procedures received. This approach is standard in the literature and ensures that we measure real services rendered to patients, purged of reimbursement (price) variation across geographic areas or hospitals (see e.g. Skinner and Staiger, 2009a; Gottlieb et al., 2010). Appendix B gives a detailed description of our baseline input measure and the sources of variation that contribute to it.[7] About 15 percent of the variation is explained by indicator variables for whether the patient received one of two surgical procedures: bypass or stent.

On average, about $16,000 worth of hospital inputs are used on one of our patients in the 30 days following a heart attack, with a standard deviation of about $12,000. As is typical in healthcare, inputs are right skewed; the median is about $12,000 and the 90th percentile is nearly $32,000. We show below that our core results are generally robust across a wide range of alternative input measures, as well as across alternative time horizons for measuring inputs.

2.3.4 Estimation Challenges

Estimating productivity in any setting is conceptually straightforward but practically involves a number of measurement challenges (Syverson, 2011b). In addition to the measurement of output and inputs discussed

[6]The *Dartmouth Atlas of Healthcare* divides the United States into HRRs which are determined at the zip code level through an algorithm that reflects commuting patterns to major referral hospitals. HRRs, which are akin to empirically defined markets for healthcare, may cross state and county borders. A complete list of HRRs can be found at `http://www.dartmouthatlas.org/`. Since defining a market is not a straightforward undertaking, in Appendix D (Table A4) we also show that our results are robust to defining markets based on Hospital Service Areas (HSAs) instead; there are about 10 times as many HSAs as HRRs.

[7]As described in Appendix B, we make an adjustment to the prior literature's approach to account for the fact that some of CMS's DRGs are defined partly based on subsequent survival status. We purge our measure of this outcome-based variation in input measurement by assigning the relevant patients the average weight across the DRGs which distinguish otherwise similar treatments based on survival. We also discuss some of the challenges in measuring inputs in other settings (such as the handling of intermediate inputs or different qualities across workers) that we avoid here, as well as shared challenges such as the appropriate weighting of different inputs.

above, we describe three other challenges to estimating the hospital production function: endogeneity of inputs, differences across hospitals in patient characteristics related to survival, and estimation error.

2.3.4.1 Endogeneity of Inputs

A general econometric concern that pervades production function estimation is the potential endogeneity of inputs. In a typical setting, productivity is the residual in a firm-level regression of outputs on inputs; therefore, the coefficient on inputs (μ in our setting) may be biased by a correlation between input choice and the residual (productivity). In our setting, however, because we observe production at the unit (patient) level, we can include hospital-year fixed effects, estimating μ solely from within-hospital-year variation. By identifying the coefficients on inputs only from variation within hospitals, we control for any tendency for hospitals with different productivity to use different amounts of inputs on average. Of course, any unobserved inputs that do not vary within the hospital (such as, for example, whether the hospital requires its staff to use checklists) will load onto our estimate of hospital productivity. This is not a problem per se; as in the productivity literature more broadly, we think of productivity as the component of output that cannot be explained by observed inputs.

However, our estimates will be biased if, within hospital-year, hospitals choose different observable input levels for patients who differ unobservably in their latent survival, or if their choice of unobservable inputs is correlated with observed inputs at the patient level. The sign of the bias of the estimate of μ is not obvious. Moreover, our focus is not on estimating μ. Our primary concern is what impact any bias in μ will have on our analysis of the relationship between estimated productivity and market share, which are the ultimate objects of interest for the analysis. We therefore evaluate below the robustness of our main results to imposing, rather than estimating, various values for the scale parameter μ. This method amounts to following the index number, or Solow residual, approach to measuring productivity in which factor elasticities are taken from auxiliary data such as factor cost shares. We are re-assured that our main results are quite insensitive to the choice of μ. This insensitivity also has an economic interpretation that we discuss below.

2.3.4.2 Differences Across Hospitals in Patient Characteristics

Even if μ is known and imposed based on auxiliary information, if patients at different hospitals differ on average in their unobserved survival probabilities, this variation will cause us to misestimate hospital productivity. As noted earlier, one of the reasons for the focus on heart attacks in the empirical literature is the belief that such patient sorting across hospitals may be less of an issue in an emergency setting. But this does not mean there is no potential for sorting; indeed, were there no mechanisms by which patients (or their surrogates) actively selected hospitals for AMI treatment, it would be difficult to view our re-allocation findings as consistent with a role for market forces.

Therefore, to try to minimize the impact of any unobserved patient health differences across hospitals, we follow the standard practice in the literature and include various risk adjusters ($R_{p,k}$) to control for observable patient characteristics that are related to health. In particular, our baseline specification controls for a full set of interactions between age (in five-year groupings), gender, and whether the patient is white, as well as various co-morbidities. Each co-morbidity is included as an indicator for whether the patient has been to the hospital for a specific condition in the year prior to the AMI admission. Table 2.1b shows that on average our patients are 78 years old (recall our sample is for the Medicare population), about half are female, and about 90 percent are white; it also presents the means for the 17 co-morbidities we include in our baseline specification. We show below that our main results are quite insensitive to using fewer or more (for a subsample of patients where they are available) risk adjusters.

2.3.4.3 Estimation Error in TFP Measures

The median hospital-year in our sample has less than 40 patients, and for 20 percent of our hospital-years we observe fewer than 15 patients. The consequence of a relatively small number of patients in some hospital-years, together with the stochastic nature of our outcome (survival), means that our key object of interest and input into all of our productivity metrics - hospital TFP, $a_{h,t}$ - may be estimated with error. Such estimation error will cause attenuation bias in our analysis of the relationship between market share and hospital productivity in equations 2.1 through 2.3.[8]

We therefore apply the standard shrinkage or "smoothing" techniques of the empirical Bayes literature (e.g. Morris, 1983) to adjust for estimation error in our estimates of hospital productivity.[9] Appendix C provides a detailed description of this procedure. The intuition behind it is that when a hospital's productivity is estimated to be far above (below) average, it is likely to be suffering from positive (negative) estimation error. Therefore, the expected level of productivity, given the estimated productivity, is a convex combination of the estimate and the mean of the underlying productivity process. The relative weight that the estimate gets in this convex combination varies inversely with the noise of the estimate (which is based on the standard error of the hospital-year fixed effect). In practice, as we show in Appendix C, our core finding that hospitals with higher estimated productivity get allocated more market share at a point in time and over time remains statistically significant without the empirical Bayes adjustment, although naturally the magnitude is attenuated. All the analyses of hospital TFP use the empirical Bayes adjustment unless explicitly noted.

[8] This small-sample problem is probably much less of an issue in more traditional settings for estimating productivity, since the number of units of output produced (the statistical analog of patients in our context) is much larger. Increasingly, however, the productivity literature is also trying to adjust for other sources of measurement error in output (e.g. Collard-Wexler, 2011; Dobbelaere and Mairesse, 2013).

[9] McClellan and Staiger (1999) introduced this approach into the healthcare literature when estimating quality differences across hospitals, and it has since been widely applied in the education literature for estimating and analyzing teacher or school value added measures (e.g. Kane and Staiger, 2001; Jacob and Lefgren, 2007).

2.3.5 Estimates of the Hospital Production Function

Table 2.2 presents our estimates of the "returns to scale" parameter (μ) from estimating equation 2.5. Column 1 presents our baseline estimates, which use our full set of risk adjusters. We estimate a coefficient on log patient inputs (μ) of 0.446 (standard error = 0.005), which suggests that every 1 percent increase in inputs per patient is associated with a 0.45 percent increase in survival days. A comparison of columns 1 through 3 indicates that our estimate of μ increases from 0.45 to 0.59 as we reduce the set of risk adjusters to just age, race and sex (column 2) or to nothing (column 3), with the age-race-sex risk adjustment accounting for most of the difference between the results with no risk adjusters and with all risk adjusters included. Our estimates of μ are in the middle of the (very wide) range of estimates that papers in this literature have produced.[10]

The key input into our productivity metrics is not our estimate of μ but rather our estimates of TFP, $a_{h,t}$. These objects are the hospital-year fixed effects from equation 2.5 and are the key right-hand-side variables in our estimating equations 2.1 through 2.3. We find a great deal of within-hospital persistence in productivity over time, with $a_{h,t}$ exhibiting an AR(1) coefficient of about 0.7.

As a validity check on whether our estimates are picking up differences in hospital productivity, we verify that these estimates correlate positively in the cross-section with observable and independently gathered hospital quality measures. This exercise is in the spirit of Bloom and Reenen (2007), who perform the reverse procedure: validating an observable measure of management quality by correlating it with estimates of firm level productivity.

The results are summarized in Table 2.3, and several are presented graphically in Figure 2-4.[11] The first two columns of Table 2.3 show the correlation between our estimates of hospital TFP and two quality measures that were first collected by the Center for Medicare and Medicaid Services (CMS) in 2003; they have been publicly reported by the agency's "hospital compare" website (`http://www.hospitalcompare.hhs.gov`) since 2005. They are calculated by hospitals and submitted to CMS independently of the data that we use.

These measures are created to indicate the fraction of patients who received the treatment(s) that CMS determined were appropriate for their medical conditions. In the regressions, we convert them to z-scores by normalizing their means and variances to 0 and 1, respectively. In Table 2.3 column 1 we look at the

[10] Skinner and Staiger (2009a) note that various papers have used different right hand side specifications or sample periods to produce estimates of the "return to spending." They re-estimate many of these alternative specifications in a within-hospital linear probability model of an indicator for one year survival on one year inputs and produce estimates ranging from -0.015 to 0.122. In our data such linear probability models produce estimates of the "return to spending" of 0.072 to 0.100, depending on the risk adjusters. Within-hospital estimates of the return to input use tend to produce a positive relationship between inputs and survival, in contrast to the cross-region or cross-hospital comparisons that tend to find no or negative association between inputs and health-related outcomes. One parsimonious explanation for this difference would be if low productivity hospitals tended to compensate by using more inputs.

[11] Table 2.3 and Figure 2-4 examine regressions of our estimates of hospital TFP in 2003 on various hospital characteristics. We omit the EB correction for hospital TFP since classical measurement error on the left-hand side does not affect the consistency of a regression. The estimates of a hospital's 2003 TFP come from our full sample estimates of equation 2.5, but we use only a single year since most of the hospital characteristics are only available cross-sectionally. We choose 2003 estimates since that is the first year that the CMS quality measures are available.

hospital's z-score for beta blockers, which are inexpensive drugs that reduce the demands on the heart and are long-established as having important benefits for AMI patients after discharge. In column 2 we look at the z-score of a combined measure that sums across the number of patients who are given each of eight consensus AMI treatments and divides by the sum of patients appropriate for each of these treatments.[12] All of these measures have been studied in the literature and are considered indicative of good quality care (e.g. Higashi et al., 2007; Skinner and Staiger, 2009a; Jha et al., 2005; and cites therein). We use the measure in the first year it was collected to minimize the chance that hospitals responded to the reporting by changing the measure and thus reducing its signal of quality.

In column 3 we use the Bloom et al. (2012a) measure of hospital management quality.[13] It is based on a survey of management practices that were administered to a sample of approximately 300 hospitals in 2009 and 2010; a higher management z-score indicates closer conformance to management best practices. This measure of management quality has been found to be significantly negatively correlated with 30 day risk-adjusted mortality for patients in cardiac units (McConnell et al., 2013a); outside the hospital sector, it has also been found to correlate positively and significantly with productivity, profitability, Tobin's Q, and firm survival (Bloom and Reenen, 2007).

Reassuringly, the results indicate a positive correlation between these "external" measures of the quality of the hospital and our estimates of hospital productivity. For example, we estimate that a one standard deviation increase in the hospital's beta blockers score is associated with a 3 percent increase in hospital productivity. The results are statistically significant for the beta blockers and composite score; the results for the hospital management measure (which are available for only a very small subsample of our hospitals) are significant at the 10% level. We also find that teaching hospitals and urban hospitals have higher estimated productivity; estimated productivity is higher for non-profit hospitals than for for-profit or public hospitals.

2.4 Main Results: Static and Dynamic Allocation

Table 2.4 presents our central results on the static and dynamic allocation of patients across hospitals. In our discussion, we focus on column 1, which presents our baseline estimates based on the full set of risk adjusters (i.e. the same specification as shown in Table 2.2, column 1); the results are not sensitive to the choice of risk adjusters (columns 2 and 3).

The first row shows our static allocation analysis based on estimation of equation 2.1, examining the correlation between a hospital-year's productivity, $a_{h,t}$, and the logged number of heart attack patients it

[12] The eight measures are 1) given aspirin at arrival, 2) given aspirin at discharge, 3) given ACE inhibitor for left ventricular systolic dysfunction (LVSD), 4) given smoking cessation advice/counseling, 5) given beta blockers at arrival, 6) given beta blockers at discharge, 7) given fibrinolytic medication within 30 minutes of arrival, and 8) given percutaneous coronary intervention (PCI) within 90 minutes of arrival.

[13] We are extremely grateful to Nick Bloom for providing us with these measures.

treats, $\ln(N_{h,t})$. Because we include market-year (HRR-year) fixed effects, this estimate is within market-year, relating a hospital's market share of heart attack patients to its TFP relative to other hospitals in its market-year. Our right-hand side measure of $a_{h,t}$ $(= \ln(A_{h,t}))$ is the estimate of productivity from estimation of the hospital production function in equation 2.5. We bootstrap the standard errors, clustering at the market level.

The results show a statistically significant positive relationship between productivity and market share, suggesting that within markets, more market share (patients) tends to be allocated to more productive hospitals at a point in time. In particular, our baseline estimate suggests that a 10 percent increase in a hospital's productivity is associated with about a 25 percent higher market share.[14] A visual presentation of the results is given in Figure 2-2a.

The second row shows our analysis of the TFP-exit relationship based on estimation of equation 2.2, which examines the within market-year relationship between a hospital-year's productivity $a_{h,t}$ and an indicator variable for whether the hospital "exits" next year. The regression's right-hand side and standard errors are calculated as in the static allocation analysis. We define the dependent variable $I[exit_{h,t+1}]$ equal to one if hospital h has less than 5 heart attack patients in each year from year $t+1$ to $t+5$.[15] We measure exit as the lack of more than 5 patients in each of five subsequent years to try to ensure that we've captured a "permanent" reduction in volume, as opposed to measurement error stemming from idiosyncratic fluctuations in the number of patients that a hospital receives.

We find a statistically significant negative relationship between hospital productivity and subsequent exit. The baseline results suggest that a 10 percent increase in hospital productivity within a market-year is associated with a statistically significant decline in the probability of exit next year of about 0.3 percentage points (about an 8 percent decline relative to the baseline exit rate of 4.4 percent).

The bottom row of Table 2.4 shows our analysis of the TFP-growth relationship based on estimation of equation 2.3, which examines the within market-year relationship between a hospital-year's productivity $(a_{h,t})$ and its subsequent one-year growth. The right-hand side and standard errors are calculated as in the prior analyses. For our left-hand side measure of the hospital's one-year growth rate $\Delta_{h,t,t+1}$ we define

$$\Delta_{h,t,t+1} = \frac{N_{h,t+1} - N_{h,t}}{\frac{1}{2}(N_{h,t+1} + N_{h,t})} \tag{2.6}$$

where $N_{h,t}$ is once again the number of heart attack patients treated by hospital h in year t. Our measure

[14]Because our sample is limited to hospital-years with at least 5 patients, there is a potential concern about selection on the dependent variable in the static analysis. (This is not a concern for the subsequent dynamic analysis). We explored the sensitivity of our static allocation results to an alternative, Tobit-style truncated regression and found that the static allocation results were slightly strengthened by this adjustment.

[15]There are a non-trivial number of hospital mergers over our time period. If hospital A merges with hospital B and physically shuts down, hospital A is coded as having 0 patients in subsequent years. If however, hospital A and B both continue to exist physically and admit their own patients (e.g. Beth Israel and Deaconess), they continue to be coded as separate hospitals with each still assigned the AMI patients whom they admit.

of the hospital's one-year growth rate thus divides the change in the number of patients between this year and next year by the average number of patients across these two years.[16]

Again, the estimates are statistically significantly different from zero. The baseline results suggest that a 10 percent increase in hospital productivity within a market-year is associated with over a 1 percent increase in the number of patients the hospital treats in the next year.[17] Figure 2-2b gives a visual presentation of this relationship between hospital productivity and growth.

2.5 Interpretation and Discussion

2.5.1 Mechanisms

The above findings indicate that more productive hospitals have statistically significantly higher market share at a point in time and are more likely to increase that market share over time. These findings contrast with the conventional wisdom - summarized in the introductory quotations - that there is little in the healthcare sector to encourage the growth of higher productivity providers or weed out lower productivity ones. Our findings place US healthcare, at least qualitatively, in the same part of the spectrum as US manufacturing, and differentiate it from many less competitive manufacturing settings where these relationships have been found to not exist or even to have the opposite sign.

What mechanisms might act to allocate more patients to higher productivity hospitals in an emergency setting like heart attacks? A definitive answer is beyond the scope of this paper. However, we try in this section to present some initial, suggestive evidence.

We begin by examining whether the positive relationship between productivity and market share is primarily driven by patients choosing hospitals that, for a given amount of inputs, are more likely to produce high survival, or hospitals that, for a given amount of survival, use fewer inputs. Figures 2-5a and 2-5b therefore show the within market-year correlation, respectively, between risk-adjusted survival and market share (conditional on risk adjusted inputs) and between risk-adjusted inputs and market share (conditional on risk adjusted survival).[18] The results suggest that the productivity-market share relationship is primarily driven by the relationship between risk-adjusted survival and market share. The positive correlation between risk-adjusted survival and market share (Figure 2-5a) is virtually the same as that between risk-adjusted

[16] This monotonic transformation of the standard percentage growth rate metric bounds growth between -2 (exit) and +2 (growth from an initial level of 0). An attraction of this transformation is that it reduces the chance that the results are skewed by a few fast-growing but initially small hospitals that would have very large percentage growth rates. This growth rate transformation has been used in other contexts to avoid unnecessary skewness in the growth rate measure; see, for example, Davis et al. (1998).

[17] Table 2.4 reports negative average annual growth; this is primarily due to the fact that our measure conditions on the hospital initially being in the market.

[18] As with our productivity estimates, we use an empirical Bayes correction to adjust our estimates of risk-adjusted survival and of risk-adjusted inputs for measurement error; our procedure accounts for the correlation in measurement error between these two objects.

productivity and market share in Figure 2-2a. The negative correlation between risk-adjusted inputs and market share (Figure 2-5b) is statistically significant but less than half the magnitude. These findings are consistent with patients and their surrogates primarily seeking out hospitals that achieve higher risk-adjusted survival (conditional on risk adjusted inputs) rather than seeking out ones that use fewer risk-adjusted inputs (conditional on risk-adjusted survival). In practice, we find that risk-adjusted survival and productivity are extremely highly correlated.

It is not immediately obvious how patients know which hospitals offer longer survival. This ambiguity is not unique to our study. Indeed, a long-standing question in the field - dating back at least to Arrow (1963) - is how patients can acquire information on provider quality. One possibility is some form of market-learning; hospitals acquire a reputation for good outcomes and this reputation spreads through physicians' professional networks and patients' social networks and influences patients, family members, physicians, and ambulance drivers to request treatment at hospitals that are better at producing survival. Indeed, in a related setting, Johnson (2011) finds that cardiac specialists who have higher risk-adjusted survival rates for their patients are less likely to stop practicing. She interprets this and related evidence as consistent with a model of market learning by the referring physician. Patients or their family members may also obtain such information themselves; there is some evidence, for example, that patients respond to provider report cards (e.g., Dranove et al., 2003 and Dranove and Sfekas, 2008).

An alternative view, however, is that there is no scope for AMI patients or their surrogates to exercise choice over hospitals because in emergency situations all (or most) patients simply get taken to the nearest hospital. This hypothesis seems particularly natural given the famous McClellan et al. (1994) use of distance as an instrumental variable for which hospital treats a given AMI patient. With mechanical assignment of many patients to the nearest hospital, our static and dynamic allocation results could be produced spuriously if, for example, within a market, more densely populated (e.g. urban) areas have both higher productivity hospitals and faster population growth.

In practice, however, this type of strict mechanical allocation rule does not seem able to explain our findings. For one thing, we estimate that slightly over half of AMI patients go to a hospital that is *not* the closest one in their market; in other words, while the McClellan et al. (1994) distance instrument has a significant first stage with respect to hospital choice, its R^2 is far from 1. There is therefore scope for demand to affect patient allocation to hospitals in the AMI context. Moreover, when we produce a counterfactual allocation of patients by assigning each patient to his nearest hospital within an HRR instead of the one at which we observe treatment, our static and dynamic allocation results either substantially attenuate or actually reverse.[19]

Of course, the presence of active hospital choice by AMI patients or their surrogates does not establish

[19] Specifically, the exit result reverses sign and is statistically insignificant; the growth result is less than 20 percent of the baseline estimate and is statistically insignificant; the static allocation result remains statistically significant but with a magnitude that is 20 percent of the baseline estimate; see Appendix Table A4 (Columns 1 vs 3) and Appendix D for more detail.

that they are choosing on the basis of hospital productivity or risk-adjusted survival as in the speculative discussion of market learning above. It is possible that the correlation between productivity and market share reflects omitted factors that independently drive demand and correlate with productivity. For example, higher productivity hospitals might also have better non-health amenities like nicer lobbies, which would in turn influence hospital demand. Alternatively, high productivity hospitals could have better managers who improve both the production process and separately increase demand for the hospital.

As one highly imperfect and indirect way to gauge what may be driving the observed correlations between productivity and market share, we briefly examine how the magnitude of these static and dynamic relationships varies across hospitals and across markets. The results, which are shown in Appendix D (especially Table A5) are mixed. For example, within a market the allocation results are stronger for hospitals facing more competition for their patients (following Gaynor and Vogt, 2003's use of distance to nearest hospital as a proxy for hospital competition); however, at the market level there is no evidence that the allocation results are stronger for more competitive markets (following Syverson, 2004b) use of population density as a proxy for market competition for a spatially differentiated product. More work is clearly needed to establish to what extent the allocation and re-allocation to more productive hospitals is a direct result of competition or the result of other factors that are correlated with both productivity and demand.

2.5.2 Magnitudes

For many economic and policy questions, the mechanism by which market share is allocated to higher productivity firms is quite important. However, the exact mechanism is less important for forecasting whether and to what extent the market is evolving in a manner that favors higher productivity firms. Here, the magnitude of the productivity-market share relationships we estimate becomes important.

To begin to try to shed some light on these magnitudes, we investigate how a hospital's productivity correlates with its within-market growth and exit over longer horizons than the one-year horizon examined in Table 2.4. Specifically, we re-estimate equations 2.2 and 2.3 replacing the dependent variables $I\left[exit_{h,t+1}\right]$ and $\Delta_{h,t,t+1}$ with $I\left[exit_{h,t+K}\right]$ and $\Delta_{h,t,t+k}$, respectively.

Table 2.5 shows the results. The first row shows the allocation relationships one year out (i.e. the results from Table 2.4 where $k = 1$), and the subsequent rows show results up to 10 years out ($k = 10$). The relationship between productivity and growth or exit strengthens in absolute value over time. For example, a 10% increase in hospital productivity is associated with about 1 percent more patients next year, 4 percent more patients in 5 years, and almost 6 percent more patients in ten years.[20]

As another way to provide a sense of magnitude, we calculate the market re-allocation associated with a standard deviation change of productivity. Our baseline estimate of the national standard deviation of

[20] Because our data on growth and exit ends in 2007, as k rises, a smaller sample of hospital-years is available for these analyses. We verified that the finding that these relationships strengthen over time also holds (with quite similar magnitudes) if we restrict our sample to the hospital-years for which we observe at least 10 years of subsequent growth data (not shown).

hospital productivity is 0.17.[21] Thus a hospital that has one standard deviation higher productivity has about 40 percent higher market share at a point in time, and grows about 6 percent more over the next five years.

On the other hand, many other factors besides hospital productivity create the observed variation in market share. We estimate a partial R^2 on productivity in the static allocation regression (equation 2.1) of about 5 percent, and in the growth regression (equation 2.3) of about 0.06 percent. Of course, noise in our productivity estimate causes us to understate the ability of (precisely measured) productivity to explain market share.

As a final way to provide a sense of the magnitudes of these relationships, we compare them to those in other industries. To do so, we produced estimates of the static and dynamic allocation analyses for the ready-mixed concrete industry, which produces a physically homogenous product. Details on the data, estimation and results can be found in Appendix D. Like healthcare, concrete is consumed and produced locally, so that spatial differentiation (i.e. physical distance) can be an important barrier to competition. Otherwise, however, concrete lacks many of the features deemed to be important impediments to competition in healthcare: prices are not set administratively, consumers are likely well informed about their choices, and they bear the financial consequences of their decisions.

Across all of our static and dynamic allocation measures, the results indicate a stronger (often an order of magnitude larger) relationship between producer productivity and market allocation for hospitals than for concrete plants. Likewise, Figure 2-1 shows that national productivity dispersion appears larger for concrete than for hospitals; we estimate a standard deviation of 0.25 in concrete, compared to 0.17 for hospitals.[22]

This comparative finding is not limited to concrete. The static and dynamic allocation analyses are not easily comparable to pre-existing estimates in other sectors. However, productivity dispersion in other U.S. manufacturing industries also tends to be similar to (indeed, somewhat larger than) our estimates for healthcare.[23]

[21] Appendix D (especially Table A6) presents the dispersion estimates and also shows that they are quantitatively stable across alternative sets of risk adjusters.

[22] We follow the tradition of the existing productivity literature and compute productivity dispersion metrics at a nationwide (within-year) level, even though the market for treating heart attacks is (like many of the manufacturing industries studied) plainly local. This standard practice arose in part because manufacturing industries, the focus of the previous literature, are often geographically broad. But the literature has also typically reported nationwide numbers even for those industries that are more locally oriented, such as ready-mix concrete (Syverson, 2004b), in part because geographic differentiation is itself one of the possible causes of productivity dispersion within an industry. In practice, we find within-market year dispersion to be only slightly lower (standard deviation about 0.16) than our national dispersion estimate. Put another way, we estimate that about 88 percent of the within-year variation in hospital productivity is within (rather than across) markets. For concrete, we estimate that about 70 percent of the variation in productivity is within market.

[23] Compared to our estimate of a standard deviation of hospital productivity of 0.17, Foster et al. (2008b) estimate an average within-industry standard deviation of productivity of 0.22 across a dozen manufacturing industries in the US selected for having physically homogeneous products (e.g. white pan bread, block ice, raw sugar cane, etc.); Bartelsman et al. (2013) estimate an average within-industry standard deviation of 0.39 across a broader range of manufacturing industries. Across 450 different narrowly defined (4-digit SIC code) US manufacturing industries, Syverson (2004a) estimates an average within-industry interquartile range of logged plant productivity of 0.29, compared to our estimate in Table A6 of 0.23 for hospitals. Although most of the work in productivity dispersion has focused on the manufacturing sector, the more limited work on productivity dispersion in service industries suggests that in general it is roughly similar to that found in manufacturing. For example, Fox and Smeets (2011a) estimate productivity dispersion in four Danish service industries and four Danish

We are not the first to perform such cross-industry comparisons in productivity dispersion. For example, looking across narrowly defined manufacturing industries, Syverson (2004a) finds that the extent of within-industry productivity dispersion is negatively correlated with proxies for the amount of substitutability or competition across firms within that industry. We caution, however, against drawing inferences about the extent of competition in such different settings as heart attack treatment and manufacturing from comparisons of productivity dispersion. Basic measurement differences - such as differences in the output definition (survival vs. revenue), how inputs are measured, and estimation error - raise real comparability concerns, albeit without creating a clear direction of bias.[24] Moreover, as noted earlier, the causal force behind reduced dispersion is unclear, and need not be competitive pressure.

Nonetheless, at a broad level, the comparison may serve as a useful benchmark against which to assess the quantitative relationships we have estimated for productivity and allocation in the US healthcare sector. They also seem inconsistent with the conventional wisdom that the variations in inputs across areas and hospitals without concomitant output gains are unique to healthcare and must therefore result from idiosyncratic features of the sector.

2.6 Robustness

We explored the robustness of our allocation and dispersion findings along a number of dimensions and were generally quite reassured by the results. Here, we briefly describe some of our robustness analysis concerning risk adjustment, measurement of inputs, measurement of output, and potential endogeneity of inputs. Appendix E presents the results in more detail.

2.6.1 Controls for Patient Health

A key concern is whether we have adequately controlled for patient characteristics that are correlated with both hospital choice and survival. We have already seen that our core results are robust to controlling for fewer observable characteristics than in our baseline specification; specifically all of our tables have shown results with no patient covariates and with only covariates for age/race/sex interactions, in addition to the "full" set of demographics and co-morbidities. In addition, for one year of our sample we have access to considerably richer data that are abstracted from patients' medical charts and contain many additionally

manufacturing industries and find generally comparable estimates. Similarly, looking at 4-digit retail industries, Foster et al. (2006) estimate an average interquartile range for logged labor productivity which is comparable to Syverson (2004a)'s estimate of the interquartile range for logged labor productivity in manufacturing.

[24] To take but one example, the extent of measurement error in output - which would serve to attenuate estimates of the correlation between productivity and market share and to increase estimated dispersion - is likely different in healthcare than in manufacturing, although the sign of the difference is unclear. On the one hand, AMI survival is an accurately recorded account of output, in contrast to manufacturing revenue which could be reported with error and may confound output variation with price variation (see Foster et al., 2008b and Foster et al., 2012). On the other hand, in manufacturing industries output is more-or-less a deterministic function of inputs, while survival in our setting is stochastic. As discussed, we use the empirical Bayes "shrinkage" estimator to try to adjust for this stochastic element and the relatively small sample size within hospital-years.

relevant clinical characteristics such as test results and medical histories. We find that our results are not sensitive to including this more extensive set of controls (see Table A8).

2.6.2 Input Measure

We face several key choices with the construction of our input measure. One is how coarsely or finely to measure inputs. There is a tradeoff between our relatively coarse baseline measure of inputs (with its associated measurement error stemming from input variation that we do not capture) and more granular measures which suffer from potential survivorship bias (a patient cannot receive many procedures if she does not survive very long); we experimented with considerably more granular input measures based on the individual procedures received and the length of hospital stay. We also explored using these inputs directly in a multi-input production function rather than aggregating them to a single index as in our baseline approach. Finally, our baseline measure follows standard practice and defines inputs based only on hospital inpatient treatments, thereby excluding physician inputs - which may occur both inside and outside the hospital - and other outpatient inputs. We tried an alternative input measure that incorporates non-hospital inputs. Again there is a trade-off; some non-hospital inputs may be closely linked (or indeed part of) the care received in the hospital, while others may be quite distinct. These alternative input measures are each described in more detail in Appendix B and the results are presented in Table A9.[25]

2.6.3 Time Horizon

Another issue concerns the time horizon over which we measure inputs and outputs. Our baseline measures use a 30 day window for inputs and a 1 year window for output (survival days). We explored the robustness of our results to shorter and longer time horizons - 7 days and 1 year on the input side, and 30 days and 5 years on the output side. Again, there are tradeoffs in the length of time horizon.[26] We find our results are generally robust to these alternative input and output horizon windows (Table A10).

[25] Estimation in more traditional settings must also deal with input measurement problems, including issues we do not confront here stemming from differential qualities across types of workers and capital, trying to capture the flow of capital services using measures of capital stocks, and intermediate inputs typically measured by expenditures rather than quantities. Additionally, and more directly to the issue here, these inputs must also be aggregated to a single-dimensional input index by weighting the individual inputs appropriately. The theoretically correct weights are the elasticities of output with respect to the respective inputs. Estimating these elasticities involves its own set of measurement challenges. Our approach in the hospital sector avoids many of these additional issues.

[26] On the input side, a shorter time horizon will miss some of the resources the patient receives, while a longer horizon creates greater scope for survival bias as well as treatments that are linked to providers other than the original hospital. On the output side, for our baseline measure we chose the relatively standard 1-year horizon since it seemed substantively more of interest than shorter-term (e.g. 30 day) survival. Analysis of a shorter horizon might capture aspects of hospital productivity that reflect only a slight postponement in death, and might not capture aspects that affect outcomes through long-term mechanisms such as the management of complications due to co-morbidities and the quality of the hospital's follow-up care. On the other hand, with a longer output horizon there is greater scope for the impact of non-hospital factors - such as patient compliance in terms of diet, smoking and medication, and the impact of doctor quality regardless of whether the doctor was associated with the initial hospital - on our productivity estimates.

2.6.4 Potential Endogeneity of Inputs

Finally, as noted earlier, a pervasive concern in the productivity literature is the potential endogeneity of inputs to producer productivity. This can bias the estimates of the returns to scale parameter μ. There is a wide range of estimates of this parameter in the literature (see e.g. Cutler et al., 1998; Fisher et al., 2003b; Baicker and Chandra, 2004b) and uncertainty as to the "right" estimate. We are therefore reassured that our results are quite robust to imposing (rather than estimating) a range of "reasonable" values of μ and then calculating productivity under different imposed values (see table A11). The lack of sensitivity of our static and dynamic allocation results to alternative values of μ is consistent with the results in Figures 2-5a and 2-5b that the correlation between market share and estimated productivity is driven primarily by the correlation between market share and risk-adjusted survival.[27]

2.7 Conclusion

This paper has examined the relationship between productivity and market allocation in healthcare, specifically for hospital treatment of Medicare patients' heart attacks. We have done so by drawing on the insights of several decades of theoretical and empirical work in productivity more broadly. Qualitatively, we find that higher productivity hospitals have greater market share at a point in time and grow more over time. Quantitatively, a hospital with a one standard deviation higher productivity has about 40 percent higher market share at a point in time, and grows about 6 percent more over the next five years.

These relationships, which are driven primarily by the relationship between risk-adjusted survival and market share, mean that over time the healthcare market evolves in a manner favorable to higher productivity producers. This qualitative pattern is generally viewed by the broader productivity literature as an empirical sign of the workings of competition; it has been consistently found within manufacturing industries in the United States but not in less competitive settings such as post-Soviet Eastern block countries or Chile prior to trade reforms. Our more speculative quantitative comparisons between healthcare and manufacturing industries in the US suggest that, if anything, these re-allocation results are stronger, and dispersion similar or smaller, in healthcare.

Taken together, our qualitative and quantitative findings indicate that the healthcare sector may not be as idiosyncratic as the conventional wisdom has claimed. In this sense, our results are in the same spirit as Skinner and Staiger (2007b)'s finding of a common "innovativeness factor" across healthcare and other sectors within a geographic area; they found that areas of the country that were early adopters of hybrid

[27] Referring back to the basic estimating equation for hospital productivity (equation 2.5), the fact that the market share-productivity covariance is not sensitive to μ must mean that there is little variance in risk-adjusted inputs and/or a low covariance between risk-adjusted inputs and market share - otherwise, changes in the value of μ, which ties risk-adjusted input variation to our estimate of hospital's productivity levels, would change the correlation between estimated productivity and market share.

corn in the 1930s and 1940s were also early adopters of beta blockers for heart attacks at the beginning of the current century.

Such findings suggest that, going forward, research on the determinants of productivity in the health care sector may benefit from more attention to the insights, both theoretical and empirical, from research about productivity and allocation in other industries. By the same token, insights from the health care sector may likewise be a useful laboratory for thinking about other industries. A recent series of papers by Bloom, Van Reenen and co-authors have begun to do just this, empirically investigating the role of such factors as management style and labor quality on hospital performance (usually survival rates; see Bloom et al., 2010; Propper and Reenen, 2010; Bloom et al., 2012a; McConnell et al., 2013a). Moreover, in our healthcare setting as in the manufacturing setting more broadly, the estimated re-allocation relationships stop far short of indicating what economic or policy forces could be unleashed to create still greater reallocation to higher productivity producers. We see a great opportunity for further work that tries to estimate the causal impact of competition - or other factors - on allocation in healthcare and in manufacturing settings.

Of course, a given amount of re-allocation to higher productivity producers - or a given improvement in this re-allocation process - may be much more valuable in healthcare than in manufacturing, not to mention of greater consequence for public sector budgets. In this case, more than healthcare having different market dynamics, perhaps it is this feature of healthcare that makes it exceptional.

Figures

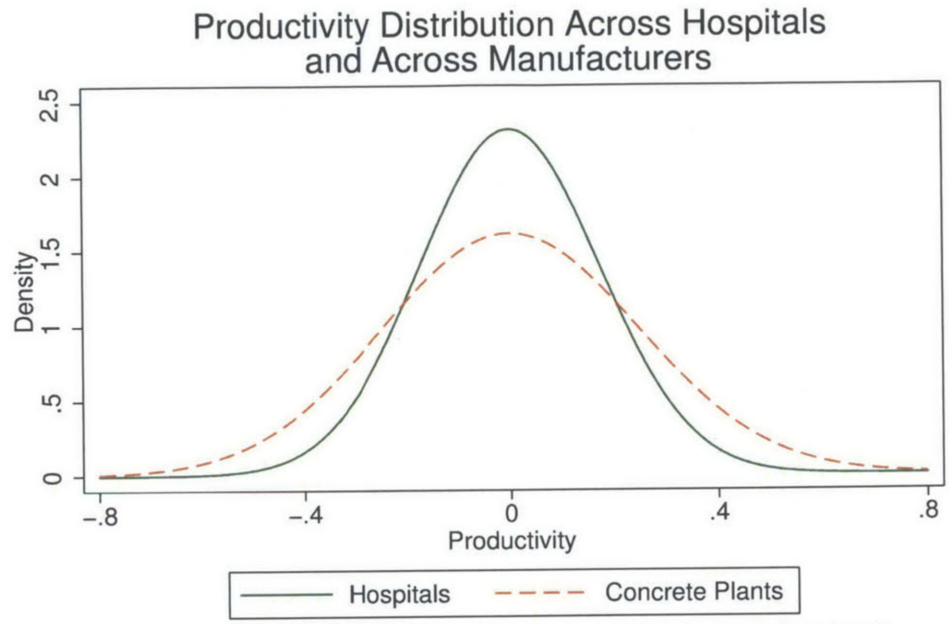
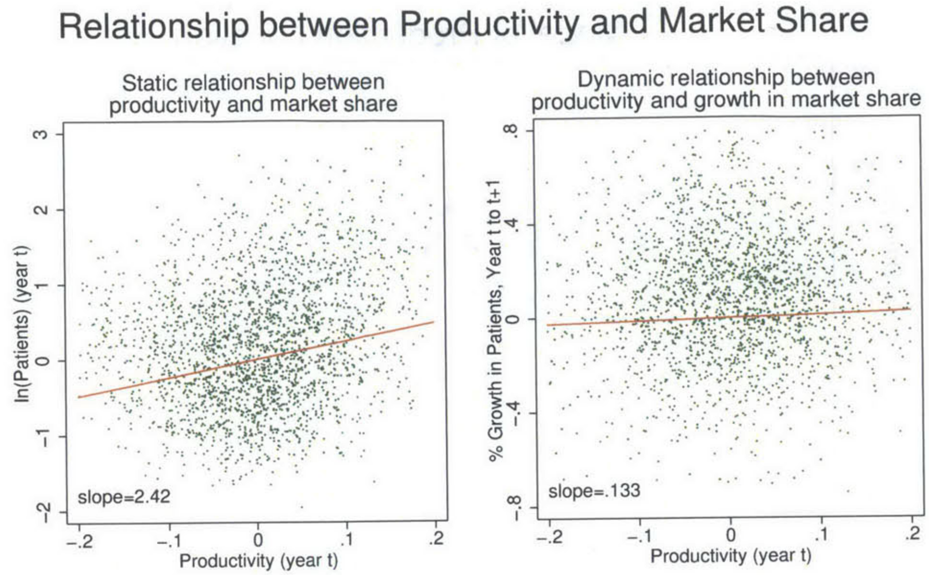


Figure shows estimated productivity dispersion across hospitals for heart attack treatments and across concrete plants for the production of ready−mixed concrete. We show the average within−year fitted normal density for each. Hospital productivity estimates (which reflect the hospital's ability to produce patient survival given a fixed set of inputs), are from our baseline specification (Table 2.2, column 1); concrete productivity estimates are from Table A7. See text for more details on the construction of these estimates.

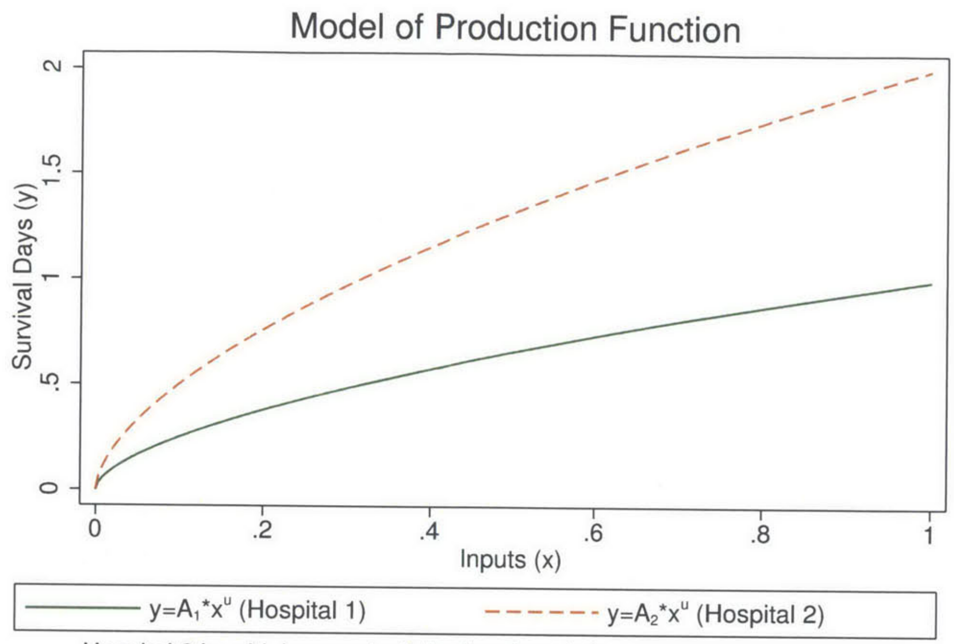
Figure 2-1



(2–2a) (2–2b)

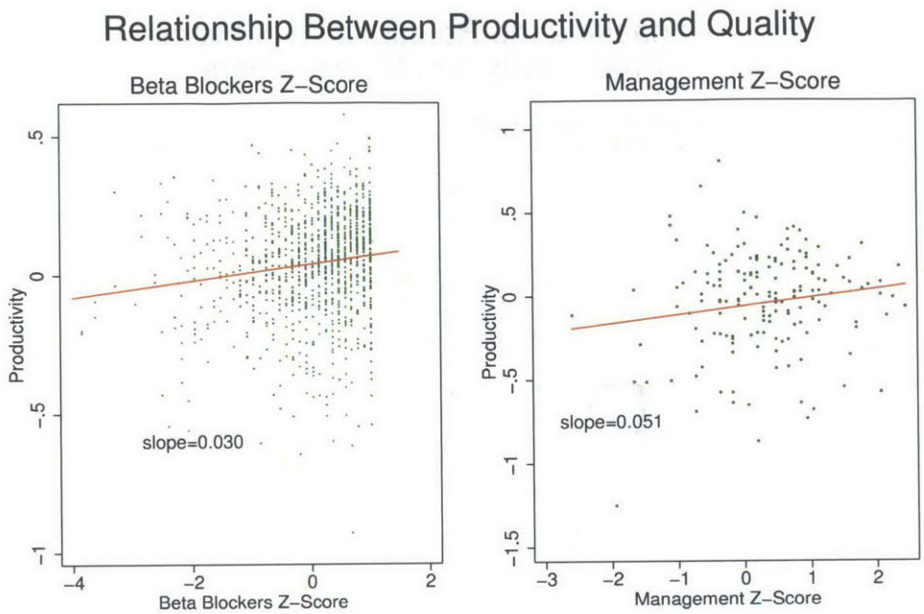
Figure shows relationship between a hospital–year's market share and productivity after partialing out market–year fixed effects. Figure 2–2a shows the static relationship between the hospital's log number of heart attack patients in year t and estimated productivity in year t; Figure 2–2b shows the dynamic relationship between the hospital's percent growth in heart attack patients between year t and t+1 (defined in equation 6) and estimated productivity in year t. Hospital productivity estimates (which reflect the a hospital's ability to produce patient survival given a fixed set of inputs) are from our baseline specification (Table 2.2, column 1). Figures show results for a random 5% of hospital–years, with hospital–years that have less than 11 patients suppressed form the scatter for confidentiality reasons. In addition, in Figure 2–2b for visual clarity the y–axis is restricted to the almost 95% of hospital–years with residual growth between –0.8 and 0.8. In both graphs, line shows the linear fit based on the whole sample (prior to any suppression).

Figure 2-2



Hospital 2 has higher productivity than hospital 1.

Figure 2-3



(2–4a) (2–4b)

Figure plots the relationship between our estimate of 2003 hospital–year TFP (from our baseline specification in Table 2.2, column 1, but without the empirical Bayes adjustment) against specific observable measures of hospital quality. Left hand panel plots relationship between the hospital's TFP and its beta–blockers z–score in 2003 for the 1,045 hospitals where we observe both (6 hospitals with outlying z–scores are not shown). Right hand panel shows the relationship between the hospital's 2003 TFP and its management z–score for the 179 hospitals where we observe both. See text for more detail on both of these z–scores. Hospitals that have less than 11 patients in 2003 are suppressed from the scatter for confidentiality reasons. Line shows the linear fit based on the whole sample (prior to any suppression and removal of outliers).

Figure 2-4

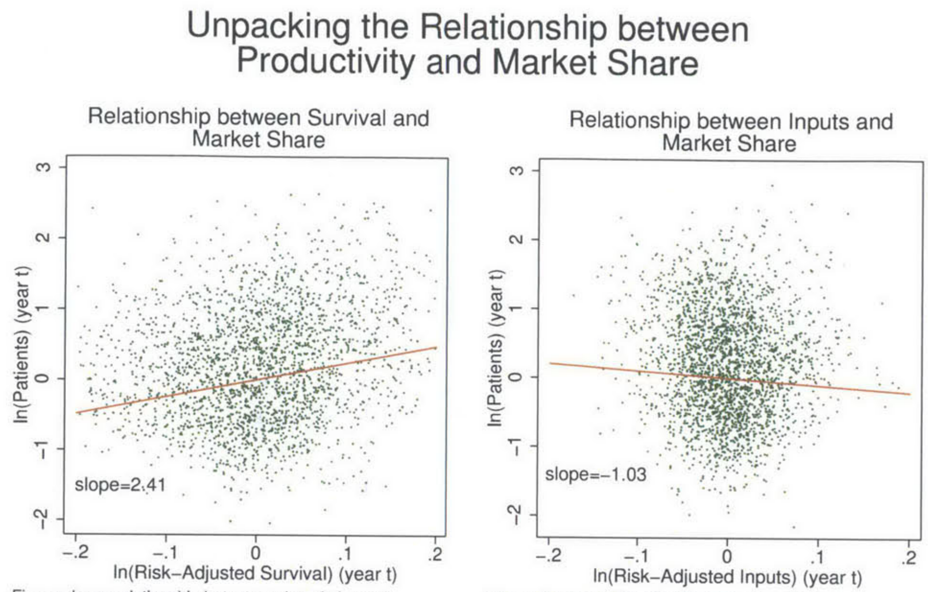


Figure shows relationship between a hospital-year's market share and risk-adjusted survival after partialing out market-year fixed effects and risk-adjusted inputs. Y-axis is the log number of heart attack patients in year t; x-axis is the hospital's risk-adjusted average log-survival in year t. Baseline risk adjusters (shown in Table 2.1b) are used. Figure shows results for a random 5% of hospital-years, with hospital-years that have less than 11 patients suppressed from the scatter for confidentiality reasons. For visual clarity, the x-axis is restricted to the 97% of hospital-years with residual survival between -0.2 and 0.2. Line shows the linear fit based on the whole sample (prior to any suppression or restriction).

(2-5a)

Figure shows relationship between a hospital-year's market share and risk-adjusted inputs after partialing out market-year fixed effects and risk-adjusted survival. Y-axis is the log number of heart attack patients in year t; x-axis is the hospital's risk-adjusted average log-input in year t. Baseline risk adjusters (shown in Table 2.1b) are used. Figure shows results for a random 5% of hospital-years, with hospital-years that have less than 11 patients suppressed from the scatter for confidentiality reasons. For visual clarity, the x-axis is restricted to the 99.9% of hospital-years with residual inputs between -0.2 and 0.2. Line shows the linear fit based on the whole sample (prior to any suppression or restriction).

(2-5b)

Figure 2-5

Tables

Table 2.1a - Hospital and market statistics

	(1) Mean	(2) SD	(3) Min	(4) Max
Hospital-Years (N=55,540)				
Patients	63.57	69.63	5	917
Market-Years (N=4,560)				
Patients	774.2	735.2	63	5,700
Hospitals	12.18	11.38	1	97

Note: The number of hospitals is 5,346.

Table 2.1b - Patient Summary Statistics

	(1) Mean	(2) SD
Outputs		
Survival (days; censored at 365)	268.1	149.4
Binary: Survival > 365 Days	0.660	0.474
Inputs		
Baseline (30 day) input measure ($)	15,996	12,172
Risk Adjusters		
Age	78.17	7.546
Female	0.507	0.500
White	0.906	0.291
Hypertension	0.207	0.405
Stroke	0.0232	0.150
Cerebovascular Disease	0.0398	0.195
Renal Failure	0.0521	0.222
Dialysis	0.00670	0.0816
COPD	0.0981	0.297
Pneumonia	0.0592	0.236
Diabetes	0.128	0.334
Protein Cal Malnut	0.0118	0.108
Dementia	0.0412	0.199
Paralysis/FD	0.0256	0.158
Periph Vasc Disease	0.0639	0.245
Metastatic Cancer	0.0117	0.107
Trauma	0.0392	0.194
Substance Abuse	0.0225	0.148
Major Psych Disorder	0.0138	0.117
Chronic Liver Disease	0.00281	0.0529

Note: The number of observations is 3,530,401.

Table 2.2 - Production Function Parameter Estimates

	(1)	(2)	(3)
Risk Adjustment:	Baseline	Age/Race/Sex	None
Parameter			
μ	0.446	0.481	0.589
	(0.00511)	(0.00523)	(0.00552)

Notes: N = 3,530,401 patients, 55,540 hospital-years, and 5,346 hospitals. Standard errors are bootstrapped with 300 replications and are clustered at the market level (304 markets). "Baseline" risk-adjustment includes a full set of interactions between age (in five year groupings), gender and whether the patient is white; it also includes indicators for the various co-morbidities shown in Table 2.1; column 2 excludes the co-morbidities and column 3 has no risk adjusters.

Table 2.3 - Relationship Between Hospital TFP and Hospital Covariates

	(1) ln(Prod)	(2) ln(Prod)	(3) ln(Prod)	(4) ln(Prod)	(5) ln(Prod)	(6) ln(Prod)
Beta-Blockers Z-Score	0.0299					
	(0.00667)					
Composite Z-Score		0.0299				
		(0.00676)				
Management Z-Score			0.0511			
			(0.0290)			
Teaching Hospital				0.0799		
				(0.0129)		
Urban					0.0696	
					(0.0160)	
For-Profit						0.0228
						(0.0266)
Non-Profit						0.101
						(0.0221)
Constant	0.0382	-0.0147	-0.0609	-0.0810	-0.102	-0.127
	(0.00577)	(0.00590)	(0.0248)	(0.00890)	(0.0144)	(0.0202)
Observations	1,045	2,183	179	3,361	3,361	3,363

Unit of observation is a hospital. Dependent variable is our estimate of 2003 hospital-year TFP from our baseline specification (Table 2.2, column 1) without empirical Bayes adjustment. Right hand side variables in columns 1 through 3 are z-scores for hospitals that reported the measure indicated. Data on beta-blockers and composite scores are from CMS Hospital Compare; beta-blockers score includes hospitals with at least 30 patients appropriate for the treatment, while composite score includes hospitals with a sum of at least 30 patients appropriate for each of the treatments within the score. Data on management score are based on a 2010 survey of management practices adminstered by Bloom et al. (2012); see text for more details. Right hand side variables in columns 4 through 6 are indicators for whether the hospital is a teaching hospital (Column 4), in an urban area (Column 5), or is a for-profit or non-profit entity (Column 6, public is the omitted category). Indicators for hospital characteristcs are coded from CMS Provider of Services and Impact files; we define a teaching hospital as one that has residents. Standard errors are bootstrapped with 300 replications and are clustered at the market level.

Table 2.4 - Main Results - Allocation Metrics

	(1)	(2)	(3)	(A)	(B)
Risk Adjustment:	All	Age/Race/Sex	None	DV Mean[a]	Observations
Static Allocation	2.418	2.496	2.618	3.641	55,540
	(0.0889)	(0.0851)	(0.0779)		
Dynamic Allocation					
Exit Regression	-0.0329	-0.0353	-0.0458	0.0438	40,379
	(0.00935)	(0.00884)	(0.00766)		
Growth Regression	0.133	0.154	0.201	-0.126	52,777
	(0.0225)	(0.0214)	(0.0184)		

Notes: "Static Allocation" reports the results from estimating the relationship between a hospital's log(patients) and TFP (i.e. productivity) within a market year given by equation (1). "Exit regression" reports the results from estimating the within-market relationship between a hospital "exit" as defined in the text and last year's productivity as given by equation (2). "Growth regression" reports the results from estimating the within-market relationship between a hospital's one-year percent growth and its base year productivity as defined in equation (3). Productivity is estimated based on the corresponding specifications from Table 2.2. Standard errors are bootstrapped with 300 replications and are clustered at the market level. Observations are hospital-years.

[a]"DV mean" reports the mean of the dependent variable for the regressions, which is ln(Patients) for the static allocation regression, 5-year exit for the exit regression, and 1-year growth for the growth regression. See text for more detailed definitions of dependent variables.

Table 2.5 - Dynamic Allocation Varying Time Horizons

	Growth from t to $t+k$				Exit in $t+k$			
Years (k)	Coeff	Std Err	Mean DV[a]	Obs	Coeff	Std Err	Mean DV[a]	Obs
1	0.133	(0.022)	-0.126	52,777	-0.033	(0.009)	0.044	40,379
2	0.207	(0.027)	-0.224	49,954	-0.056	(0.014)	0.077	36,864
3	0.270	(0.038)	-0.314	46,961	-0.085	(0.019)	0.108	33,163
4	0.345	(0.047)	-0.392	43,742	-0.122	(0.023)	0.137	29,338
5	0.365	(0.052)	-0.462	40,379	-0.147	(0.028)	0.166	25,359
6	0.397	(0.062)	-0.530	36,864	-0.165	(0.030)	0.195	21,320
7	0.477	(0.068)	-0.598	33,163	-0.203	(0.037)	0.226	17,226
8	0.526	(0.070)	-0.666	29,338	-0.224	(0.040)	0.255	13,050
9	0.573	(0.074)	-0.735	25,359	-0.242	(0.049)	0.284	8,761
10	0.587	(0.077)	-0.807	21,320	-0.212	(0.060)	0.313	4,412

These results report the coefficient and its standard error from the regressions of growth or exit on productivity, controlling for market-year fixed effects. These are modified versions of equations (2) and (3) where the time horizon over which growth or exit is measured is now k years rather than 1 year. Each row considers a different time horizon k. Longer horizons have smaller samples because data on growth ends in 2007 and data on exit ends in 2003. Standard errors are bootstrapped with 300 replications and are clustered at the market level.

[a] "Mean DV" refers to the mean of the dependent variable (growth or exit) in the sample over the time horizon indicated.

Chapter 3

Do Incentives Affect Treatment Decisions in the Hospital? Evidence from a Medicare Reform*

3.1 Introduction

A central question in public policy is how to set prices optimally in regulated industries. The health care sector is perhaps the most salient among such industries, accounting for nearly one-fifth of output in the United States. Price-setting offers a lever to improve the efficiency of this sector; depending on how health care providers respond to alternative payment regimes, properly set prices could reduce wasteful spending and nudge providers to practice more cost-effectively (McClellan, 1996). How providers alter their treatment patterns in response to changes in relative payments – and how patient outcomes are affected in turn – is therefore a key topic of both academic research and a significant parameter for health policy (Cutler, 1995; Gilman, 2000; Ellis and McGuire, 1996).

I utilize a recent Medicare payment reform to isolate shocks to payments to hospitals for treating lung cancer patients. These shocks affected the relative reimbursement rates for the three broadly defined treatment approaches that Medicare recognizes for these patients. By exploiting exogenous variation in relative prices, I am able to estimate the substitution matrix of treatment approaches with respect to relative treatment

*I am grateful to Amy Finkelstein, Michael Greenstone, and Jon Gruber for their advice and guidance on this project. I thank Isaiah Andrews, David Chan, Arun Chandrasekhar, Manasi Deshpande, Kate Easterbrook, Ben Feigenberg, Matt Fiedler, Eliza Forsythe, Greg Leiserson, Conrad Miller, David Molitor, Iuliana Pascu, Maxim Pinkovskiy, Maria Polyakova, Miikka Rokkanen, Annalisa Scognamiglio, Brad Shapiro, Mark Shepard, Henry Swift, Melanie Wasserman, and participants in the Harvard Works in Progress Seminar and MIT Public Finance lunch for their comments and suggestions. I would also like to thank Jean Roth for her assistance with all aspects of the Medicare data. I gratefully acknowledge funding from the National Institute on Aging grant T32-AG000186.

prices – the full set of own- and cross-price elasticities. I find estimates of these elasticities that are right-signed and economically meaningful. For example, I show that increasing the relative reimbursement for minor surgery (the treatment approach of intermediate intensity) by $1,000 would increase the surgery rate by about 1.6pp, with patients drawn partly from medical management and partly from major surgery (the low and high intensity approaches, respectively) – though this estimate is marginally insignificant at the 10% level. This variation also allows me to explore whether the changes in treatment patterns had effects on patient outcomes like readmission and mortality rates, though issues with statistical power make it difficult to reach conclusions in these specifications.

This research speaks to several open questions in health economics and policy. First, though it may be *a priori* obvious that hospitals respond to prices, the scale of that response is not at all clear from theory. I show that the response appears to be substantial. Second, the own- and cross-price elasticities are highly policy relevant, relating to the potential effects of Accountable Care Organizations (ACOs) and other proposed reforms to payment systems. In a policy simulation, I show that a reform that eliminates incentives for treatment at the margin would reduce treatment intensity, though the magnitude of the reduction is measured with so much noise that it is not statistically significant. Finally, the health effects of manipulating treatment patterns with prices are ambiguous, but are clearly central in evaluating the costs and benefits of these policies. I attempt to isolate a measurable effect of the reform on health outcomes, but weak identification limits my power in these specifications.

As public and private insurers have increasingly focused on paying for quality and value rather than volume, the effects of powering down incentives for intensity have become a key policy input. Public agencies play an outsize role in price-setting, controlling payments to the health care providers of the millions who receive insurance through government programs and providing a baseline against which private insurers set their own policies (Clemens and Gottlieb, 2013). The Affordable Care Act (ACA) has placed a renewed emphasis on using payment policy to improve the efficiency and value of the US health care system. The results from this study provide important evidence to predict some of the effects of reform law projects like ACOs, which lower the incentives for treatment intensity at the margin.

At the same time, insurers are considering refinements to the existing prospective payment systems which pay health care providers based on the expected but not realized costs of treatment. Real-world prospective payment systems deviate from full prospectivity by also allowing payments to depend on the broadly defined treatment approach chosen by the provider, but the granularity of the treatment approaches differs from system to system. Insurers – and state Medicaid programs in particular – frequently consider switching to payment systems with more narrowly defined categories of treatment. This study provides evidence that changes in incentives due to treatment category refinements will influence treatment patterns.

I exploit variation in reimbursements due to a 2008 Medicare hospital payment reform.[1] I focus on lung

[1]All years are federal fiscal years unless otherwise noted. Federal fiscal years start on October 1 of the preceding calendar year. Therefore the FY2008 Medicare reform took effect on October 1, 2007.

cancer because the Medicare payment system recognizes just three treatment approaches for these patients, reducing the number of own- and cross-price elasticities to be estimated. Figure 3-1 shows the evolution of reimbursement for major treatment and minor treatment relative to medical management due to the reform – the key variation that the study will exploit. The reform changed how payments to hospitals were adjusted for the severity of illness of patients – conditional on treatment approach, the reform tended to raise payments for treating relatively sick patients and reduce payments for treating relatively healthy patients. Crucially, these changes differed by treatment approach.

The econometric strategy that I utilize is an instrumental variables differences-in-differences approach. To address concerns about the endogeneity of prices, I construct instruments for the returns to treatment approaches. Research has demonstrated that hospitals respond to severity adjustment by providing better documentation of high-severity conditions for marginal patients (Dafny, 2005a; Sacarny, 2014). Documentation and coding improvements induce a correlation between the return to treatments following the reform and a patient's unobserved appropriateness for those treatments. My instruments take advantage of persistent differences across hospitals in the average severity of illness of their patients. In this strategy, the instruments are hospital-level predictions of the returns to treatment approaches. To eliminate the possibility that endogenous documentation improvements drive variation, the predictors are constructed using patients in years prior to the analysis sample; these patients are re-priced under the payment rules of the analysis sample.

The differences-in-differences strategy exploits within-hospital over-time variation in the instruments; this variation is driven by base year variation in patient severity of illness across hospitals interacted with the changes in the price of severity induced by the reform. This strategy makes two strong assumptions: that the sorting of patients to hospitals along the severity dimension does not change systematically due to the reform, and that hospitals do not respond to the reform by adjusting their propensity to admit patients to the inpatient setting. I perform specification checks to test whether these margins were operative, but the checks are inconclusive. The four specifications mostly fail to reject a null consistent with instrument exogeneity, but do so with low statistical power.

This work adds a new perspective to existing literature studying the effects of prospective payment reforms on how patients are treated in the hospital and how their outcomes are subsequently influenced. Cutler (1995), perhaps the most notable work of this literature, finds that the Medicare Inpatient Prospective Payment System (IPPS) had slight effects on the timing of death, and bases its hypothesis on a view of IPPS as reducing marginal incentives for treatment intensity across the board. Ellis and McGuire (1996), similarly, derive and test predictions of prospective payment assuming it eliminates incentives for intensity of treatment.

In contrast, I follow Gilman (2000) in considering the implications of the partial non-prospectivity of the IPPS. Gilman looks at the introduction of new payment categories for HIV inpatients covered by Medicaid in New York state in the mid 1990s. The New York reform recognized certain surgical approaches as distinct payment categories, creating a return to surgery; the study estimates that the partial effect of the incentives for treatment at the margin was an increase in treatment intensity, though other incentives created by the reform made the total effect of the policy on intensity negative. More recently, Dafny (2005a) uses a reform to the Medicare inpatient PPS severity adjustment system as a shock to the price of different treatments, but finds no evidence that hospitals alter their "real" treatment patterns in response.

I differ from earlier work by focusing directly on the choice of treatment approach that the patient receives, estimating the full substitution matrix of own- and cross-price elasticities over the approaches; to do so I limit my analyses to lung cancer patients, for whom the treatment approaches are limited and the estimation therefore tractable. This analytical approach is novel in the literature and I have heretofore been unable to find any economic work that has used it in the context of prospective payment.

I proceed as follows: in section 3.2, I provide an overview of prospective payment and the 2008 reform and discuss the data. In section 3.3, I explain the instrumental variables differences-in-differences approach and its identification assumptions. Section 3.4 presents the results. In section 3.5 I present specification tests, and I conclude in section 3.6.

3.2 Background and Data

Lung cancer accounts for about one-sixth of cancer deaths worldwide and about 160,000 deaths in the United States annually (Hoffman et al., 2000). In my data, I observe about 90,000 hospital admissions each year for lung cancer among elderly Medicare patients. Patients with lung cancer may present in the hospital for a variety of reasons and their treatment may take courses of varying intensity. Treatment approaches include relatively minor non-operating room procedures like fine needle biopsies, which can be useful in the diagnosis of the cancer; more intensive diagnostic methods like mediastinoscopy, in which a diagnostic scope is introduced through a chest incision; and extremely invasive surgeries involving the opening of the chest cavity, such as the removal of an entire lobe of the lung. Treatment guidelines depend on the type of cancer suspected and are evolving as technology improves and new research is released (Goldstraw et al., 2011).

This study focuses on the interaction of two key aspects of payment for patients with lung cancer: treatment approach *intensity* and patient *severity* of illness. The former term refers to the level of resources that tend to be utilized in the treatment approach – a higher intensity approach involves more inputs, including time at the hospital, labor from nurses and physicians, and use of operating rooms. For example, treating a lung cancer patient outside the operating room with a fine needle biopsy is of much lower intensity than treating the patient with an operating room procedure like a lobectomy (the removal of a lobe of the lung).

Likewise, severity of illness refers to the patient's mortality risk and loss of function. Patients with higher severity, or equivalently with a greater level of illness, are likely to require greater hospital resources conditional on treatment approach (3M Health Information Systems, 2007). High severity conditions (as defined by Medicare's payment mechanisms) common among lung cancer patients include pneumonia and respiratory failure. These conditions are associated with greater resource utilization in the hospital. Common low severity conditions include hypertension and type II diabetes – patients with these conditions do not tend to have greater utilization than those without them.

The analyses in this study concern the incentives for different treatment approaches under the Medicare Acute Inpatient Prospective Payment System (IPPS), the $111 billion program that pays most hospitals in the United States for inpatient stays (MEDPAC, 2012a). A 2008 reform to the IPPS – the most substantial change to the program since it was created – refined how hospital payments depended on patient severity of illness. In this section, I explain how the IPPS generates a financial return to broadly defined treatment approaches and how this return was changed by the 2008 reform.

3.2.1 Prospective Payment and the Return to Treatment Approaches

Under prospective payment, hospitals are paid a fixed, largely predetermined sum for treating each patient. The word prospective is meant to indicate that payments are set *ex ante* independent of the intensity of treatment, in contrast to earlier retrospective systems in which facilities were fully reimbursed for all so-called "reasonable and customary" charges realized in the course of care. By reducing incentives for overtreatment at the margin, prospective payment offered the potential to improve the efficiency of the health care sector (Newhouse, 1996; Weisbrod, 1991).

Since 1984 Medicare has paid most acute care hospitals prospectively under the IPPS. The unit of payment in the IPPS is the diagnosis-related group (DRG). Each DRG is associated with a "weight", or resource intensity. To determine reimbursement, DRG weights are converted into dollars using a national conversion factor adjusted for attributes of the local labor market and certain hospital characteristics (Centers for Medicare and Medicaid Services, 2013). Since each hospital has one conversion factor, the relative prices of DRGs are the same across facilities.

The IPPS deviates from an idealized model lacking any incentive for treatment at the margin. DRGs are partitioned into two branches, medical and surgical. Medical DRGs classify patients who were treated without an operating room procedure using their primary diagnosis, the diagnosis ruled responsible for the patient's admission to the hospital. Surgical DRGs handle patients who received such a procedure and classify patients on the basis of the highest intensity procedure that was performed. Hospitals that opt to treat a patient with a qualifying procedure receive the surgical DRG payment rather than the medical DRG payment, which can create an incentive (or counter-incentive) to perform such procedures.

The presence of surgical DRGs helps to ensure that hospitals that treat patients relatively intensively are reimbursed their expected costs (McClellan 1996 discusses the implications of this aspect of the IPPS for productive efficiency). These categories are defined broadly, and within a category additional treatments do not generally result in greater reimbursements to the facility. In other words, the IPPS reimbursement depends on the treatment *approach*, but hospitals that treat intensively conditional on the approach will not systematically receive greater payments from Medicare. For example, for the vast majority of lung cancer patients, running additional laboratory tests and diagnostic scans will not raise the hospital's reimbursement regardless of the treatment approach that the patient receives.[2]

Return to Surgery for Lung Cancer Patients

This study focuses on patients with a primary diagnosis listed in the "Respiratory Neoplasms" medical DRG - a set of diagnoses covering benign and malignant cancers throughout the lung and other parts of the respiratory system. If these patients receive medical management or low-intensity non-operating room procedures, they fall into the Respiratory Neoplasms DRG and hospitals receive the payment associated with that DRG's weight.

However, when these patients receive operating room procedures, the IPPS classifies them into a surgical DRG. The vast majority fall into two such DRGs: "Other Respiratory Procedures", which I refer to as minor surgery, and "Major Chest Procedures", which I refer to as major surgery.[3] Patients in the minor surgery category receive procedures like mediastinoscopy and closed (i.e. not requiring chest incisions) biopsies. Procedures in the major surgery category include higher intensity operations like lobectomies and open (i.e. requiring chest incisions) biopsies. A patient who receives procedures in both categories is classified into the "major" DRG.

This arrangement of medical and surgical DRGs implies that the IPPS recognizes three key treatment approaches for lung cancer: medical management, minor surgery, and major surgery. Relative to medical management, the returns to major and minor surgery are given by the reimbursement for these surgeries less the reimbursement for medical management.

3.2.2 Patient Severity and Payment Reform

Much of the difficulty of designing health care payment systems centers on how to reduce the incentive for providers to cherry-pick inexpensive patients and how to ensure that providers that treat expensive patients are compensated sufficiently that they remain solvent. The IPPS addresses these concerns by adjusting its payments for patient severity of illness. Many DRGs "split" on the basis of severity, yielding a higher weight

[2]The exception to this rule comes from outlier payments, which reimburse hospitals for 80% of their costs above a fixed loss threshold. These payments act as insurance for hospitals against unusually expensive cases.

[3]There are three other categories of classification, accounting for 3% of lung cancer patients. The first contains patients who receive extraordinarily intensive procedures like lung transplants. The second includes patients receiving mechanical ventilation. The last comprises patients who have lung cancer but receive a major surgical procedure unrelated to the respiratory system. I exclude patients who fall into these categories from my analysis.

– and thus a higher reimbursement – for higher severity patients. Conditional on treatment approach, a hospital's reimbursement for a patient depends almost entirely on the patient's severity of illness.

A patient's severity is defined as the highest severity diagnosis that was submitted by the hospital on its insurance reimbursement claim (though the severity of the patient's primary diagnosis is not counted in the calculation). The severity system therefore comprises a mapping from the approximately 13,000 ICD-9 diagnosis codes to a set of severity levels.

The 2008 IPPS reform updated and refined the severity system to address policymakers' concerns about patient selection incentives and underpayment of providers. Prior to the reform, the IPPS recognized two levels of severity. 7% of lung cancer patients were classified in the low severity (called "no complication or comorbidity") category and 93% were high severity ("with complication or comorbidity"). With so many patients crowded into the upper category, the capacity for severity adjustment to improve the matching of reimbursements to expected costs was curtailed.

To address policymakers' concerns, the reform carved out an additional severity level for patients who were particularly ill and rewrote the mapping of diagnosis codes to severity levels, demoting many codes to low severity. The result was a more even distribution of patients into the categories: in the year after the reform, 13% of lung cancer patients were low severity ("no complication or comorbidity"), 49% were medium severity ("with complication or comorbidity"), and 38% were high severity ("with major complication or comorbidity"). With more granular categories, the IPPS was able to raise reimbursements for sicker patients while lowering them for healthier ones.[4]

Table 3.1 shows the levels of reimbursement for the three lung cancer treatment approaches and how they changed following the reform. Since the reform was phased in over 2 years, the table shows payment rates in 2007, the last year prior to the reform, and compares them to payment rates in 2009, the first year in which it was fully phased in. It shows that payments for treating the sickest patients rose substantially, while payments for treating relatively healthy patients fell.[5]

One can define the return to surgery for minor and major surgery as the payment for the surgical DRG less the payment for the medical DRG. For a patient p of severity $s(p)$ in year t, this return is:

$$ret^{minor}_{s(p)t} = payment^{minor}_{s(p)t} - payment^{medical}_{s(p)t} \tag{3.1}$$

$$ret^{major}_{s(p)t} = payment^{major}_{s(p)t} - payment^{medical}_{s(p)t} \tag{3.2}$$

[4]To reflect the role of severity in the DRGs, the name of the IPPS methodology was changed from CMS-DRGs (CMS, or Centers for Medicare and Medicaid Services, is the agency that administers the Medicare program) to MS-DRGs – Medicare Severity DRGs.

[5]While reimbursements for performing minor surgery on low severity patients ostensibly rose from $7,645 to $8,640, the low severity definition was expanded due to the reform. Many patients in the post-reform low severity category would have been categorized as high severity before the reform. Reimbursements for treating these patients therefore fell substantially.

The shocks to reimbursements differed substantially across treatment approaches and depended on patient severity. As a result, the returns to surgery given in equations 3.1 and 3.2 were also dramatically altered. Table 3.2 lists these returns before and after the reform, and figure 3-1 depicts the evolution of these returns for three types of patients: low severity (low before and after the reform), medium severity (high before the reform, medium after it), and high severity (high before and after the reform). By these definitions, 6% of patients were low, 49% were medium, and 35% were high. Most of the remaining patients were classified as medium severity before the reform and low severity after it.

The right pane of figure 3-1 shows that the severity adjustment raised the return to major surgery for the sickest patients and lowered it for the healthiest. The left pane tells a somewhat more complicated story – the return to minor surgery for the healthiest patients actually rose. This counterintuitive result occurred because the low severity category was expanded to include more patients, which meant including sicker and thus more costly patients. On the other hand, the return to minor surgery for medium severity patients fell, while the return rose for high severity patients.

3.2.3 Data and Summary Statistics

My study analyses the effect of incentives for treatment approaches using a dataset of hospital insurance claims for reimbursement from Medicare. The dataset is called the MEDPAR RIF, which is a 100% sample of these claims. The MEDPAR file includes all of the information needed to determine a patient's DRG, including the patient's primary and secondary diagnoses and the major procedures they received during their hospital stay. I use the data to construct two datasets: an analysis sample of patients in the years immediately before and after the reform, and an instrument sample of patients in earlier years.

Both the analysis and instrument samples are limited to lung cancer patients. To isolate these patients, I use a piece of software called a grouper that takes information about the reimbursement claim and maps it to the patient's DRG.[6] I define lung cancer patients as those who, when their surgical procedures are removed from the claim, are classified by the grouper into the Respiratory Neoplasms DRGs. These patients have a primary diagnosis – the diagnosis ruled responsible for their admission to the hospital – of benign or malignant cancer in the lung or other parts of the respiratory system.

I limit to short-term US hospitals that are paid according to the federal IPPS. Hospitals that are exempt from this system include Critical Access Hospitals, which are small rural facilities, and Maryland hospitals, which use a special PPS system administered by the state government; I exclude these facilities from the analysis. At the patient level, I limit to individuals who were at least 65 years old and whose stays were covered by Original (i.e. fee-for-service) Medicare, not private Medicare Advantage plans. The non-elderly covered by Medicare are a unique population and may not be comparable to the elderly. Medicare Advantage stays

[6] The grouper was purchased from `http://www.drggroupers.com`.

may be poorly observed in the data; in addition, these plans have latitude to use their own reimbursement methods, so incentives for treatment are not well measured for these stays.

In addition, because the econometric approach requires data about the patients at the hospital in a base period before the reform, I also limit to hospitals in which I observe lung cancer patients in 2003, 2004, and 2005. The lung cancer patients discharged in these years form the instrument sample, consisting of 288,791 patients discharged across 2,946 hospitals.

The analysis sample consists of lung cancer patients discharged from 2006 through 2009 – two years before and after the reform – that fit the listed criteria. The sample contains 337,198 patients across 2,946 hospitals. Summary statistics on these patients are given in Table 3.3. Lung cancer patients are 76 years old on average and about half are female. 57% are managed medically, 13% receive minor surgery, and 30% receive major surgery. By the pre-reform severity definitions 93% were high severity, but by the post-reform definitions only 35% were at the top level.

Outcomes for lung cancer patients in the analysis sample are generally poor. Almost 10% die in the hospital, over 30% die within 30 days of discharge, and nearly half die within 90 days of discharge. About one-fifth are readmitted to an inpatient facility within 30 days of discharge; that share increases to about one-third when the readmission window is expanded to 90 days.

3.3 Research Design

This study seeks to estimate the elasticities of treatment approaches with respect to reimbursement.[7] I specify the minor and major surgery treatment rates as linear in the returns to the treatments given in equations 3.1 and 3.2:

$$treat_{pt}^{minor} = \gamma_{own}^{minor} \cdot ret_{s(p)t}^{minor} + \gamma_{cross}^{minor} \cdot ret_{s(p)t}^{major} + \epsilon_{pt}^{minor} \tag{3.3}$$

$$treat_{pt}^{major} = \gamma_{cross}^{major} \cdot ret_{s(p)t}^{minor} + \gamma_{own}^{major} \cdot ret_{s(p)t}^{major} + \epsilon_{pt}^{major} \tag{3.4}$$

Where p indexes patients, t indexes years, and $s(p)$ is the patient's severity. $treat_{pt}^{X}$ indicates that the patient received treatment X, while ret_{st}^{X} is the return for treatment X in \$1000s relative to medical management. The γ_{own}^{X} coefficients are therefore the percentage point increase in the treatment rate of X following an increase in the return to treatment X of \$1000. The γ_{cross}^{X} coefficients are the percentage point increase in the treatment rate of X when the return to treatment $-X$ is raised by \$1000. Economic theory suggests that the own-reimbursement coefficients should be positive and the cross-price coefficients should be negative (under the assumption that they are substitutes).

[7] In section 3.4.4 I consider how my instruments can identify the effects of treatments on health outcomes.

Due to the design of the DRGs, the estimation of these equations raises significant endogeneity issues. In this section, I explain the sources of endogeneity and how my identification strategy attempts to resolve them. I also explain potential concerns with this strategy.

3.3.1 Sources of Endogeneity

There are two key omitted variables from the equations of interest that prevent a simple regression from revealing the causal effects of reimbursements on treatment and of treatment on outcomes: unobserved patient appropriateness for the treatments and endogenous exaggeration of patient severity.

The first unobservable in equations 3.3 and 3.4 is a measure of the patient's appropriateness for the treatment. In these equations, the source of variation in the return to treatment within a year is due to the patient's severity of illness $s(p)$. There is good reason to think that this severity is correlated with appropriateness: a patient's level of illness can affect the probability that she survives the surgery and the expected benefit of the treatment conditional on surviving the operation.

To provide some suggestive evidence on differences in appropriateness, figure 3-2 breaks out the treatment approaches that patients receive by their post-reform severity categories and reveals dramatic differences along that dimension. Sicker patients are more likely to be treated with medical management and minor surgery; they are less likely to receive major surgery. Though these statistics reflect equilibrium treatment choices, the incentives for intensive treatment are rising with severity even though intensity falls along this dimension. The figure is therefore strongly suggestive of differences in appropriateness for treatments across the categories.

Within a year, controlling for the full set of observables about a patient would effectively absorb all the variation in the returns to treatment – that return is solely a function of the patient's severity, a report of which is observed in the data. The severity adjustment reform is therefore attractive, because conditional on a patient's severity, it induces time variation in the return to treatment. This intuition suggests that simply adding time-constant severity controls to equations 3.3 and 3.4 could solve the issue of unobserved appropriateness. For example, the equations could be augmented with year fixed effects γ_t and a set of dummies for the patient's severity classification by the pre- and post-reform rules $\mathbf{S}_p$:

$$treat_{pt}^{X} = \gamma_t + \mathbf{S}_p\Theta + \gamma_{own}^{X} \cdot ret_{s(p)t}^{X} + \gamma_{cross}^{X} \cdot ret_{s(p)t}^{-X} + \epsilon_{pt}^{X} \tag{3.5}$$

In this specification, patients that were low severity before and after the reform would receive one shock to the returns to treatment, while patients that were high severity before and medium severity after would receive another, and so on. An identifying assumption is that conditional on severity, there is no residual correlation between patient unobservables and the return to the treatment approach:

$$\mathbb{E}\left[\epsilon_{pt}^{X} \cdot ret_{s(p)t}^{X} | ret_{s(p)t}^{-X}, \ldots, \gamma, \mathbf{S}_p\right] = 0$$

This identifying assumption is unlikely to be satisfied due to the endogenous documentation and coding response of hospitals. It has been well established that when hospitals are paid more to treat sick patients, they go to greater lengths to find evidence of illness and submit higher-severity codes on their reimbursement claims (Dafny, 2005a; Sacarny, 2014; Silverman and Skinner, 2004). Figure 3-3 provides evidence of this phenomenon in the analysis sample. The left pane shows how patients in 2006 (prior to the reform) would have been classified according to the post-reform severity system. The right pane shows how patients in 2009 were actually classified following implementation and the documentation and coding response. The share of low and medium severity patients fell, with a nontrivial share of patients moving into the high severity category.

The coding response implies that patient unobservables within a severity category are changing over time – for example, the strong incentive to move patients into the high-severity category likely drove down the average underlying (unobservable) level of illness of patients in this category. However, since the return to surgery also varies within a severity category, it is likely that the return and the unobservables were correlated. This violates the identifying assumptions of the specification .

3.3.2 Instrumental Variables Strategy

The endogeneity of returns to treatment even after severity group controls suggests a need for an instrument. The instruments that I propose are predictions for a patient's return to surgery based on patients who were treated at her hospital in a base year prior to the reform. The return to treatment in year t is predicted by re-pricing the constant set of base year patients under the year t payment rules. This method isolates the change in the return to treatments that is due solely to the changing IPPS rules, not to the innovations in how patients are reported by hospitals to Medicare.

The intuition behind this strategy is that if average severity differs and is persistent across the hospitals, then high-severity hospitals will experience price changes similar to high severity patients and low-severity hospitals will experience price changes like low severity patients. Calculating the returns using base year patients – whose severity levels were coded before even the announcement of the reform – constructs these returns in a way that is not confounded with the reform's incentives for better documentation and coding of severity.

I propose predicted return instruments, called $\overline{ret}_{ht|t_0}^{major}$ and $\overline{ret}_{ht|t_0}^{minor}$, which are the average returns to surgery that patients from year t_0 in hospital h would have experienced under the IPPS payment rules of year t. The return is calculated by passing patients from year t_0 through a DRG grouping program initialized with grouping rules from year t.[8]

To exploit only within-hospital over-time variation in the predicted return, the specification includes hospital and time fixed effects. This approach yields the OLS and first stage equations for treatment X:

$$treat_{pt}^{X} = \gamma_t + \gamma_{h(p)} + \gamma_{own}^{X} \cdot ret_{s(p)t}^{X} + \gamma_{cross}^{X} \cdot ret_{s(p)t}^{-X} + \epsilon_{pt}^{X} \tag{3.6}$$

$$ret_{s(p)t}^{X} = \delta_t + \delta_{h(p)} + \rho_{own}^{X} \cdot \overline{ret}_{h(p)t|t_0}^{X} + \rho_{cross}^{X} \cdot \overline{ret}_{h(p)t|t_0}^{-X} + \upsilon_{pt}^{X} \tag{3.7}$$

This specification is an instrumental variables differences-in-differences approach. Consider the reduced form regression of the receipt of treatment X on the predicted return to treatments X and $-X$:

$$treat_{pt}^{X} = \tau_t + \tau_{h(p)} + \kappa_{own}^{X} \cdot \overline{ret}_{h(p)t|t_0}^{X} + \kappa_{cross}^{X} \cdot \overline{rev}_{h(p)t|t_0}^{-X} + \xi_{pt} \tag{3.8}$$

In this setup, low-severity *hospitals* (not patients) act as controls for medium- and high-severity hospitals. In the extreme, if each hospital only received patients of one severity category, the regression would match equation 3.5. Figure 3-4 provides a visualization of the first stage and reduced form of the experiment. In the figure, the hospitals are divided into terciles on the basis of the severity of their lung cancer patients in 2003-2005. To construct the terciles, the 2003-2005 patients are classified under the 2008 severity rules. Hospitals are ranked according to their average severity, with patients getting 1, 2, or 3 points depending on whether they are low, medium, or high severity, respectively.

The upper row of figure 3-4 visualizes the first stage of the experiment, showing the realized returns to minor and major surgery according the base period hospital severity. Following the reform, bottom and middle severity tercile hospitals saw the return to minor surgery fall relative to top severity tercile hospitals, and bottom tercile hospitals saw the largest decline in the return. Likewise, while the return to major treatment rose for all terciles, the increase was particularly large for hospitals in the top base period severity tercile. These results show that there is persistence in the types of patients that hospitals treat – the severity of the hospital's base period patients is predictive of the shocks to treatment incentives that the hospital will ultimately experience.

The lower row of figure 3-4 depicts the reduced form of the experiment. While it is hard to discern a pattern in these graphs, they are instructive in how the experiment works. For example, recall that the upper row shows that bottom tercile hospitals received relatively large decreases in the return to minor

[8] From year-to-year, ICD-9 diagnosis and procedure codes are sometimes deleted. Processing patients under DRG rules that post-date their discharge thus entails translating the handful of deleted codes to the appropriate current codes. I construct mappings from deleted codes to current codes by finding the current code that contains the deleted code's diagnosis or procedure. If no unique current code matches, I use the most popular current code that includes the diagnosis or procedure.

surgery and relatively small increases in the return to major surgery. That shock to incentives is the forcing to which the regression will attribute the relative changes in treatment rates depicted in the reduced form graphs. Of course, by aggregating to hospital terciles rather than using all of the information about base period severity interacted with the evolving pricing rules, this visualization throws away useful potentially exogenous variation in the returns to treatment. The results in section 3.4 have greater statistical power to identify relationships between the instruments and the treatment approaches utilized.

3.3.3 Identifying Assumptions

The key identifying assumption of this strategy is that beyond year effects and time-constant hospital effects, the instruments $\overline{ret}^{major}_{h(p)t|t_0}$ and $\overline{ret}^{minor}_{h(p)t|t_0}$ are uncorrelated with the error terms ϵ^{major}_{pt} and ϵ^{minor}_{pt}. Since the assumption concerns variation in these values above and beyond the time and hospital effects, the assumption can only be violated by within-hospital, over time correlations between the instruments and the error term.

The error terms contain unobservable patient appropriateness for the surgeries, so one necessary condition is that the hospital's change in incentives from the new payment rules is uncorrelated with changes in the unobservable appropriateness for the surgeries of the hospital's patients. There are two key scenarios that would violate this condition.

Changes in Admitting Patterns

The first scenario concerns a systematic change in admitting patterns by hospitals that is correlated with the shock to treatment incentives. In response to the incentives, some hospitals may seek to move less intensive surgical procedures to outpatient facilities or perform medical management in physicians' offices or even a hospice setting. Since hospitals often have their own or affiliated outpatient facilities, the decision of where to treat a patient is likely as salient to administrators as how to treat her. Indeed, the rapid rise of outpatient facilities like ambulatory surgical centers has allowed hospitals to treat patients more quickly and potentially reduce costs; research suggests that these centers have grown in response to prospective payment for inpatients (Cullen et al., 2009).

The reform under study changed the prices of inpatient major and minor surgical treatment and inpatient medical treatment. It also changed the price of these courses relative to outpatient treatment because outpatient reimbursements more or less held steady over the period. Recall that the inpatient payments became more finely severity adjusted, raising the payments of all treatments for the very sick and lowering them for the very healthy. Even though the payments for the approaches were often moving in the same direction, the return to major surgery relative to medical management for the very sick rose substantially because the payment for the former increased more than the latter.

Outpatient payments were not severity adjusted at all. Thus the payment for major inpatient surgery relative to outpatient treatment rose even more than the payment for major inpatient surgery relative to inpatient medical management. This is because the risk adjustment of major surgery was partially offset by the risk adjustment of inpatient medical management. Given the long-term growth in outpatient clinics in response to the incentives of prospective payment, there is good reason to think that hospitals were sensitive to the price of inpatient treatment relative to outpatient treatment.

Re-Sorting of Patients

A related violation of the identifying assumptions would occur if the matching of patients to hospitals were altered by the reform in a way that was systematically correlated with the shocks to incentives. This concern would be operative even if the set of patients who were treated in the inpatient setting remained fixed – a pure re-sorting of patients to hospitals that was due to the reform would introduce a correlation between the instrument and unobserved hospital-level appropriateness for the treatments.

For example, the hospitals that received the largest shock to the incentive for major surgery might look to draw patients appropriate for these surgeries to the hospital. Likewise, hospitals that received small or negative shocks to the incentive for major surgery could downsize their operating rooms and send marginal patients to other facilities. The appropriateness for major surgery within a hospital over time would therefore change in a way that was correlated with the shocks to incentives due to the reform.

Testing for Endogeneity

If the predicted returns to treatments are correlated with changes in unobservables of patients who are treated at the hospital, they are invalidated as instruments. The admitting patterns and re-sorting concerns discussed in this section emit two potential ways to test for endogeneity. First, changes in patient volume at the hospital level that are correlated with the instruments would suggest that admitting patterns were modified or re-sorting occurred. A regression of patient counts on the instruments can explore this concern. Second, observable correlates of unobserved patient appropriateness for surgery could be regressed on the instruments. For example, patient age is correlated with health status and is well observed in the data.

In section 3.5 I consider a variety of specification tests along these lines, the results of which are inconclusive. Most of the specification tests fail to reject null hypotheses consistent with instrument validity, though one of the four rejects at the 10% level. In light of the plausible scenarios under which the instruments may not be valid, the specification tests suggest that the results of this study should be interpreted with some caution.

3.4 Results

This section presents the results taking the specifications of equations 3.6 and 3.7 to the data. I make several changes to these equations that improve the power of the instruments. First, I include time effects at the quarterly level, rather than the yearly level. I add controls for well-observed patient characteristics: five age categories (5-year bins from age 65 through 84, and 85+) interacted with sex. Lastly, rather than use two instruments to predict the returns to surgery, I use six: the predicted returns for minor and major surgery based on the hospital's patients in 2003, 2004, and 2005 repriced under the contemporaneous payment rules. Each year yields a prediction for minor treatment and a prediction for major treatment. Allowing the prediction from each year to enter the first stage separately improves the explanatory power of the regressions, though the results are robust to using only the two instruments that are based on 2005 patients.

3.4.1 Baseline Specification: First Stage and IV

The first stage results are given in table 3.4. In column 1, I show the first stage for $ret^{minor}_{s(p)t}$ and in column 2 I show the first stage for $ret^{major}_{s(p)t}$. The explanatory power of the instruments is substantially greater for the return to major surgery than for the return to minor surgery. This is perhaps unsurprising: the minor surgery DRG was risk adjusted (albeit coarsely) in the old IPPS, damping the variation in prices that was induced by the reform.

Five of the six "own" price instruments are positive and significant. Since the first stage includes quarterly fixed effects and hospital fixed effects, this implies that, fixing pre-period patients and exploiting only the change in prices due to the reform, the predicted change in the return to treatment is positively associated with the realized change in the return to treatment.[9]

The instruments constructed from 2005 patients are particularly strong predictors of realized prices in the analysis sample. Given that the types of patients visiting the hospital may be evolving over time, predictors based on more recent samples of patients may be stronger than those constructed on earlier samples. Still, the instruments based on 2003 and 2004 patients are frequently significant and have coefficients of greater magnitude than their 2005 counterparts, suggesting that the decay in power from using earlier samples of patients to predict returns is not large.

Table 3.5 presents the results of the instrumental variables regression using the six predicted returns variables as instruments. On the diagonals of the table are the own-price elasticities and on the off-diagonals are the cross-price elasticities. The coefficients should be interpreted as the percentage point increase in the treatment rate due to a $1000 increase in the reimbursement for the treatment relative to medical management.

[9]Some of the "cross" price instruments are also significant. One story consistent with this fact is that the instruments are becoming weaker predictors of the return over time, contrary to the specification with time constant coefficients. As a result, the cross price instrument, which is highly correlated with the own price instrument, may effectively allow the coefficient to attenuate over time.

Thus raising the reimbursement for minor treatment by $1000 is estimated to raise the minor treatment rate by 1.6pp (own-price) and reduce the major treatment rate by 1pp (cross price). Since medical management is the omitted category, the remaining 0.6pp of the shift to minor treatment would come from this treatment approach. Likewise, raising the reimbursement for major treatment by $1000 is estimated to raise the major treatment rate by 1pp (own price) and reduce the minor treatment rate by 0.8pp (cross price) – the remaining 0.2pp of the increase in the major treatment rate comes from patients who would have been medically managed.

These coefficients are right-signed according to economic theory – the own-price elasticities are positive and the cross-price elasticities are negative (consistent with the treatment "goods" being substitutes, not complements). However, one concern with interpreting the coefficients is that the effects with respect to the return to minor treatment are measured imprecisely and are not statistically significant, a result of the relatively weak first stage for this variable. In addition, the results should be read with caution due to the concerns about identification discussed in section 3.3.3.

3.4.2 Effects by Hospital Control

Previous literature on hospital payment under the IPPS has found that for-profit facilities are more responsive to incentives for documentation and coding. Dafny (2005a) shows that when the IPPS was reformed to pay more for patient severity for some patients, for-profit hospitals were more likely to respond to the incentive to list patients as high severity than their non-profit and public peers. Silverman and Skinner (2004) show that for-profit hospitals were more likely to manipulate diagnosis codes to push patients into higher-paying DRGs. They also find that non-profits in markets with greater for-profit penetration are more likely to manipulate the codes.

This research showed a differential "nominal" coding response by hospital control; Dafny looked for, but found no evidence of a differential effect of incentives on "real" treatment decisions. I pursue evidence for such a response by interacting the endogenous returns and return instruments with indicators for whether the hospital has for-profit or government ownership. If the interaction effects were significant, it would suggest that elasticities to incentives differ by hospital control.

Table 3.6 shows the results of the IV regression. Columns 1 and 4 replicate the baseline results, while columns 2 and 5 add the hospital control interactions. The tripling of the instrument set causes the Kleibergen-Paap Weak Identification F-statistic to fall to 6.6, well below the rough cutoff of 10 suggested by Stock et al. (2002), so in columns 3 and 6 I estimate the equations using LIML, which is approximately median-unbiased in weakly identified specifications (Angrist and Pischke, 2009).

None of the interaction effects is significant in these specifications, though with such weak identification the power to detect these parameters is limited. The question of whether the response to treatment incentives differs by control clearly remains an open one.

3.4.3 Policy Simulation

One policy question is how treatment approaches would be affected if incentives were powered down. For example, in fully capitated arrangements, provider groups receive a fixed payment per beneficiary in exchange for assuming responsibility for that beneficiary's care – hospital payment is no longer a function of the treatment approach at all. Following provisions of the Affordable Care Act, Medicare is currently implementing Accountable Care Organizations (ACOs), which create novel contracts with provider groups to remove some of the incentives of traditional fee-for-service payment.

The two new ACO programs, Medicare Shared Savings and Pioneer, make provider groups more responsible for the overall cost of caring for patients. Under Medicare Shared Savings, the provider group receives additional payments at the end of the year if its assigned Medicare beneficiaries incurred lower than expected costs; these upside payments are larger if the group opts to accept downside risk, in which the group's payments are reduced if beneficiaries incurred greater than expected costs. The Pioneer program goes further and introduces the potential for partial capitation: the group's fee-for-service payments are reduced in exchange for a monthly payment for taking responsibility for the beneficiary's care (Boyarsky and Parke, 2012). These policies have the effect of reducing the return to treatments, since the additional payments the hospital would receive for choosing, for example, a high intensity treatment approach would be partly clawed back from the provider group at the end of the year (in the case of Medicare Shared Savings) or attenuated (in the case of Pioneer).

In table 3.7, I simulate the effects of a policy that goes further than these programs and eliminates the return to treatment approaches entirely. In 2009, the average return to minor surgery was \$7,830 and the return to major surgery was \$13,232. The first row shows the overall effect of reducing the incentives for the treatment approaches by these averages according to the elasticities estimated in the baseline specification. Therefore the point estimates for the effects on the minor and major surgery rates reported in the table are, respectively:

$$d^{minor} = -7.83 * \hat{\gamma}^{minor}_{own} - 13.23 * \hat{\gamma}^{minor}_{cross} \quad (3.9)$$

$$d^{major} = -7.83 * \hat{\gamma}^{major}_{cross} - 13.23 * \hat{\gamma}^{major}_{own} \quad (3.10)$$

The $\hat{\gamma}$ are the estimates of the elasticities from the baseline specification (see table 3.5). The standard errors are for the hypotheses that $d^{minor} = 0$ and $d^{major} = 0$.[10]

[10]The instrumental variables regression produces a local average treatment effect (LATE) estimate of the γ, and one issue that these results raise is whether the LATE would be valid for a large change in incentives. The policy prediction relies on the assumption that the γ are constant or can be effectively treated as such for the range of policies being studied.

The point estimates suggest that the share of lung cancer patients receiving minor surgery would fall by 2pp and the share receiving major surgery would fall by 5pp – though neither estimate is statistically significant. The subsequent rows show how treatment rates would change for each severity group assuming that the 2009 returns to treatment displayed in table 3.2 were reduced to zero. The higher severity groups are estimated to have larger reductions in their surgery rates, a result of their greater returns to the surgical treatment approaches. However, none of these estimates is statistically significant.

3.4.4 Effects on Outcomes

The equation of interest for patient outcomes considers how mortality and readmission respond to the treatment approach that the patient receives:

$$outcome_{pt} = \pi_{major} \cdot treat_{pt}^{major} + \pi_{minor} \cdot treat_{pt}^{minor} + \eta_{pt} \tag{3.11}$$

Where the π_X coefficient is the percentage point increase in the mortality rate or readmission rate due to the receipt of treatment X. Of course, a simple regression of outcomes on treatment approach will not recover causal effects of the treatments since treatments are chosen endogenously. The 2008 reform offers the possibility of resolving this endogeneity by using the shocks to the return to treatment induced by the reform as instruments for the realized treatment approach. In this setup, the reduced form of the baseline specification becomes the first stage:

$$treat_{pt}^{X} = \tau_t + \tau_{h(p)} + \kappa_{own}^{X} \cdot \overline{ret}_{h(p)t|t_0}^{X} + \kappa_{cross}^{X} \cdot \overline{rev}_{h(p)t|t_0}^{-X} + \xi_{pt} \tag{3.12}$$

The OLS equation is augmented with time and hospital fixed effects:

$$outcome_{pt} = \lambda_t + \lambda_{h(p)} + \pi_{major} \cdot treat_{pt}^{major} + \pi_{minor} \cdot treat_{pt}^{minor} + \eta_{pt} \tag{3.13}$$

The identifying assumption is that the within-hospital, over-time variation in the instruments is uncorrelated with patient unobservables. For example, a necessary condition is that unobservables – the underlying propensity of patients to die or be readmitted – were not, within a hospital, evolving in a manner that was correlated with the shock to the returns to treatments. In this sense, the concerns of section 3.3.3 remain operative. If hospitals systematically change their admitting patterns of patients in response to the reform, or if patients re-sort to facilities due to the shocks, then the identifying assumptions are invalidated. Section 3.5 presents mixed results on whether these assumptions hold.

Table 3.8 presents the 2SLS results on outcomes. Column 1 puts in-hospital mortality on the left-hand side, column 2 considers mortality within 30 days of discharge, and column 3 considers mortality within 90 days of discharge. Columns 4-6 look at whether the patient was transferred (admitted to the same or another inpatient facility on the day of discharge), readmitted within 30 days of discharge, or readmitted within 90 days of discharge, respectively. The coefficients can be interpreted as the percentage point effect of receiving the treatment on the probability of the outcome occurring.

Though some of the results have economically substantial coefficients, the regressions suffer from severe weak identification – the Kleibergen-Paap F-statistic is 0.5, suggesting that the IV coefficients are likely to be biased in the direction of the OLS coefficients. With such a low F-statistic, there is little power to identify any effect. However, following standard suggestions in this situation, in table 3.9 I estimate the regressions with LIML. The coefficients here are all statistically insignificant, and the magnitudes are often greater than 1, a result that is suggestive of low statistical power in a linear probability model with dummy variables on the right-hand side. In light of the weak identification and endogeneity concerns, it is hard to draw conclusions from these results.

3.5 Specification Tests

The key issue regarding the identification strategy of this study is whether patient unobservables are evolving in a manner that is correlated with the instruments. In section 3.3.3, I proposed two mechanisms of particular concern. First, hospitals could be responding on the admission decision margin, with hospitals that received positive shocks to incentives loosening their admission standards and hospitals that received negative shocks moving marginal patients to the outpatient setting. Second, patients could re-sort due to the reform: even if the pool of patients treated in the inpatient setting did not change, sicker patients could move to hospitals that got positive shocks and healthier patients could move to hospitals that got negative shocks.

These concerns emit two falsification tests. The first, a covariate balance test at the patient level, asks if the instruments are correlated with observable factors that are likely to be related to unobservable severity and appropriateness. If a correlation exists, it suggests that further endogeneity may remain in the error term even when the observables are used as controls. To run this test, I repeat the reduced form specification of equation 3.8 replacing the left-hand side with patient age and log-age – this differences-in-differences specification relates within-hospital over-time variation in patient age to the instruments for the return to treatment. The age variables are likely correlates of unobservable severity and appropriateness for the treatments.

The second test is at the hospital level and asks whether hospitals that received shocks to the treatment returns systematically changed their admitting patterns. A correlation in this test would suggest that the instrument was correlated with a change in admission standards. To run this test I take the specification of equation 3.8 and place patient counts on the left-hand side. This regression is at the hospital-quarter level and includes hospital effects, time effects, and the instruments on the right-hand side.

These specification tests are presented in table 3.10. Columns 1 and 2 regress age and log-age, respectively, on the instruments. Four of the 12 coefficients in these regressions are significant at the 10% level, but in neither regression does the test that all the coefficients are zero reject. Column 3 regresses the log of the hospital's lung cancer patient count on the instruments. In the linear regression specification, the patient count is incremented so that it is never zero. Column 4 repeats the specification as a fixed-effects Poisson counting model. This model can accommodate zeroes on the left-hand side, so it is not necessary to increment the count. The linear regression and Poisson models return somewhat conflicting results – the joint test that all coefficients are zero easily fails to reject in column 3, but rejects at the 10% level in column 4.

The interpretation of these results depends crucially on one's prior about the validity of the instruments. Under the prior that the instruments are valid, the findings are broadly consistent with this belief. In columns 1 through 3, the joint tests that the coefficients on the instruments are zero fail to reject. Only in column 4 does the test reject, but this is only at the 10% level and is inconsistent with the test results of the similar specification of column 3.

On the other hand, if one's prior is that the patient selection and re-sorting channels were operative, the results of these regressions may not be convincing enough to invalidate that belief. First, the confidence intervals of the coefficients in columns 1 and 2 include nontrivial potential effects. In column 1, the 95% confidence interval on the 2004 minor return to surgery instrument ranges from -.01 to 0.21. The standard deviations of the return instruments are all around 1, so the effect at the upper end of the confidence set would be that a hospital that receives a one standard deviation increase in the 2004 minor return to surgery instrument would begin to admit patients several months older.

In addition, the fact that the Poisson model rejects the null that all coefficients are zero is consistent with a response on the admissions margin – and even though the linear model of column 3 fails to reject the null, the point estimates of its coefficients are economically nontrivial. For example, the point estimate of 0.015 on the return to minor surgery predicted with 2004 patients suggests that a 1 standard deviation increase in that return is associated with a 1.5pp increase in the number of patients.

The results of this section are clearly mixed, and they help to color the credibility of the estimates of section 3.4. Under the presumption that the instruments are valid, the specification tests broadly fail to reject and when they do, the rejection is sensitive to the specification. On the other hand, if one presumes that the concerns about endogeneity are valid, the tests are not highly powered enough to refute these concerns.

3.6 Conclusion

In this study, I have considered how hospitals alter the approach they use to treat lung cancer patients in response to incentives. My identification strategy instruments for the relative returns of the treatment approaches using a 2008 reform that differentially shocked these returns across hospitals. I estimate a sensible substitution matrix of own- and cross-price elasticities for the treatments, with positive own-price elasticities and negative cross-price elasticities. Though the point estimates are economically substantial, they are often statistically insignificant – a result of an identification strategy that exploits only limited variation in the endogenous variables. The instruments also produce a strategy for identifying the effects of receiving a treatment approach on patient outcomes, though the power issues are multiplied in this specification; I unsurprisingly fail to detect any real effects on outcomes. Furthermore, concerns about the identification strategy indicate that my results should be interpreted with some caution.

The question of how hospitals alter their treatments in response to changes in incentives remains of first-order importance in health policy. Current policy discussion revolves around reducing or eliminating the incentives for high intensity treatment common in fee-for-service systems – as this paper has shown, these incentives are common even in prospective systems like the IPPS. As insurers expand their involvement in Accountable Care Organizations and other programs with lower-powered payments for treatment, it will be crucial to track the resulting treatment intensity response by providers and its further effects on patient outcomes. While this study demonstrates significant elasticities of treatment to prices, they are estimated with some imprecision, rely on strong identification assumptions, and are limited to lung cancer patients; estimates with greater precision and for other treatments could be useful for optimal policy.

Studying treatment decisions by hospitals raises further questions about how payment incentives for facilities translate into treatment decisions taken by physicians in consultation with patients. Research on documentation and coding responses by hospitals has pointed to organizational mechanisms by which physicians were pressured to exaggerate the illnesses of their patients (Dafny, 2005a; Silverman and Skinner, 2004) or to improve the level of detail in their notetaking (Sacarny, 2014). These studies suggest, but do not prove, that organizational factors play a role in moving the treatment decisions of physicians into line with the hospital's incentives. The mechanisms that allow facilities to transmit their incentives to decisionmakers in the treatment process remain under-explored, and a potentially fruitful avenue for further research.

Figures

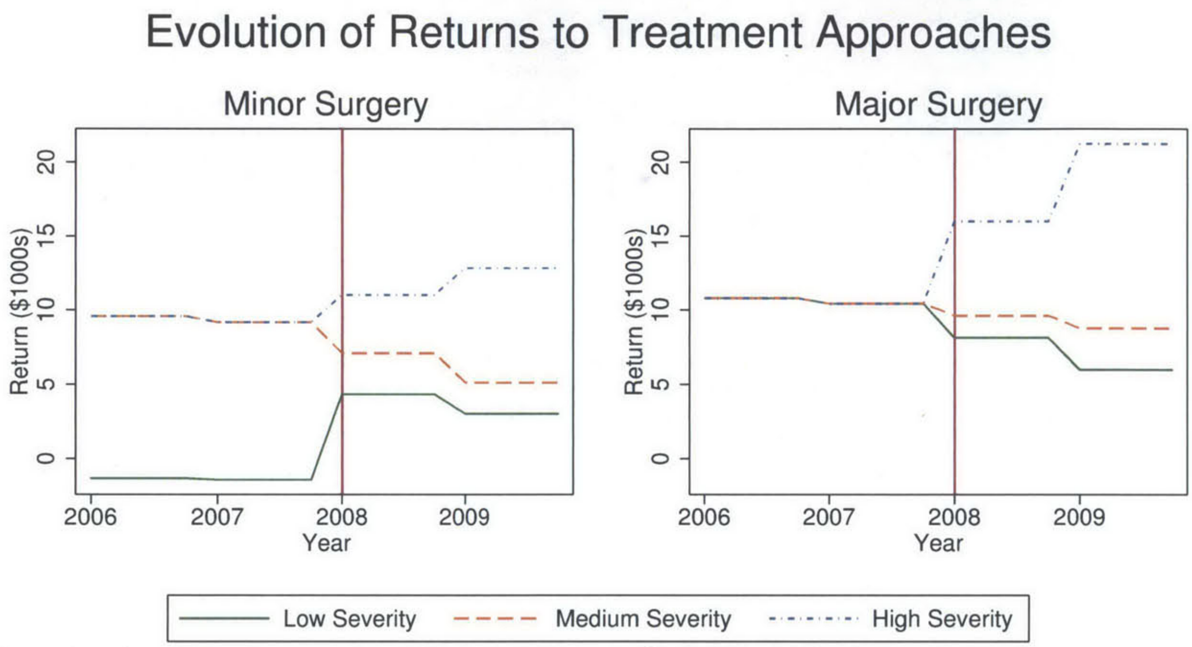
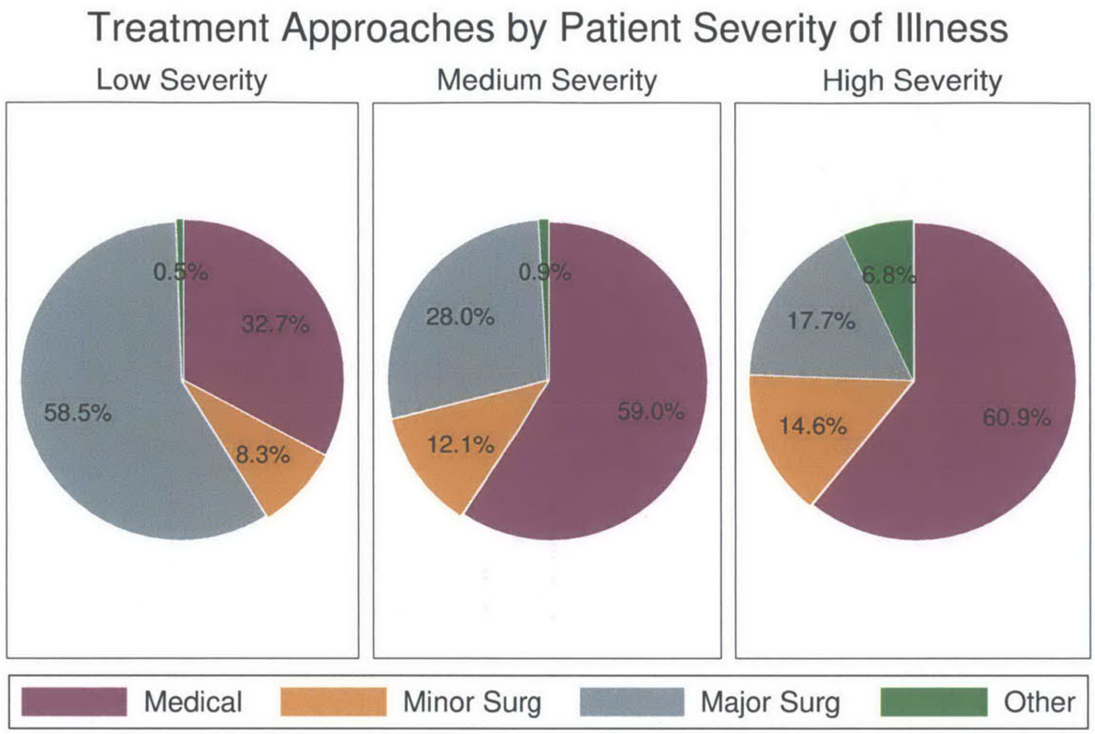


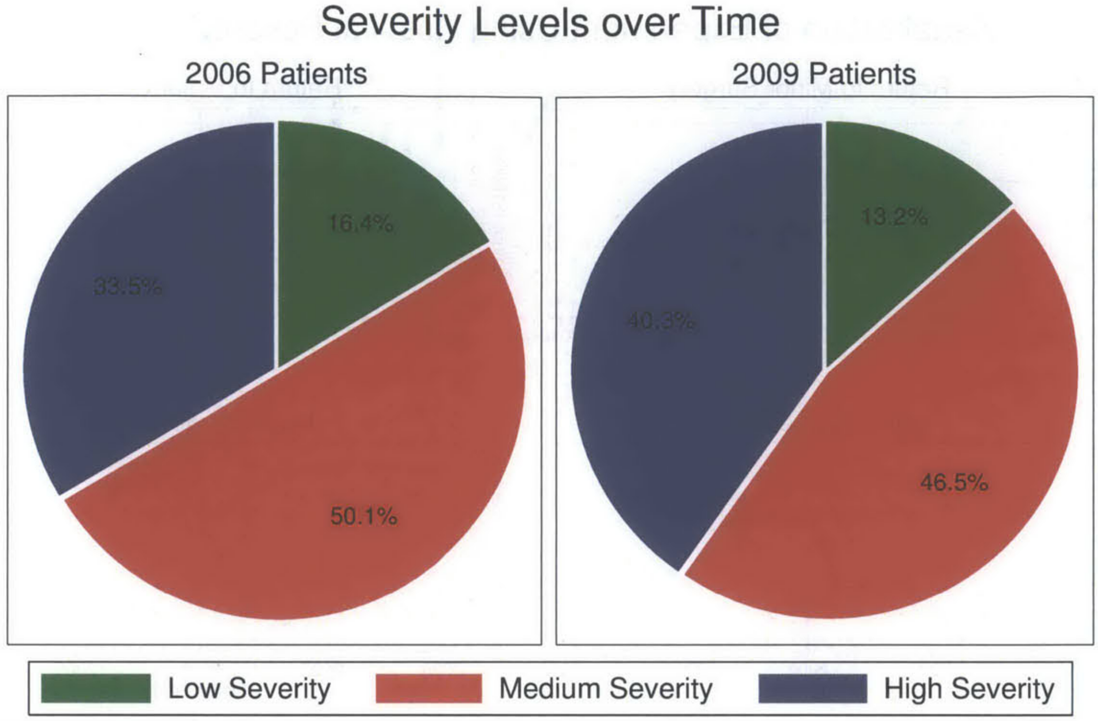
Figure plots the additional revenue that a hospital would receive for treating a patient with major or minor surgery instead of medical management. The returns are broken out by patient severity of illness. The red line denotes the reform date. See text for more details.

Figure 3-1



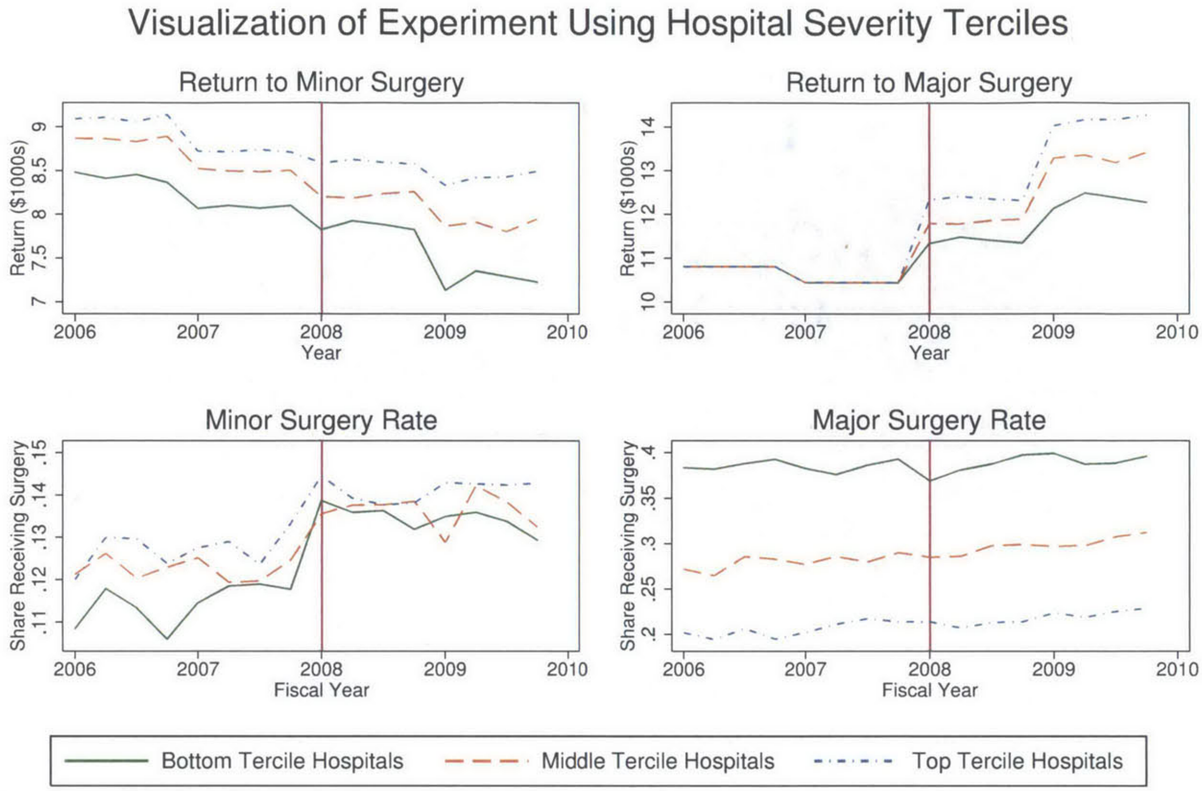
Figures reflect 2006–2009 lung cancer patients. Patients receiving "Other" treatment excluded from analysis. See text for more details.

Figure 3-2



Figures reflect 2006–2009 lung cancer patients. Patients are classified according to the post–reform severity rules. See text for more details.

Figure 3-3



Severity terciles are constructed using average severity of hospital's patients in 2003–2005. The upper row plots the realized returns to surgery relative to medical management by severity tercile. The lower row plots treatment rates by severity tercile. The red line denotes the reform date. See text for more details.

Figure 3-4

Tables

Table 3.1 - Reimbursements for Treatment Approaches

Pre-Reform (2007)		Post-Reform (2009)	
Severity	Payment	Severity	Payment
Medical Management ("Respiratory Neoplasms")			
Low*	9.08	Low	5.62
High*	9.08	Medium	7.92
		High	10.90
Minor Surgery ("Other Respiratory Procedures")			
Low	7.65	Low	8.64
High	18.26	Medium	13.03
		High	23.74
Major Surgery ("Major Chest Procedures")			
Low*	19.52	Low	11.60
High*	19.52	Medium	16.69
		High	32.15

* DRG was not adjusted for patient severity. Payments are in $1000s for an average US hospital according to 2009 conversion rates from DRG weights to $.

Table 3.2 - Returns to Treatment Approaches by Severity

Pre-Reform (2007)		Post-Reform (2009)	
Severity	Return	Severity	Return
Return to Minor Surgery - Medical Management			
Low	-1.44	Low	3.02
High	9.18	Medium	5.11
		High	12.84
Return to Major Surgery - Medical Management			
Low	10.44	Low	5.98
High	10.44	Medium	8.77
		High	21.24

Returns are in $1000s for an average US hospital according to 2009 conversion rates from DRG weights to $.

Table 3.3 - Summary Statistics

	(1) Mean	(2) SD
Patient Characteristics		
Age	75.65	6.972
Female	0.496	0.500
Treatment Approach Taken		
Medical Management	0.574	0.494
Minor Surgery	0.129	0.335
Major Surgery	0.297	0.457
Return to Treatment vs. Medical Management ($1000s)		
Minor Surgery	8.337	3.036
Major Surgery	11.52	3.635
Patient Severity (Pre-Reform Rules)		
Low ("Non-CC")	0.0689	0.253
High ("CC")	0.931	0.253
Patient Severity (Post-Reform Rules)		
Low ("Non-CC")	0.150	0.357
Medium ("CC")	0.496	0.500
High ("MCC")	0.354	0.478
Patient Mortality		
In Hospital	0.0910	0.288
30 Day Mortality	0.309	0.462
90 Day Mortality	0.463	0.499
Readmission and Transfer		
Transferred	0.0132	0.114
30 Day Readmission	0.200	0.400
90 Day Readmission	0.319	0.466

The sample includes 337,198 patients across 2,946 hospitals. CC refers to Complication or Comorbidity. MCC refers to Major Complication or Comorbidity.

Table 3.4 - First Stage Regressions

Patient Received:	(1) Minor Surgery	(2) Major Surgery
Return to Minor Surgery	0.00451	-0.180***
(Predicted, 2003 Patients)	(0.0297)	(0.0430)
Return to Minor Surgery	0.130***	-0.0972**
(Predicted, 2004 Patients)	(0.0289)	(0.0450)
Return to Minor Surgery	0.0840***	-0.206***
(Predicted, 2005 Patients)	(0.0302)	(0.0465)
Return to Major Surgery	0.0385*	0.231***
(Predicted, 2003 Patients)	(0.0217)	(0.0313)
Return to Major Surgery	0.0223	0.269***
(Predicted, 2004 Patients)	(0.0207)	(0.0339)
Return to Major Surgery	0.0476**	0.334***
(Predicted, 2005 Patients)	(0.0198)	(0.0309)
Observations	337,198	337,198
R^2 (adjusted)	0.017	0.098
Joint F-test P-value	0	0

Standard errors clustered at the hospital level. Regressions include quarterly fixed effects, hospital fixed effects, and controls for 5 age categories interacted with sex. F-test null is that all 6 predicted return coefficients are 0. Returns are in $1000s.
*** significant at 1% level; ** significant at 5% level; * significant at 10% level

Table 3.5 - Instrumental Variables Regressions

	(1)	(2)
Patient Received:	Minor Surgery	Major Surgery
Return to Minor Surgery	0.0155	-0.0101
	(0.00952)	(0.0135)
Return to Major Surgery	-0.00758**	0.00977**
	(0.00305)	(0.00453)
Observations	337,169	337,169
Joint F-test P-value	0.0399	0.0419
KP Weak ID F-statistic	19.30	19.30

Standard errors clustered at the hospital level. Regressions include quarterly fixed effects, hospital fixed effects, and controls for 5 age categories interacted with sex. F-test null is that both return coefficients are 0. Returns are in $1000s.

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

Table 3.6 - Instrumental Variables Results by Hospital Control

Patient Received:	(1) Minor Surgery	(2) Minor Surgery	(3) Minor Surgery	(4) Major Surgery	(5) Major Surgery	(6) Major Surgery
Return to Minor Surgery	0.016	0.013	0.013	-0.010	-0.015	-0.013
	(0.0095)	(0.0098)	(0.010)	(0.014)	(0.014)	(0.015)
* Government-Run		0.0044	0.0045		0.015	0.015
		(0.0074)	(0.0075)		(0.0094)	(0.0095)
* For-Profit		0.0011	0.0011		0.012	0.012
		(0.0078)	(0.0079)		(0.012)	(0.012)
Return to Major Surgery	-0.0076**	-0.0077**	-0.0078**	0.0098**	0.0099**	0.0097**
	(0.0031)	(0.0030)	(0.0031)	(0.0045)	(0.0045)	(0.0045)
* Government-Run		0.0013	0.0013		0.0021	0.0020
		(0.0020)	(0.0020)		(0.0030)	(0.0031)
* For-Profit		0.0027	0.0027		-0.00011	-0.00015
		(0.0023)	(0.0023)		(0.0033)	(0.0033)
Observations	337,169	337,169	337,169	337,169	337,169	337,169
KP Weak ID F-statistic	19.3	6.58	6.58	19.3	6.58	6.58
Model	2SLS	2SLS	LIML	2SLS	2SLS	LIML

Standard errors clustered at the hospital level. Regressions include quarterly fixed effects, hospital fixed effects, and controls for 5 age categories interacted with sex. The omitted interaction category is Non-Profit. Returns are in $1000s.

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

Table 3.7 - Policy Simulation

Patient Received:	Minor Surgery	Major Surgery
Overall	-0.021	-0.050
	(0.049)	(0.073)
Low Severity Patients Only	-0.002	-0.028
	(0.018)	(0.027)
Medium Severity Patients Only	-0.013	-0.034
	(0.032)	(0.047)
High Severity Patients Only	-0.039	-0.078
	(0.082)	(0.120)

Results predict the effect of setting the returns to minor and major surgery to zero on the probability patient receives surgery, relative to 2009 payment rules.
*** significant at 1% level; ** significant at 5% level; * significant at 10% level

Table 3.8 - Patient Outcomes - 2SLS Model

Patient Received:	(1) Died in Hospital	(2) 30 Day Mortality	(3) 90 Day Mortality	(4) Transferred	(5) 30 Day Readmission	(6) 90 Day Readmission
Received Minor Surgery	0.303	-0.311	0.142	0.307	-0.673	-0.472
	(0.594)	(0.803)	(0.622)	(0.299)	(0.678)	(0.645)
Received Major Surgery	-0.573	-1.130*	-0.654	0.218	-0.561	-0.258
	(0.443)	(0.606)	(0.479)	(0.225)	(0.494)	(0.492)
Observations	337,169	337,169	337,169	337,169	337,169	337,169
Joint F-test P-value	0.0260	0.0468	0.0476	0.577	0.515	0.764
KP Weak ID F-statistic	0.495	0.495	0.495	0.495	0.495	0.495

Standard errors clustered at the hospital level. Regressions include quarterly fixed effects, hospital fixed effects, and controls for 5 age categories interacted with sex.
*** significant at 1% level; ** significant at 5% level; * significant at 10% level

Table 3.9 - Patient Outcomes - LIML Model

Patient Received:	(1) Died in Hospital	(2) 30 Day Mortality	(3) 90 Day Mortality	(4) Transferred	(5) 30 Day Readmission	(6) 90 Day Readmission
Received Minor Surgery	-0.194	-0.940	4.516	0.411	-2.591	-3.285
	(4.694)	(2.300)	(181.0)	(0.453)	(6.189)	(15.96)
Received Major Surgery	-1.104	-1.682	2.494	0.295	-1.974	-2.290
	(3.492)	(1.723)	(134.3)	(0.342)	(4.609)	(11.89)
Observations	337,169	337,169	337,169	337,169	337,169	337,169
Joint F-test P-value	0.177	0.254	0.892	0.656	0.912	0.973
KP Weak ID F-statistic	0.495	0.495	0.495	0.495	0.495	0.495

Standard errors clustered at the hospital level. Regressions include quarterly fixed effects, hospital fixed effects, and controls for 5 age categories interacted with sex.
*** significant at 1% level; ** significant at 5% level; * significant at 10% level

Table 3.10 - Specification Tests

	(1) Age	(2) ln(Age)	(3) ln(1+Pats)	(4) Patients
Return to Minor Surgery	-0.0179	-0.000238	0.00573	0.0156
(Predicted, 2003 Patients)	(0.0531)	(0.000690)	(0.00904)	(0.00966)
Return to Minor Surgery	0.102*	0.00135*	0.000414	0.000995
(Predicted, 2004 Patients)	(0.0550)	(0.000715)	(0.0108)	(0.0104)
Return to Minor Surgery	-0.0424	-0.000548	0.0153	0.0133
(Predicted, 2005 Patients)	(0.0530)	(0.000687)	(0.00972)	(0.00963)
Return to Major Surgery	0.0172	0.000233	0.000536	-0.00789
(Predicted, 2003 Patients)	(0.0355)	(0.000462)	(0.00643)	(0.00700)
Return to Major Surgery	-0.0649*	-0.000871*	0.000833	-0.00483
(Predicted, 2004 Patients)	(0.0371)	(0.000482)	(0.00705)	(0.00722)
Return to Major Surgery	0.0329	0.000427	-0.0102	-0.0116*
(Predicted, 2005 Patients)	(0.0359)	(0.000466)	(0.00667)	(0.00661)
Observations	337,198	337,198	11,220	11,142
R^2 (adjusted)	0.001	0.001	0.059	
Joint F-test P-value	0.677	0.650	0.682	0.0709
Model	OLS	OLS	OLS	Poisson

Standard errors clustered at the hospital level. Regressions include quarterly fixed effects and hospital fixed effects. F-test null is that all predicted return coefficients are 0. The age regressions are at the patient level and the patients regressions are at the hospital-quarter level. Coefficients in column 4 are raw Poisson coefficients, not incidence-rate ratios. Returns are in $1000s.

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

References

3M Health Information Systems. 2007. *Medicare Severity Diagnosis Related Groups (MS-DRGs) Definitions Manual.* Version 25.0.

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.

Andersen, Henning R., Torsten T. Nielsen, Klaus Rasmussen, Leif Thuesen, Henning Kelbaek, Per Thayssen, Ulrik Abildgaard, Flemming Pedersen, Jan K. Madsen, Peer Grande, Anton B. Villadsen, Lars R. Krusell, Torben Haghfelt, Preben Lomholt, Steen E. Husted, Else Vigholt, Henrik K. Kjaergard, Leif Spange Mortensen, and DANAMI-2 Investigators. 2003. "A Comparison of Coronary Angioplasty with Fibrinolytic Therapy in Acute Myocardial Infarction." *The New England Journal of Medicine*, 349(8): 733–42.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton: Princeton University Press.

Arrow, Kenneth J. 1963. "Uncertainty and the Welfare Economics of Medical Care." *The American Economic Review*, 53(5): pp. 941–973.

Asplund, Marcus, and Volker Nocke. 2006. "Firm Turnover in Imperfectly Competitive Markets." *The Review of Economic Studies*, 73(2): 295–327.

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

Balasubramanian, Natarajan, and Jagadeesh Sivadasan. 2009. "Capital Resalability, Productivity Dispersion, and Market Structure." *The Review of Economics and Statistics*, 91(3): 547–557.

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

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

Bartelsman, Eric J., and Mark Doms. 2000. "Understanding Productivity: Lessons from Longitudinal Microdata." *Journal of Economic Literature*, 38(3): 569–594.

Bloom, Nicholas, Christos Genakos, Raffaella Sadun, and John Van Reenen. 2012a. "Management Practices Across Firms and Countries." Working Paper 17850, National Bureau of Economic Research, URL: http://www.nber.org/papers/w17850.

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

Bloom, Nicholas, Carol Propper, Stephan Seiler, and John Van Reenen. 2010. "The Impact of Competition on Management Quality: Evidence from Public Hospitals." Working Paper 16032, National Bureau of Economic Research, URL: http://www.nber.org/papers/w16032.

Bloom, Nicholas, and John Van Reenen. 2007. "Measuring and Explaining Management Practices across Firms and Countries." *The Quarterly Journal of Economics*, 122(4): 1351–1408.

Boyarsky, Victoria, and Rob Parke. 2012. "The Medicare Shared Savings Program and the Pioneer Accountable Care Organizations." Milliman Health Care Reform Briefing Paper, Milliman, URL: http://us.milliman.com/insight/healthreform/The-Medicare-Shared-Savings-Program-and-the-Pioneer-Accountable-Care-Organizations/.

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.

Centers for Medicare and Medicaid Services. 2013. "Acute Care Hospital Inpatient Prospective Payment System." Payment System Fact Sheet Series, Medicare Learning Network, URL: http://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/downloads/AcutePaymtSysfctsht.pdf.

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, URL: http://www.nber.org/papers/w19200.

Chandra, Amitabh, Jonathan S. Skinner, and Douglas O. Staiger. 2010. "Saving Money (and Lives)." In *The Healthcare Imperative: Lowering Costs and Improving Outcomes.* eds. Pierre L. Young, Robert S. Saunders, and LeighAnne Olsen: Institute of Medicine.

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.

Clemens, Jeffrey P., and Joshua D. Gottlieb. 2013. "Bargaining in the Shadow of a Giant: Medicare's Influence on Private Payment Systems." Working Paper 19503, National Bureau of Economic Research, URL: http://www.nber.org/papers/w19503.

Clemens, Jeffrey P., and Joshua D. Gottlieb. 2014. "Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?" *The American Economic Review*, 104(4): 1320–1349.

Collard-Wexler, Allan. 2011. "Productivity Dispersion and Plant Selection in the Ready-Mix Concrete Industry." Working Paper 11-25, Center for Economic Studies, URL: http://ideas.repec.org/p/cen/wpaper/11-25.html.

Collard-Wexler, Allan, and Jan De Loecker. 2013b. "Reallocation and Technology: Evidence from the U.S. Steel Industry." Working Paper 18739, National Bureau of Economic Research, URL: http://www.nber.org/papers/w18739.

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

Cullen, Karen A., Margaret J. Hall, and Aleksandr Golosinskiy. 2009. "Ambulatory Surgery in the United States, 2006." *National Health Statistics Reports*, (11).

Cutler, David M. 1995. "The Incidence of Adverse Medical Outcomes Under Prospective Payment." *Econometrica*, 63(1): pp. 29–50.

Cutler, David M. 2010. "Where Are the Health Care Entrepreneurs?" *Issues in Science & Technology*, 27(1): 49–56.

Cutler, David M., and Mark McClellan. 2001a. "Is Technological Change in Medicine Worth It?" *Health Affairs*, 20(5): 11–29.

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

Davis, Steven J., John C. Haltiwanger, and Scott Schuh. 1998. *Job Creation and Destruction*. 1: The MIT Press, , URL: http://ideas.repec.org/b/mtp/titles/0262540932.html.

Disney, Richard, Jonathan Haskel, and Ylva Heden. 2003. "Restructuring and Productivity Growth in UK Manufacturing." *Economic Journal*, 113(489): 666–694.

Dobbelaere, Sabien, and Jacques Mairesse. 2013. "Panel Data Estimates of the Production Function and Product and Labor Market Imperfections." *Journal of Applied Econometrics*, 28(1): 1–46.

Doms, Mark, and Eric J. Bartelsman. 2000. "Understanding Productivity: Lessons from Longitudinal Microdata." *Journal of Economic Literature*, 38(3): 569–594.

Dranove, David, Daniel Kessler, Mark McClellan, and Mark Satterthwaite. 2003. "Is More Information Better? The Effects of "Report Cards" on Health Care Providers." *Journal of Political Economy*, 111(3): 555–588.

Dranove, David, and Andrew Sfekas. 2008. "Start Spreading the News: A Structural Estimate of the Effects of New York Hospital Report Cards." *Journal of Health Economics*, 27(5): 1201–7.

Ellis, Randall P., and Thomas G. McGuire. 1996. "Hospital Response to Prospective Payment: Moral Hazard, Selection, and Practice-Style Effects." *Journal of Health Economics*, 15(3): 257 – 277.

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

Ericson, Richard, and Ariel Pakes. 1995. "Markov-Perfect Industry Dynamics: A Framework for Empirical Work." *The Review of Economic Studies*, 62(1): 53–82.

Escribano, Alvaro, and J. Luis Guasch. 2005. "Assessing The Impact Of The Investment Climate On Productivity Using Firm-Level Data : Methodology And The Cases Of Guatemala, Honduras, And Nicaragua." World Bank Policy Research Working Paper 3621, World Bank, URL: http://dx.doi.org/10.1596/1813-9450-3621.

Fisher, Elliott S., David E. Wennberg, Thérèse A. Stukel, Daniel J. Gottlieb, F. L. Lucas, and Etoile L. Pinder. 2003a. "The Implications of Regional Variations in Medicare Spending. Part 1: The Content, Quality, and Accessibility of Care." *Annals of Internal Medicine*, 138(4): 273–87.

Fisher, Elliott S., David E. Wennberg, Thérèse A. Stukel, Daniel J. Gottlieb, F. L. Lucas, and Etoile L. Pinder. 2003b. "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.

Foster, Lucia, John C. Haltiwanger, and Chad Syverson. 2012. "The Slow Growth of New Plants: Learning about Demand?" Working Paper 17853, National Bureau of Economic Research, URL: http://www.nber.org/papers/w17853.

Foster, Lucia, John Haltiwanger, and C. J. Krizan. 2006. "Market Selection, Reallocation, and Restructuring in the U.S. Retail Trade Sector in the 1990s." *The Review of Economics and Statistics*, 88(4): 748–758.

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2008b. "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *The American Economic Review*, 98(1): 394–425.

Fox, Jeremy T., and Valérie Smeets. 2011a. "Does Input Quality Drive Measured Differences in Firm Productivity?" Working Paper 16853, National Bureau of Economic Research, URL: http://www.nber.org/papers/w16853.

Garber, Alan M., and Jonathan Skinner. 2008. "Is American Health Care Uniquely Inefficient?" *The Journal of Economic Perspectives*, 22(4): 27–50.

Gawande, Atul. 2009. "The Cost Conundrum." *The New Yorker*, June 1.

Gaynor, Martin, and William B. Vogt. 2003. "Competition among Hospitals." *The RAND Journal of Economics*, 34(4): 764–785.

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. Robert Gibbons, and John Roberts, Chap. 17, Princeton: Princeton University Press, 680–731.

Gilman, Boyd H. 2000. "Hospital Response to DRG Refinements: The Impact of Multiple Reimbursement Incentives on Inpatient Length of Stay." *Health Economics*, 9(4): 277–294.

Goldstraw, Peter, David Ball, James R. Jett, Thierry Le Chevalier, Eric Lim, Andrew G. Nicholson, and Frances A. Shepherd. 2011. "Non-Small-Cell Lung Cancer." *The Lancet*, 378(9804): 1727 – 1740.

Gottlieb, Daniel J., Weiping Zhou, Yunjie Song, Kathryn Gilman Andrews, Jonathan S. Skinner, and Jason M. Sutherland. 2010. "Prices Don't Drive Regional Medicare Spending Variations." *Health Affairs*, 29(3): 537–43.

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." Special Advisory Bulletin, Office of Inspector General, URL: https://oig.hhs.gov/fraud/docs/alertsandbulletins/gainsh.htm.

Higashi, Takahiro, Neil S. Wenger, John L. Adams, Constance Fung, Martin Roland, Elizabeth A. McGlynn, David Reeves, Steven M. Asch, Eve A. Kerr, and Paul G. Shekelle. 2007. "Relationship Between Number of Medical Conditions and Quality of Care." *The New England Journal of Medicine*, 356(24): 2496–504.

Hoffman, Philip C., Ann M. Mauer, and Everett E. Vokes. 2000. "Lung Cancer." *The Lancet*, 355(9202): 479 – 485.

Hortaçsu, Ali, and Chad Syverson. 2007. "Cementing Relationships: Vertical Integration, Foreclosure, Productivity, and Prices." *Journal of Political Economy*, 115(2): 250–301.

Jacob, Brian A., and Lars Lefgren. 2007. "What Do Parents Value in Education? An Empirical Investigation of Parents' Revealed Preferences for Teachers." *The Quarterly Journal of Economics*, 122(4): 1603–1637.

Jha, Ashish K., Zhonghe Li, E. John Orav, and Arnold M. Epstein. 2005. "Care in U.S. hospitals–The Hospital Quality Alliance Program." *The New England Journal of Medicine*, 353(3): 265–74.

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

Kane, Thomas J., and Douglas O. Staiger. 2001. "Improving School Accountability Measures." *National Bureau of Economic Research Working Paper Series*, No. 8156.

Keeley, Ellen C., Judith A. Boura, and Cindy L. Grines. 2003. "Primary Angioplasty versus Intravenous Thrombolytic Therapy for Acute Myocardial Infarction: A Quantitative Review of 23 Randomised Trials." *Lancet*, 361(9351): 13–20.

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.

Martin, Ralf. 2005. "Productivity Dispersion, Competition and Productivity Measurement." CEP Discussion Paper dp0692, Centre for Economic Performance, URL: http://ideas.repec.org/p/cep/cepdps/dp0692.html.

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. 1996. "Medicare Reimbursement and Hospital Cost Growth." In *Advances in the Economics of Aging*. ed. David A. Wise: University of Chicago Press, 149–190.

McClellan, Mark B., 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.

McClellan, Mark B., and Douglas O. Staiger. 1999. "The Quality of Health Care Providers." Working Paper 7327, National Bureau of Economic Research, URL: http://www.nber.org/papers/w7327.

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. David M. Cutler: University of Chicago Press, 93–112.

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

MEDPAC. 2012a. "Health Care Spending and the Medicare Program." June 2012, URL: http://www.medpac.gov/documents/Jun12DataBookEntireReport.pdf.

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

Melitz, Marc J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica*, 71(6): 1695–1725.

Morris, Carl N. 1983. "Parametric Empirical Bayes Inference: Theory and Applications." *Journal of the American Statistical Association*, 78(381): 47–55.

Newhouse, Joseph P. 1996. "Reimbursing Health Plans and Health Providers: Efficiency in Production versus Selection." *Journal of Economic Literature*, 34(3): 1236–1263.

Office of Management and Budget. 2009. "McAllen Redux." URL: http://www.whitehouse.gov/omb/blog/09/06/04/McAllenRedux/.

Olley, G. Steven, and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica*, 64(6): 1263–97.

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.

Pavcnik, Nina. 2002. "Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants." *The Review of Economic Studies*, 69(1): 245–276.

Pear, Robert. 2009. "Health Care Spending Disparities Stir a Fight." *New York Times*, June 8, 2009, URL: http://www.nytimes.com/2009/06/09/us/politics/09health.html.

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.

Propper, Carol, and John Van Reenen. 2010. "Can Pay Regulation Kill? Panel Data Evidence on the Effect of Labor Markets on Hospital Performance." *Journal of Political Economy*, 118(2): 222–273.

Sacarny, Adam. 2014. "Technological Diffusion Across Hospitals: The Case of a Revenue-Generating Practice." MIT Mimeo.

Scarpetta, Stefano, Philip Hemmings, Thierry Tressel, and Jaejoon Woo. 2002. "The Role of Policy and Institutions for Productivity and Firm Dynamics: Evidence from Micro and Industry Data." Working Paper 329, OECD Economics, URL: http://ideas.repec.org/p/oec/ecoaaa/329-en.html.

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

Skinner, Jonathan S. 2011. "Causes and Consequences of Regional Variations in Health Care." In *Handbook of Health Economics*. eds. Mark V. Pauly, Thomas G. Mcguire, and Pedro P. Barros, Volume 2: Elsevier, 45–93.

Skinner, Jonathan S., Daniel J. Gottlieb, and Donald Carmichael. 2011. "A New Series of Medicare Expenditure Measures by Hospital Referral Region: 2003-2008." Dartmouth Atlas Project Report, The Dartmouth Institute, URL: http://www.dartmouthatlas.org/downloads/reports/PA_Spending_Report_0611.pdf.

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

Skinner, Jonathan S., and Douglas O. Staiger. 2009a. "Technology Diffusion and Productivity Growth in Health Care." Working Paper 14865, National Bureau of Economic Research, URL: http://www.nber.org/papers/w14865.

Skinner, Jonathan S., Douglas O. Staiger, and Elliott S. Fisher. 2006a. "Is Technological Change in Medicine Always Worth It? The Case of Acute Myocardial Infarction." *Health Affairs*, 25(2): w34–47.

Sloan, Frank A. 2000. "Not-for-Profit Ownership and Hospital Behavior." In *Handbook of Health Economics*. eds. 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.

Stock, James H., Jonathan H. Wright, and Motohiro Yogo. 2002. "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments." *Journal of Business & Economic Statistics*, 20(4): 518–29.

Syverson, Chad. 2004a. "Product Substitutability and Productivity Dispersion." *The Review of Economics and Statistics*, 86(2): 534–550.

Syverson, Chad. 2004b. "Market Structure and Productivity: A Concrete Example." *Journal of Political Economy*, 112(6): 1181–1222.

Syverson, Chad. 2011b. "What Determines Productivity?" *Journal of Economic Literature*, 49(2): 326–365.

Trude, Sally. 1992. "Physicians' Hospital Visits: A Description of Rates of Visits Under Medicare." 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.

Weisbrod, Burton A. 1991. "The Health Care Quadrilemma: An Essay on Technological Change, Insurance, Quality of Care, and Cost Containment." *Journal of Economic Literature*, 29(2): 523–52.

Yasaitis, Laura, Elliott S. Fisher, Jonathan S. Skinner, and Amitabh Chandra. 2009. "Hospital Quality and Intensity of Spending: Is there an Association?" *Health Affairs*, 28(4): w566–72.

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-electronic-health-records.

Appendices for Chapter 2

Appendix A

Analytical Framework

As mentioned in the text, models of reallocation mechanisms among heterogeneous-productivity producers have found applications in a number of fields, including industrial organization, trade, and macro-economics. While these models differ considerably in their specifics, they share an archetypal mechanism that connects the extent of competition in the market to the shape of the productivity distribution among market producers. We describe this central mechanism here.

Producers (indexed by i) earn profits which depend positively on their idiosyncratic productivity levels A_i - more productive firms earn higher profits due to their lower costs - and negatively on the number (or mass, in models with a continuum of firms) of producers in the industry N.[1] Hence $\pi_i = \pi\left(A_i, N\right)$, with $\frac{\partial \pi}{\partial A_i} > 0$ and $\frac{\partial \pi}{\partial N} < 0$. The monotonic relationship between productivity and profits implies that, for any given N, there is a critical cutoff productivity level $A^*\left(N\right)$ at which firm profits are zero. Only producers with productivity levels at or above $A^*\left(N\right)$ will operate in equilibrium.

The zero-profit cutoff productivity $A^*\left(N\right)$ is endogenously determined by a free entry condition, where ex-ante identical potential entrants consider whether to pay a sunk cost σ to take an idiosyncratic productivity draw from a known distribution, $G\left(\cdot\right)$ with upper bound $\bar{A}$. The expected value of entry, which equals zero by the free entry condition, is:

$$V^e = \int_{A^*(N)}^{\bar{A}} \pi\left(A, N\right) g\left(A\right) dA - \sigma = 0$$

[1]Standard presentations of these models consider profit-maximizing firms. Although we keep this terminology to be more familiar relative to the existing literature, we note that in the context of hospitals, it might be more appropriate to consider firms as earning (and maximizing) "surplus" rather than "profits". This more general terminology recognizes that many hospitals are legally structured as nonprofits and does not affect the qualitative comparative statics. Nonprofit hospitals are often modeled in the literature as having an objective function that is a convex combination of profits and other objectives; therefore on the margin they should respond qualitatively the same way as for-profit hospitals to factors like competition. Moreover, even if a hospital's objective is not profit maximization, it is likely that for any given level of output(s) the hospital produces (in order to meet whatever outcomes are in its objective function), surplus will be larger if the hospital's costs are lower. In practice, a large empirical literature finds essentially no evidence of differential behavior across for-profit and non-profit hospitals, calling into question whether the non-profit label has any substantive meaning for behavioral responses (see Sloan, 2000 for a recent review of this literature).

The expected profits from entry depend upon the equilibrium number of entrants N in two ways. First, an increase in N shifts upward the zero-profit cutoff productivity level $A^*(N)$, reducing the probability that the entrant's productivity draw is high enough to earn nonnegative profits and thus making successful entry less likely. Second, a higher number of firms N also reduces the producer's profits if it does enter. Thus expected profits fall monotonically in N. In equilibrium, the number of firms choosing to pay the entry cost yields a number of entrants N that, through these two effects, exactly equates the expected profit from taking a productivity draw to the sunk entry cost.

The endogeneity of $A^*(N)$ means the industry productivity distribution observed in the data is determined in equilibrium. Specifically, it is a truncation of $G(\cdot)$, the underlying productivity distribution from which potential entrants take productivity draws, where the truncation point is $A^*(N)$. Changes in market primitives that shift the equilibrium location of $A^*(N)$ therefore shift the observed productivity distribution as well.

The primitive that we are interested in here is the extent of competition, as reflected in how easily consumers can (or how willing consumers are to) substitute to alternate producers. The specific mechanism through which primitives map into substitutability may vary, from changes in the differentiation of firms' products, to shifts in openness to trade, to movements in the size of transport costs. The particulars of the mechanism aren't important here; what matters are the effects on the equilibrium.

Higher substitutability has three effects that can be examined empirically. First, it makes it more difficult for higher-cost (lower-productivity) firms to earn positive profits, as demand is now more responsive to their cost and price differential relative to other firms in the industry.[2] In turn, the zero-profit cutoff productivity level $A^*(N)$ rises: the threshold for operation is greater than before. This truncates the equilibrium productivity distribution, reducing observed *productivity dispersion.*[3] Second, higher substitutability means that, among operating firms, market shares are more sensitive to productivity differences. Purchases are reallocated to more productive firms, raising the correlation between productivity and market share at a point in time (*"static allocation")*. Third, over time more productive firms are likely to grow in market share (*"dynamic allocation")*.[4]

[2]In the case of hospitals, this demand response can be manifested either directly in patients' choices in response to out-of-pocket costs, or indirectly through insurers' decisions to include the hospital in its covered network.

[3]This dispersion implication requires some additional regularity assumptions on the underlying productivity distribution. Most "standard" distributions exhibit declining second moments as they are truncated from below. The exponential distribution, however, is an example of one that does not. Nevertheless, if we assume the productivity distribution is bounded at the top (i.e., there is some maximum productivity level), as we do here, then all distributions will eventually exhibit decreased dispersion as they are truncated from below.

[4]The model just described is static, so the effects of changes in competition on equilibrium should be thought of as comparing two different markets or the same market across different long-run steady states. However, several of the models in the literature are explicitly dynamic and have similar predictions about the effect of competition on the productivity of entrants and growth of incumbents (e.g. Hopenhayn, 1992; Asplund and Nocke, 2006).

Appendix B

Measuring Inputs

Our baseline input measure (as well as many of the alternative measures discussed below) is derived from the formulas used to determine Medicare's Hospital (Part A) reimbursement. Some alternative measures also use information derived from the formulas used to determine Medicare's reimbursement of physicians and outpatient facilities (Part B). It is therefore useful to begin with a very brief overview of the key features of Medicare hospital reimbursement needed to understand the construction and composition of our baseline and alternative input measures. Considerably more detail can be found in Centers for Medicare and Medicaid Services (2011).

The amount Medicare reimburses a hospital is determined by the patient's Diagnosis Related Group (DRG), national factors, and hospital-specific factors. A patient's DRG is a function of his principal diagnosis, procedures performed, and secondary complications and comorbidities. Some DRGs also depend on whether the patient died in the hospital.

Each DRG is assigned a (national) weight based on how much it costs to treat the nationwide average patient with that DRG; a national conversion factor is used to convert these DRG weights into dollar payments. The weights and the conversion factor are updated annually. The national rate is then adjusted for hospital-specific considerations. The major adjustments are due to geographic factors (e.g. the local wage rate) and characteristics of the hospital (such as whether it operates a resident training program or has a disproportionate share of patients on Medicare or SSI).

For most stays the hospital will receive payments solely based on the patient's DRG. However, in certain extraordinarily costly cases hospitals receive additional "outlier payments" covering 80 percent of costs beyond a threshold level. To compute costs, the hospital's billed charges are deflated by a hospital-specific cost-to-charge ratio. If a patient has a short stay and is transferred to another hospital, Medicare reduces payments to the transferring hospital but pays the receiving hospital as it would for a standard inpatient stay. For our purposes, we assign all inputs for the patient in the time horizon (30 days for our baseline measure) back to the initial hospital.

B.1 Baseline Input Measure: Part A "Resources"

Our baseline input measure follows the approach of Gottlieb et al. (2010) and Skinner and Staiger (2009a) to purge the "price" variation in the reimbursement formula from the "input" variation. Specifically, our starting point is the DRG weight (multiplied by a national conversion factor to convert it to a dollar metric) plus outlier payments (also in dollars). It does not reflect any variation in reimbursement prices across hospitals due to geographic factors or specific characteristics of the hospital.

According to this measure, the inputs a patient receives equal the sum of his converted DRG weights and outlier payments at all hospital stays in the 30 days following his AMI. Variation across patients in the input measure therefore comes from 3 sources: variation in the patient's DRG(s); whether there are (and the extent of) outlier payments; and the number of hospital stays during the 30 day window. We discuss each in turn.

B.1.1 Variation across Index Event DRGs

To give a sense of the nature and variation across DRGs, Table A1 lists the top 20 DRGs for the index event (initial AMI hospital stay), their patient share and their weights in 2000.[1] The top five DRGs account for over 90 percent of the index events, and the top 20 account for virtually 100 percent.

Looking within the top five we see substantial differences in weight based on whether an invasive procedure is performed. There are two separate DRGs for invasive procedures (#107, "Coronary Bypass with Cardiac Catheterization" and #116, "Other Permanent Cardiac Pacemaker Implant or PTCA with Coronary Artery Stent Implant") and they respectively have weights of 5.46 and 2.47. By contrast, the other three DRGs in the top five are medical DRGs (i.e. do not involve invasive procedures) and have weights ranging from 1.11 to 1.51. For the year 2000, two dummies for these two surgical DRGs (bypass and stent) explain 15 percent of the total variation in our 30 day input measure.

Within the three most common medical DRGs, we see that there is variation for a medically treated AMI based on whether or not the patient died (#123), survived following a stay with major complications (#121) or survived following a stay without major complications (#122). This variation has, to our knowledge, not previously been noted by the large empirical literature on the relationship between inputs for heart attacks and subsequent survival which has used the variation in inputs stemming from survival. However, this source of variation in the standard input measure seems suspect: it partly causes in-hospital death - not inputs, per se - to explain survival, an association that must exist trivially.

[1] For presentation purposes, we limit Table A1 to one year because DRG weights and classifications change slightly from year to year.

Therefore, for these three DRGs that refer to the same diagnosis but differ on the basis of patient survival, we eliminate the variation in inputs across DRGs within this group at the hospital-year level. We assign each DRG the patient-weighted average of the different DRG weights. The averaging weights are equal to the share of patients in the DRG in that year. Almost three-quarters of hospital stays were grouped into DRGs that were affected by this fix.[2]

B.1.2 Variation from Outlier Payments

Approximately 8.2% of our patients trigger outlier payments due to unusually costly cases. These payments are triggered when a hospital's cost of treating a patient exceeds a national threshold. Conditional on receiving an outlier payment, the average outlier payment as a share of DRG reimbursement without outlier payments is 53.9; the standard deviation of outlier payments is 13,154.8. (All statistics calculated for patients in the year 2000.)

B.1.3 Variation Due to Number of Hospital Stays

Even ignoring outlier payments, the total variation coming from DRGs is in fact larger than that indicated in Table A1 because of the possibility of multiple (and potentially non AMI) hospital stays in the 30 days following the index event (AMI). Our baseline input measure is constructed for the 30 days following the initial AMI, meaning that it includes all hospital stays in these 30 days. On average, an AMI patient has 1.07 stays in this window. Conditional on having multiple stays, the average patient visits the hospital 2.07 times in the month following the AMI.

If a hospital stay straddles the end of the time window (e.g. a patient stays in the hospital for 10 days and is admitted on day 25 days following the heart attack), the inputs attributed to that hospital are reduced; in particular, we multiply our input measure by the share of days in the hospital that were inside the 30 day analysis window. We adjusted all DRGs (not just those associated with index events) to purge variation stemming from mortality in the manner described above.

[2]Note that this "fix" also purges the variation across the three most common medical DRGs in whether the patient had a major complication or not. Although the case in question is the only one where different DRGs are assigned based on patient survival, there are other cases where separate DRGs are assigned based on the presence of complicating conditions (CCs). For example, the 6th-ranked DRG #110, "Major Cardiovascular Procedures with CC" (weight 4.16) and the 18th-ranked DRG #111, "Major Cardiovascular Procedures without CC" (weight 2.23) differ only on this basis. It is a priori unclear to us whether we want to purge variation due to the presence of CCs. On the one hand, conditional on a rich set of patient risk adjusters, the presence of a CC may be a useful measure of the intensity of resources required to treat the condition; on the other hand, with imperfect risk adjusters, it may also capture correlates of mortality (our outcome of interest).

As noted, in practice our approach to purging mortality-based variation across DRGs also purges complications-based variation in the most common DRGs. We experimented with an alternative measure that purged variation due to CCs in all DRGs. The procedure took DRGs that were identical but for the CC requirement and assigned them the same DRG weight within each hospital-year. This DRG weight was a weighted average of the component DRG weights; the averaging weights were the shares of patients in each DRG in the hospital-year. For example, in 2000, DRGS #110 and #111 were assigned the same weight in each hospital-year. This correction affected only a few percent more patients and made no noticeable difference to our findings (results available on request).

Table A2 lists the top 20 DRGs across all stays in the 30 day window following the index event. The index events are included in this table. As expected, there is more variation across these DRGs.

B.1.4 Empirical Variation in Baseline Input Measure

The panels of Figure A1 show the variation in the input measures across patients for one year (2000). Figure A1a shows the variation in the DRG index events (using our "collapsed" DRG measure that purges mortality variation). Figure A1b shows the variation from the DRG index events plus outlier payments in the index event. Figure A1c shows the total 30 day variation (our baseline input measure), which adds in additional hospital stays (their DRGs and outlier payments) within the 30 days. As would be expected, the input distribution gets less "lumpy" at each step.

B.2 Alternative Input Measures

We confronted a number of choices in defining our baseline input measure. We therefore constructed several other alternative input measures. This section describes them.

B.2.1 Alternative Measures of Hospital Inputs

A central tension in our choice of input measurement is how coarse or detailed we make our input measure. The tradeoff is between the survival bias that can occur with finer input measures - since the longer a patient survives, the more can be done to a patient - and the measurement error which occurs at coarser definitions of inputs. Our baseline measure, following standard practice, is aggregated to a relatively high level, and may therefore measure inputs with a non-trivial amount of error.

We experimented with two alternative hospital-based input measures. One measures Part A spending rather than Part A inputs; it therefore includes variation in reimbursement rates stemming from hospital specific factors like geographic location or type of hospital. As shown in Figure A1d the distribution of Part A reimbursement is less "lumpy" than our baseline input measure; the correlation between the two is 0.90.

The other measure is designed to be more detailed than our baseline measure to reflect that fact that input use may vary substantially within the relatively coarse DRGs. We used data on the length of hospital stay and the procedures performed during the stay (up to six may be listed). Procedure codes are themselves available at different levels of granularity; there are 3 levels of CCS procedure codes ranging from the least granular level 1 to the most granular level 3; the much larger set of ICD-9 procedure codes is more granular still. The ICD-9 codes account for over 3878 possible procedures that may be performed on patients.

To reduce the dimensionality of the set of procedures, we use the following algorithm. We start with the coarsest set of procedures (level 1 CCS codes, of which there are 16) and move iteratively to the finest set of procedure codes (ICD-9). At each step we aggregate codes that are rare and disaggregate codes that are very common. Thus, beginning with CCS level 1 codes, we include indicators for level 1 procedures that were performed on less than 10% of patients; if the level 1 procedure was performed on 10% or more of patients, we disaggregate it by looking at CCS level 2 components.

In similar fashion, if the CCS level 2 procedures were performed on 1-10 percent of patients, we include an indicator for it. Within a level 1 code, all level 2 codes performed on less than 1 percent of patients are grouped together and included as one indicator. If the level 2 procedure was performed on 10 percent or more of patients, we disaggregate by looking at its level 3 components.

We follow the same process for level 3 components; when we disaggregate these codes we look at the component ICD-9 codes. If the ICD-9 code was performed on at least 1 percent of patients we include an indicator for it. Within a level 3 code, all ICD-9 codes that were performed on less than 1 percent of patients are grouped together and included as one indicator.

This algorithm results in 60 procedure indicators: 18 for ICD-9 codes, 6 for level 3 CCS codes, 22 for level 2 CCS codes and 14 for level 1 CCS codes.

B.2.2 Incorporating Non-Hospital Inputs

A limitation of our input measures thus far is that, following standard practice in the heart attack literature, they reflect only inpatient hospital inputs. Notably, they do not include physician inputs, which may occur in an inpatient or outpatient setting. They also do not include outpatient tests and procedures like MRIs.

Many of these inputs are directly related to the treatment of the AMI. For example, the work of physicians who treat the patient surgically or medically in the hospital is obviously an input that may bear on the patient's survival. Likewise, an MRI done in an outpatient facility that is closely affiliated with the hospital will inform treatment decisions and influence mortality.

There are two reasons why we follow most of the literature on heart attacks and do not include inputs by physicians or outpatient facilities in our baseline measure. First, while some of these inputs are closely linked to the care received in the hospital, many of the payments reflect care that is independent of the hospital. In particular, doctor visits and outpatient diagnostic tests at long time horizons from the initial AMI admission may be less dependent on initial treatment decisions. The second reason is practical: data on much of these other input measures are only available for 20 percent of the sample and only since mid-2000, reducing the set of hospital-years in which we can observe at least 5 AMI patients by 70.0%.

Still, we sought to evaluate the sensitivity of our results to including physician and outpatient services. Medicare reimburses physicians based on their assessment of the "Relative Value Units" (RVUs) of the services the physician provided; the RVU of a service is intended to reflect the resources required to provide that service. The RVUs attributed to procedures are constant across geographic areas and practitioners, although Medicare makes further adjustments based on geography and provider type to derive reimbursement rates (see MedPAC, 2010b or Clemens and Gottlieb, 2014 for more details). We construct our measure of physician inputs by summing all RVUs associated with the patient in the 30 days following his initial hospital admission. We multiply the RVUs by a national conversion factor to convert them to a dollar metric; the national conversion factor eliminates variation due to Medicare's geographic price adjustments.

Calculating outpatient contributions to the production function is significantly more complicated than calculating physician or inpatient contributions. While physician services and inpatient stays are each reimbursed using a single payment system that is designed to reflect resource utilization, different outpatient services are covered by different types of systems (MedPAC, 2010a provides more details). Some outpatient services are covered prospectively - although the payment groups are so fine that treatment decisions may be reimbursed at the margin. Providers are paid for other services according to a fee schedule that is geographically adjusted. Some services are reimbursed according to local prices.

For the portion of outpatient services covered prospectively, there is a series of classification groups (Ambulatory Payment Classification groups or APCs) which function analogously to DRGs. Each APC is given a weight that is based on its expected resource costs; we translate these weights into a dollar basis using a national conversion factor that is an analogous to the procedure we use to convert DRG weights. For services that are reimbursed on a fee schedule, we mimic the method used for physician inputs by applying the fee schedule prior to geographic adjustments.

These adjustments eliminate much of the variation in outpatient prices that is region- or provider-specific. Still, some payments, like those for certain prescription drugs and new technologies, do not have an associated national fee schedule and are included unadjusted.

Appendix C

Empirical Bayes Adjustment

C.1 Introduction

In this appendix we describe the empirical Bayes (EB) procedure we use to adjust our estimates of hospital productivity for measurement error. This procedure is based on Morris (1983). For another example see Jacob and Lefgren (2007).

The exponentiated productivity of hospital h at time t is A_{ht} and its productivity is $a_{ht} = \ln(A_{ht})$. These objects are the "true" productivities and their distribution is the "underlying" distribution of productivity. We denote by $\hat{a}_{ht}$ the estimate of productivity; it equals productivity plus an error term η_{ht}:

$$\hat{a}_{ht} = a_{ht} + \eta_{ht}$$

The goal of the EB procedure is to adjust the estimates of productivity so that the presence of the error term does not introduce bias into our regressions, which use our estimate of productivity ($\hat{a}_{ht}$) as a key right hand side variable. The procedure adjusts the estimates by shrinking them toward the mean of the true, underlying productivity distribution.

True productivity is not observable, but we show in this appendix that its distribution is estimable. We also show how this shrinkage estimator fixes the attenuation bias that measurement error would otherwise introduce into our regressions.

C.2 Background on Empirical Bayes Procedure

C.2.1 Statistical Background

We start with an overview of the EB procedure assuming that all parameters of the distributions are known, and refer to the EB-adjusted estimated productivity as a_{ht}^{EB}. We then describe the feasible EB-adjusted estimate, which we denote $a_{ht}^{EB(f)}$.

Suppose that the estimated productivities are independently normally distributed around the true productivities with known variance π_{ht}^2:

$$\hat{a}_{ht}|a_{ht}, \pi_{ht}^2 \sim N\left(a_{ht}, \pi_{ht}^2\right) \text{ independently}$$

One can think of π_{ht}^2 as the variance of the measurement error of the estimate.

We also assume that the true productivities a_{ht} are independently normal with underlying mean $x_{ht}\beta_t$ (a known, year-specific linear function of the hospital-year's covariates) and underlying variance $\sigma_{a,t}^2$ (known and common across hospitals within a year).

The *prior distribution* of the productivity a_{ht} – the distribution before conditioning on the estimated productivity – is therefore:

$$a_{ht}|x_{ht}, \beta_t, \sigma_{a,t}^2 \sim N\left(x_{ht}\beta_t, \sigma_{a,t}^2\right) \text{ independently}$$

Conditioning on the estimated productivity $\hat{a}_{ht}$ produces the *posterior distribution* of a_{ht}:

$$a_{ht}|\hat{a}_{ht}, x_{ht}, \beta_t, \sigma_{a,t}^2, \pi_{ht}^2 \sim N\left(a_{ht}^{EB}, \pi_{ht}^2\left(1 - B_{ht}\right)\right) \tag{C.1}$$

a_{ht}^{EB} denotes the EB adjusted productivity. This object is the expected value of a_{ht} conditional on the estimated value $\hat{a}_{ht}$ and the parameters $\beta_t, \sigma_{a,t}^2$, and π_{ht}^2 and is given by the formula:

$$\begin{aligned} a_{ht}^{EB} &= \left(1 - B_{ht}\right)\hat{a}_{ht} + B_{ht}x_{ht}\beta_t \\ B_{ht} &= \pi_{ht}^2 / \left(\pi_{ht}^2 + \sigma_{a,t}^2\right) \end{aligned}$$

The adjustment amounts to attenuating the estimate $\hat{a}_{ht}$ toward the mean $x_{ht}\beta_t$. As the variance of the measurement error π_{ht}^2 rises, the EB correction increasingly disregards the value of the estimate and closes in on the mean.

C.2.2 Feasible Version of Procedure

This section describes how we implement the EB procedure when parameters must be estimated.

The productivity estimate $\hat{a}_{ht}$ is the estimated coefficient on a hospital-year fixed effect from equation 2.5. The regression that produces the estimated coefficient also yields a standard error for it – an estimate of the standard deviation of the asymptotic distribution of $\hat{a}_{ht}$. We estimate π^2_{ht} by squaring the standard error and call this value $\hat{\pi}^2_{ht}$.

We estimate β_t and $\sigma^2_{a,t}$ using a method outlined in section 5 of Morris (1983) which we reproduce here. Fix yearly estimates:

$$\begin{aligned}
\hat{\beta}_t &:= \left(X_t' W_t X_t\right)^{-1} X_t' W_t A_t \\
\hat{\sigma}^2_{a,t} &= \max\left\{0, \frac{\sum_h W_{ht}\left\{\left(\frac{N_{ht}}{N_{ht}-N_x}\right)\left(\hat{a}_{ht} - x_{ht}\hat{\beta}_t\right)^2 - \hat{\pi}^2_{ht}\right\}}{\sum_h W_{ht}}\right\} \\
W_{ht} &:= \frac{1}{\hat{\pi}^2_{ht} + \hat{\sigma}^2_{a,t}}
\end{aligned}$$

X_t is the stacked x_{ht} for year t, W_t is a diagonal matrix of the W_{ht} for year t, and A_t is the stacked $\hat{a}_{ht}$ for year t. N_{ht} is the number of hospitals, or equivalently the number of $\hat{a}_{ht}$, in year t. N_x is the number of regressors, i.e. the dimensionality of x_{ht}.

$\hat{\beta}_t$ is a WLS regression of the $\hat{a}_{ht}$ on x_{ht}. $\hat{\sigma}^2_{a,t}$ is the weighted average of the squared deviations of $\hat{a}_{ht}$ from $x_{ht}\hat{\beta}_t$ less the weighted average of $\hat{\pi}^2_{ht}$. The weights are W_{ht}, giving more weight to observations with less measurement error. The max operator ensures that $\hat{\sigma}^2_{a,t}$ is always nonnegative in finite samples.

$\hat{\beta}_t$ and $\hat{\sigma}^2_{a,t}$ are simultaneously determined in these equations, so for each year they are estimated by the following iterative procedure. We by fixing $W_{ht} = 1 \,\forall\, h$, then iterate the following to convergence:

1. Compute $\hat{\beta}_t$ and then a new estimate $\hat{\sigma}^2_{a,t}$
2. If $\hat{\sigma}^2_{a,t}$ has converged, exit. Otherwise, fix new weights W_{ht} and return to step 1

With a degrees of freedom correction, the (feasible) best estimate of the posterior mean $a^{EB(f)}_{ht}$ is:

$$\begin{aligned}
a^{EB(f)}_{ht} &= \left(1 - \hat{B}_{ht}\right)\hat{a}_{ht} + \hat{B}_{ht} x_{ht}\hat{\beta}_t \\
\hat{B}_{ht} &= \left(\frac{N_{ht} - N_x - 2}{N_{ht} - N_x}\right)\left(\frac{\hat{\pi}^2_{ht}}{\hat{\pi}^2_{ht} + \hat{\sigma}^2_{a,t}}\right)
\end{aligned}$$

The variance of productivity unconditional on covariates, called $\hat{\varsigma}^2_{a,t}$, is given by the following formula:

$$
\begin{aligned}
\hat{\varsigma}^2_{a,t} &= \max\left\{0, \frac{\sum_h W_{ht}\left\{\left(\frac{N_{ht}}{N_{ht}-1}\right)(\hat{a}_{ht} - \bar{A}_t) - \hat{\pi}^2_{ht}\right\}}{\sum_h W_{ht}}\right\} \\
\bar{A}_t &= \frac{\sum_h W_{ht}\hat{a}_{ht}}{\sum_h W_{ht}}
\end{aligned}
$$

Where $\bar{A}_t$ is the weighted mean productivity.

C.3 Implementation of Empirical Bayes Adjustment

We assume that the underlying mean of productivity is equal to a market-year fixed effect, i.e. $x_{ht}\beta_t = \tau_{M,t}$, where M indexes markets. Thus x_{ht} becomes a vector of 304 indicators for whether hospital h was in each of the 304 markets and β_t is a vector of the 304 market fixed effects for year t.

We perform the EB procedure separately year-by-year, producing estimates of the underlying market-year means $\hat{\beta}_t$ and year-specific conditional – i.e. within-market – variance $\hat{\sigma}^2_{a,t}$. Running the procedure also yields EB-adjusted estimated productivities $a^{EB(f)}_{ht}$ and also can be used to produce the unconditional – i.e. national – estimated variance $\hat{\varsigma}^2_{a,t}$, as described below.

Our procedure ensures that when the EB-adjusted productivities are used in our main regressions (equations 2.1 through 2.3 in the main text) which have market-year fixed effects, all regressors are orthogonal to the measurement error term.

C.4 Reported productivity metrics

C.4.1 Standard Deviation

To estimate the standard deviation of productivity using the EB adjusted values, we rely on the estimates of the yearly underlying national variance of productivity $\hat{\varsigma}^2_{a,t}$ that the procedure computes.[1] The root of these estimates is taken, forming $\hat{\varsigma}_{a,t}$. The yearly values are then averaged together.

The EB adjustment produces $\hat{\varsigma}^2_{a,t}$ by taking the weighted empirical variance of the $\hat{a}_{ht}$ and subtracting the weighted average squared standard error $\hat{\pi}^2_{ht}$. Hospital-years with larger standard errors receive lower weights. In effect, this process takes the variance of the noisy productivity estimates and subtracts off the variance due to measurement error.

[1]While it might seem natural to instead estimate the standard deviation of the EB-adjusted values, this would cause us to erroneously under-estimate dispersion. Underlying productivity is composed of a best prediction (the EB-adjusted productivity) and the prediction error. These two components are orthogonal. The variance of true productivity is thus strictly greater than the variance of EB-adjusted productivity (see Jacob and Lefgren, 2007).

C.4.2 90:10 and 75:25

We define the 90:10 ratio of productivity as $F^{-1}(0.9) - F^{-1}(0.1)$ and the 75:25 ratio as $F^{-1}(.75) - F^{-1}(.25)$ where F^{-1} is the inverse CDF of the productivity distribution. The 90:10 is the 90th percentile value of the distribution minus the 10th percentile value, and likewise for the 75:25. Exponentiating these ratios would produce the 90:10 ratio of the exponentiated productivity distribution (that is, an actual ratio: p90 / p10).

As with the standard deviation, it is not possible to estimate these ratios using the distribution of the $a_{ht}^{EB(f)}$. The EB correction does not produce a variable with the same asymptotic distribution as the underlying process. The procedure is only intended to estimate the parameters of an underlying normal distribution and correct for measurement error in regressions.

To estimate these ratios we use the inverse CDF of the underlying normal distribution that the EB procedure uncovers, so the yearly 90:10 and 75:25 are:

$$\begin{aligned} F^{-1}(0.9) - F^{-1}(0.1) &= \hat{\varsigma}_{a,t}\left[\Phi^{-1}(0.9) - \Phi^{-1}(0.1)\right] \\ F^{-1}(0.75) - F^{-1}(0.25) &= \hat{\varsigma}_{a,t}\left[\Phi^{-1}(0.75) - \Phi^{-1}(0.25)\right] \end{aligned}$$

Where $\Phi(\cdot)$ is the standard normal CDF.

C.4.3 Allocation Metrics (Patient, Growth, and Exit Regressions)

The allocation metrics use noisy estimates of productivity on the right-hand side of regressions, and rely on EB adjustment to correct for measurement error. Jacob and Lefgren (2007) show that with the adjustment, these regressions are estimated consistently.

Suppose that there is a relationship between growth g_{ht}, market-year fixed effects $\gamma_{M,t}$, and productivity a_{ht}:

$$g_{ht} = \gamma_{M,t} + \delta a_{ht} + \epsilon_{ht}$$

where $\mathbb{E}\left[\epsilon_{ht} | x_{ht}, a_{ht}\right] = 0$ (x_{ht} is a vector of indicators for the market-years – the design matrix for the market-year fixed effects.) The left-hand side variable could alternatively be the number of patients or an indicator for hospital exit.

Since we do not observe true productivity a_{ht}, we use the estimate $\hat{a}_{ht} = a_{ht} + \eta_{ht}$, where η_{ht} is measurement error. Then substituting into the equation:

$$g_{ht} = \gamma_{M,t} + \delta \hat{a}_{ht} + (\epsilon_{ht} - \delta \eta_{ht})$$

This regression produces a biased and inconsistent estimate of δ due to the correlation between $\hat{a}_{ht}$ and η_{ht} in the error term. We use the EB-adjusted productivity a_{ht}^{EB} to eliminate this correlation. Equation C.1 implies:

$$\mathbb{E}\left[a_{ht}|\hat{a}_{ht}, x_{ht}, \beta_t, \sigma_{a,t}^2, \pi_{ht}^2\right] = a_{ht}^{EB}$$

We represent the prediction error of the EB procedure as v_{ht}:

$$a_{ht} = a_{ht}^{EB} + v_{ht}$$

By construction the prediction error is orthogonal to a_{ht}^{EB} and any regressor included in x_{ht} – i.e. the market-year fixed effects:

$$\mathbb{E}\left[v_{ht}|a_{ht}^{EB}, x_{ht}, \beta_t, \sigma_{a,t}^2, \pi_{ht}^2\right] = 0$$

($\hat{a}_{ht}$ is replaced by a_{ht}^{EB} because given the parameters, knowing one determines the other)

The regression of g_{ht} on market-year effects and a_{ht}^{EB} adds only v_{ht} to the original error term ϵ_{ht}:

$$g_{ht} = \gamma_{M,t} + \delta a_{ht}^{EB} + (\epsilon_{ht} - \delta v_{ht})$$

Therefore there is no correlation between any of the regressors and the new error term. The consistency of δ follows.

C.5 Comparison of estimates

We run all of our regression analyses with the EB-adjusted productivities $a_{h,t}^{EB(f)}$ and calculate our dispersion metrics using the EB-adjusted dispersion estimates as described above. Table A3 explores the impact of the EB correction on our main results. The first column reproduces the EB-adjusted main results from Tables 2.2, 2.4, and A6. The second column shows the results without the EB correction.

To produce the uncorrected allocation metrics, we use the estimates $\hat{a}_{ht}$ rather than $a_{h,t}^{EB(f)}$ in our regressions. Due to measurement error in the estimates, the allocation metrics computed without the EB correction will be attenuated. We calculate the uncorrected dispersion metrics in the same manner as the corrected versions, but using uncorrected estimates of productivity. For example, to calculate the standard deviation, the empirical weighted standard deviation of the estimated productivities – $SD(\hat{a}_{ht})$ – is taken year-by-year, then averaged (we use the same weights that were used to calculate $\hat{\varsigma}_{h,t}^2$ so that the statistics are comparable.) Likewise, the 90:10 and 75:25 ratios are calculated by fitting a normal distribution to the estimated, uncorrected productivities and reporting the ratios implied by it (the ratios are calculated year-by-year, then averaged). Due to measurement error, the dispersion metrics computed without the EB correction will overstate the true dispersion.

The results show that the EB correction has a substantial effect on our baseline estimates, and moves them in the expected direction. Comparing our baseline (EB-adjusted) estimates in column 1 with the un-adjusted version in column 2, we see that the allocation results are substantially larger and the dispersion estimates are substantially lower with the correction. For example, we find that measurement error explains nearly half of the dispersion of the productivity estimates; without correcting for measurement error, these estimates have an average yearly standard deviation $\mathrm{SD}(\hat{a}_{ht})$ of 0.293, while the EB procedure estimates that the underlying productivity process has an average yearly standard deviation $\hat{\varsigma}_{a,t}$ of 0.173.

A quantiatively large impact of the EB correction (i.e. a large amount of measurement error) is not surprising in light of results from other applications. For example, looking at estimates of teacher fixed effects in value added regressions, Jacob and Lefgren (2007) estimate a ratio of the unadjusted standard deviation to the EB-adjusted estimate of the standard deviation of about 1.3 to 1.6. We find ratios of about 1.7.

Appendix D

Additional Results

D.1 Counterfactual allocation rule

An alternative explanation for our findings is that patients go the nearest hospital to treat their AMI, and it happens that areas with higher productivity hospitals are both higher in population density and higher in population growth. If this story held, a mechanical allocation rule that assigned patients to their nearest hospital would spuriously produce our static and dynamic allocation results. In practice, based on geocoding hospital addresses and patient zip codes to latitudes and longitudes, we estimate that less than half of our AMI patients go to the nearest hospital in their market. Moreover, we examined what the static and dynamic allocation results would look like if (counter-factually) each AMI patient did go to the nearest hospital within his market. We would be concerned if this mechanical rule produced similar static and dynamic allocation results, as that would suggest the result could be generated without any role for patient demand. In fact, as shown in Table A4 (column 3 vs. column 1), with this assignment rule the dynamic allocation results are either the wrong sign or an order of magnitude smaller (and not statistically significant) and the static allocation result declines to 20 percent of the baseline estimate.

D.2 Static and Dynamic Allocation For Different Hospitals and Markets

Appendix Table A5 looks at how the static and dynamic allocation results vary across different types of hospitals within a market, and how they vary across different markets. The results are mixed. Within a market, the allocation results are stronger for hospitals facing more competition for their patients (using distance to the nearest hospital as a proxy for competition as in Gaynor and Vogt, 2003); the allocation relationships are also weaker for public (compared to private) hospitals. However, at the market level, there is no evidence that the allocation results are stronger in more competitive markets (using population density as a proxy for competition for a spatially differentiated product as in Syverson, 2004b); there is also no evidence that the allocation result is stronger in markets with more educated consumers.

D.3 Productivity Dispersion Across Hospitals

Appendix Table A6 shows our estimates of productivity dispersion across hospitals. The calculation of the metrics was described in Appendix C.

D.4 Static and Dynamic Allocation in Concrete and Health Care

We use data on ready-mixed concrete from the Census of Manufactures, which we have for every five years from 1972 - 1997. We observe approximately 2,500 ready-mixed concrete plants per data year; by way of comparison, we have approximately 3,700 hospitals per year. We use these data to estimate plants' physical total factor productivity levels. A plant's physical total factor productivity is the number of cubic yards of concrete it produces per unit input, where inputs are a weighted composite of labor, capital, and intermediates. The weights are the inputs' cost shares. These weights are theoretically correct, equaling the elasticities of output with respect to each input assuming cost minimization and no adjustment costs in inputs. Our market definition is the Bureau of Economic Analysis' Component Economic Areas, which are approximately 350 mutually exclusive and exhaustive groupings of economically interrelated U.S. counties. (See, e.g. Syverson, 2004b for more details on productivity and market measurement in ready-mixed concrete.) To reduce the influence of outliers, we trim the top and bottom 1% of the industry's productivity distribution in each Census of Manufactures.

Table A7 reports the results. Across all of our static and dynamic allocation measures, the results indicate a stronger relationship between market allocation and producer productivity for hospitals than for concrete plants. The first row reports the results for static allocation. We estimate a slight variant of equation 2.1; as before, the specification regresses output on productivity (both measures are in logarithms) and market-year fixed effects. However, we now use lagged productivity on the right-hand side to facilitate comparisons between hospitals and concrete plants.[1] Strikingly, the correlation between output and lagged productivity is an order of magnitude larger in healthcare than in concrete.

[1] Due to how productivity is measured for concrete plants, regressing output on contemporaneous productivity would yield spuriously expanded coefficients: for concrete, output is effectively the numerator of the productivity measure. To fix the bias, we use the productivity measure from 5 years earlier on the right-hand side, rather than contemporaneous productivity. The lag is 5 years for both sectors because data for concrete plants is only available at that frequency.

The second row reports our exit analysis, based on equation 2.2 but modified to account for the fact that in concrete we only have data every five years; therefore, for purposes of comparability, we look at exit five years later for both hospitals and for concrete. However, comparability is limited by the fact that "exit" is defined quite differently in the two data sets.[2]

The final row reports our growth analysis. To make the analysis comparable across the two industries, for both we run the following regression:

$$\frac{N_{h,t+10} - N_{h,t+5}}{\frac{1}{2}\left(N_{h,t+10} + N_{h,t+5}\right)} = \beta_0 + \beta_1 a_{h,t} + \gamma_{M,t} + \varepsilon_{h,t} \tag{D.1}$$

Here, "size" (N) is defined as the number of patients in hospitals or the amount of physical output for concrete plants.[3]

[2] In the concrete data, exit is directly observed; in the hospital data we infer "exit" based on the hospital having less than 5 patients for five consecutive years. Therefore, for concrete we regress an indicator for whether the firm has exited at year $t+5$ on productivity in year t (and market-year fixed effects). For hospitals, we regress an indicator for whether the hospital has less than five patients in every year from year $t+5$ to year $t+9$ on productivity in year t (and market-year fixed effects).

[3] In order to make the growth analysis comparable for hospitals and for concrete, this regression differs from our baseline growth regression (equation 3) in two ways. First, because the concrete data is only available every five years, it looks at growth between 5 year periods rather than 1 year periods. Second, it lags the productivity estimate on the right hand side back another time period. As in the static allocation metric, we do this because in manufacturing, our measure of size is output, which also enters the numerator of the productivity estimate; if there is mean reversion in output and we had $a_{h,t+5}$ on the right-hand side instead, this would create negative bias in the β_1 coefficient.

Appendix E

Robustness Analysis

E.1 Additional risk adjusters

For approximately one year of patients, we have access to even more detailed information on health than in the Medicare claims data. These data comes from the Cooperative Cardiovascular Project (CCP), which abstracted information from patient charts to create an extremely detailed dataset of clinically relevant characteristics, like test results and medical histories, for a nationally representative sample of Medicare AMI patients in 1994 and 1995. These data, which are described in more detail in Chandra and Staiger (2007), are considered superior to administrative data because of the much more specific and reliable information available on patient charts than in claims data. In Table A8 we re-run our analyses on this subset of the data and show that the results are not sensitive to adding this additional, more extensive, set of controls.

Column (7) shows the results for the CCP sample with the all the information abstracted from the patient chart. These results are very similar to results from the CCP data that use fewer risk adjustors (columns 8 and 9). Results with fewer risk adjustors in the CCP data (columns 8 and 9) look roughly similar to results in one year (1994) of Medicare claims data with the same risk adjustors (column 5 and 6), which are also roughly similar to the results on our full set of Medicare claims data (columns 1-3).

E.2 Alternative Input Measures

Appendix Table A9 explores the robustness of our results to alternative input measures; more detail on their construction is provided in Appendix B. Column 1 replicates our baseline results. As noted in Section 2.6, there is a tradeoff between our relatively coarse baseline measure of inputs (with its associated measurement error) and more granular measures which suffer from potential survivorship bias (a patient cannot have a lot of procedures done if he does not survive very long). Columns 2 and 3 explore the sensitivity of our estimates to more granular measures which use as inputs a series of approximately 60 indicators for whether various procedures were performed as well as a continuous variable measuring the log of the number of days in the hospital during our 30 day window.

We incorporate this more granular input measure in two different ways. In column 2 we explore a multi-input production function; specifically, we replace our single index measure with all of the procedure indicators as well as the log hospital days variable. In column 3 we return to a single-input production function but one that is based on this more granular input measure; we create the single input by regressing log hospital charges on these same procedure indicators and the log hospital days variable, as well as hospital-year fixed effects.[1] We use the coefficients from this regression - ignoring the hospital-year effects - to produce an estimate of predicted charges for each patient in our data. The correlation between this predicted log charges measure and our baseline log input measure is 0.77 (with actual log charges it would be 0.75). As would be expected from survivorship bias, the returns to scale coefficient μ in column 3 is substantially higher than that in our baseline column 1.

Yet another alternative approach to inputs is to measure Medicare reimbursement to the hospital for a patient, rather than the hospital's use of inputs per se. Like our baseline approach, this approach is also often used in the literature (e.g. Cutler et al., 1998; Skinner and Staiger, 2009a). Medicare reimbursement depends not just on the patient's DRGs (our baseline resource measure) but also characteristics of the hospital (such as whether it is a teaching hospital or whether it treats a disproportionate share of low income patients) and its location (MedPAC, 2011). Part A Medicare spending per AMI patient is the standard measure used in the economics literature in studying the relationship between heart attack treatment and outcomes (e.g. Cutler et al., 1998; Skinner and Staiger, 2009a). The results in column 4 use this Medicare reimbursement measure; the returns to scale parameter μ is therefore interpreted here as the return to federal expenditures (in the form of post-AMI survival) rather than real inputs. The correlation between our baseline resources measure and the reimbursement measure is 0.90. The main results are all quite robust to this alternative measure.

A final input measure incorporates physician and outpatient inputs for the subsample of hospital years beginning in 2001 (see Appendix B for more details; our sample starts in 2001 because it is the first full year with data). Column 5 shows our baseline results limited to the sample where we can observe these other input measures; this cuts our sample of hospital-years substantially, by about 70 percent. Column 6 shows the results for this same "overlap" sample with our expanded input measure. For the overlap sample, the correlation between our baseline input measure and the expanded measure is 0.98.[2]

[1] Hospital "charges" are accounting charges for rooms and procedures and do not reflect transacted prices. They have been used in the literature as a convenient, price-weighted summary of treatment, albeit at somewhat artificial prices (Card et al., 2009; Finkelstein et al., 2012). The hospital-year fixed effects in the log charges regression eliminate variation across hospital-years in the charge-to-cost ratio (i.e. differential hospital markups of list prices above costs).

[2] This high correlation reflects the fact that outpatient resources are, on average, about one-fifth the size of the inpatient resources devoted to one of our patients; in addition there is a high (about two-thirds) correlation between outpatient and inpatient resources devoted to a patient.

Looking across the columns, the basic qualitative findings concerning the role for competition in allocating more market demand to more productive firms both at a point in time and over time are quite robust to alternative input measures. In particular, the static allocation analysis and the growth analysis remain statistically significant in virtually all alternative specifications. The statistical significance of the exit-based regression results is more sensitive to the choice of input measure. Perhaps not surprisingly, the magnitudes of the static and dynamic allocation analyses vary somewhat across the specifications. The dispersion estimates are remarkably robust to alternative input measures.

E.3 Alternative Time Frames for Measuring Inputs and Outputs

Appendix Table A10 considers how our metrics are affected by alternative time windows for measuring survival and inputs. Our baseline specification looks at survival over 1 year and at inputs over 30 days. A shorter time horizon for inputs will miss some of the resources provided to the patient. There is also a practical limitation to very short horizons; we observe resources at the level of a hospital stay, not a hospital day or hour; 96% of hospital stays are at most 30 days long, but a measure like 7 day utilization would require arbitrary spreading of resources across the 7 days for the 33% of patients who spend more than 7 days in the hospital. Longer time horizons have their own limitations: issues of survival bias (the longer the patient lives, the more that can be done) and the fact that as time passes since the first incident, the treatments that are undertaken are increasingly linked to providers outside the original hospital. Columns 2 and 3 show, respectively, that the results are robust to a longer (one year) survival horizon and a shorter (7 day) survival horizon, rather than our baseline 30 day time frame.

In terms of the time horizon for outcomes, we choose a 1-year survival window because it is of more interest than short-term survival, which may reflect only a few days postponement of mortality. As a practical matter, censoring is also less prevalent at 1 year than at shorter horizons. Finally, another advantage of our 1-year window is that it will pick up aspects of hospital productivity that affect outcomes through longer-term mechanisms such as the management of complications due to co-morbidities like congestive heart failure or diabetes. Longer time windows will also better capture the quality of continuing care like the prescribing of statins and the follow up to make sure the patient is taking these medications. Such inputs are less likely to affect survival at much shorter horizons but can be quite important over longer intervals. On the other hand, the longer measurement horizon introduces greater scope for patient autonomy (e.g. in terms of changes in behavior such as diet and smoking, compliance with recommended medications, follow-up visits, etc.) and for the impact of doctors (regardless of which hospital the patient went to) or admissions to other hospitals to affect survival. Longer horizons may therefore attenuate differences across hospitals in measured productivity. Our results are robust to moving away from our baseline 1 year survival to 30 day survival (column 4) or to 5 year survival (column 6); the 5 year horizon requires that we limit the sample to heart attacks through 2003 so that we observe the patient for 5 subsequent years; column 5 shows our baseline 1 year survival measure on this sample.

E.4 Alternative Market Definition

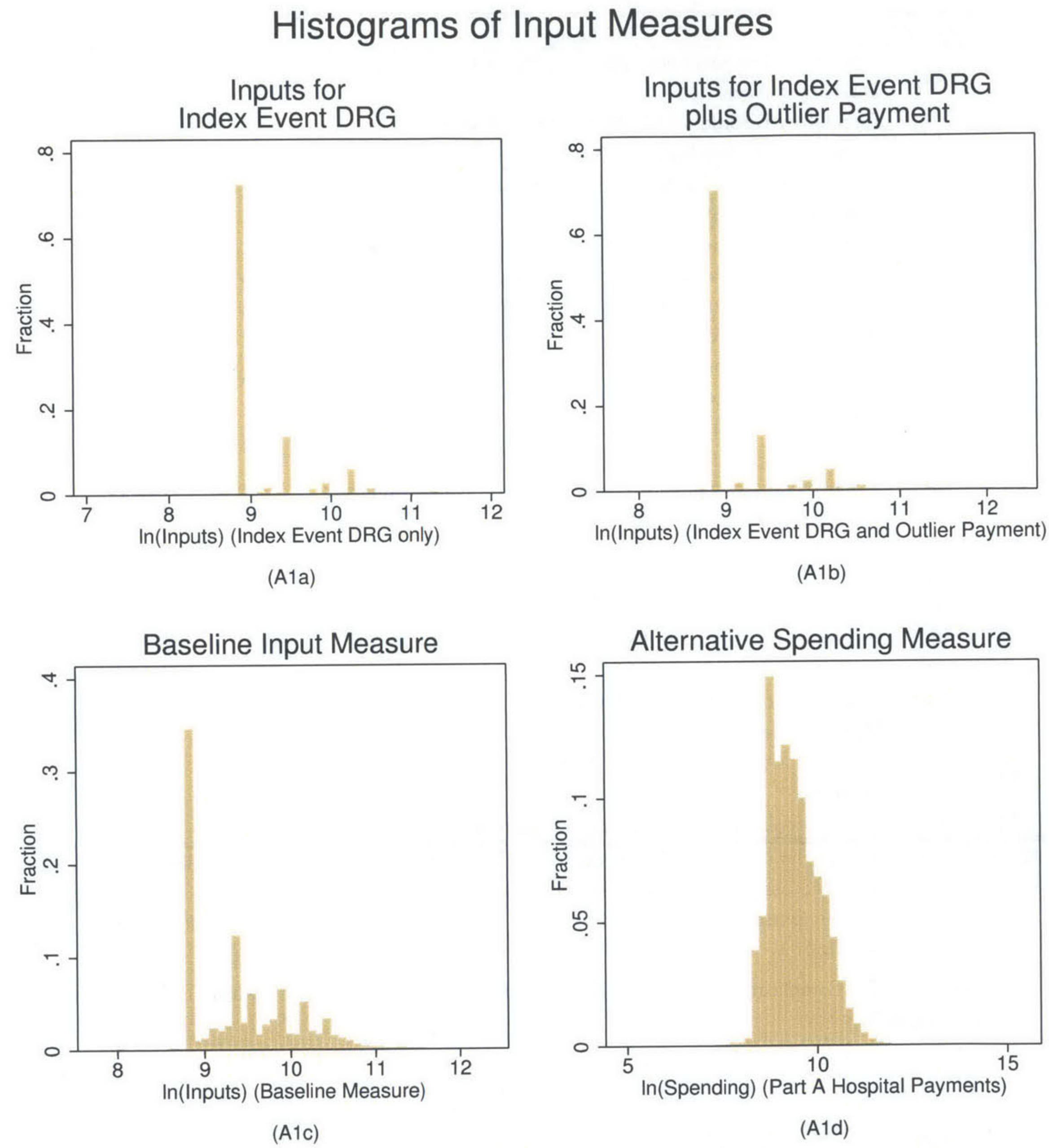
Our analysis looks at within-market static allocation and dynamic re-allocation. The baseline results use a Hospital Referral Region (HRR) as the market definition. An alternative definition of the hospital market which is sometimes used is a Hospital Service Area (HSA). HSAs are partitions of HRRs; there are about 10 times as many HSAs as HRRs.[3] Table A4 shows that our core static and dynamic allocation results are robust - indeed, they become slightly larger in magnitude - when using this alternative market definition.

E.5 Imposing scale parameter μ

We evaluated the robustness of our main results to imposing, rather than estimating, various values for the scale parameter . This method amounts to following the index number, or Solow residual, approach to measuring productivity in which factor elasticities are taken from auxiliary data such as factor cost shares. We impose a μ of 0.1, 0.3, and 0.9. These results are shown in Table A11.

[3] For more information see http://www.dartmouthatlas.org

Appendix Figures



Top row shows the component of our baseline input measure that is attributable to the patient's "index event", or initial hospitalization for the AMI. Bottom row shows the distribution of the baseline input measure and compares it to an alternative measure that captures actual payments to hospitals. Specifically, Figure A1a shows the component of the baseline measure due to the patient's index event DRG weight. Figure A1b adds index event outlier payments. Figure A1c, our baseline input measure, adds inputs (due to DRGs and outlier payments) from subsequent hospital stays within 30 days of the index event. Figure A1d shows the Part A (hospital-based) spending measure, an alternative input measure which incorporates the same hospital stays as the baseline measure but adds in geographic and hospital-specific price adjustments to capture actual Medicare payments to providers. See Appendix B for more details. All measures are in logarithms and are for the year 2000 only.

Figure A1

Appendix Tables

Table A1 - List of Top DRGs for Index Events (Initial Hospital Stays for the AMI Episode) in 2000

Rank	Number	DRG Name[a]	Weight	Share	Cum. Share
1	121	Circulatory Disorders with AMI and Major Complications, Discharged Alive	1.63	41.2%	41.2%
2	122	Circulatory Disorders with AMI, without Major Complications, Discharged Alive	1.11	20.9%	62.1%
3	116	Other Permanent Cardiac Pacemaker Implant or PTCA with Coronary Artery Stent Implant	2.47	13.0%	75.1%
4	123	Circulatory Disorders with AMI, Expired	1.51	10.9%	86.0%
5	107	Coronary Bypass with Cardiac Catheterization	5.46	5.4%	91.4%
6	110	Major Cardiovascular Procedures with CC	4.16	2.0%	93.4%
7	112	Percutaneous Cardiovascular Procedures	1.92	1.6%	95.0%
8	115	Permanent Cardiac Pacemaker Implant with AMI, Heart Failure or Shock, or AICD Lead or Generator Procedure	3.47	1.0%	96.0%
9	104	Cardiac Valve and Other Major Cardiothoracic Procedure with Cardiac Catheterization	7.24	0.8%	96.8%
10	483	Tracheostomy except for Face, Mouth, and Neck Diagnoses	16.12	0.5%	97.3%
11	106	Coronary Bypass with PTCA	7.33	0.4%	97.7%
12	109	Coronary Bypass without PTCA or Cardiac Catheterization	4.04	0.4%	98.1%
13	144	Other Circulatory System Diagnoses with CC	1.15	0.3%	98.4%
14	478	Other Vascular Procedures with CC	2.35	0.3%	98.7%
15	468	Extensive OR Procedure Unrelated to Principal Diagnosis	3.64	0.3%	99.0%
16	120	Other Circulatory System OR Procedures	2.01	0.2%	99.2%
17	108	Other Cardiothoracic Procedures	5.77	0.2%	99.4%
18	111	Major Cardiovascular Procedures without CC	2.23	0.1%	99.5%
19	477	Non-Extensive OR Procedure Unrelated to Principal Diagnosis	1.77	0.1%	99.6%
20	145	Other Circulatory System Diagnoses without CC	0.65	0.1%	99.7%

Notes: "Rank" refers to the share of patients with the DRG; "Number" refers to CMS's assigned number for that DRG; "Weight" is a CMS-assigned value that is designed to be proportional to the average cost of treatment and is used to determine reimbursement - the weights are set by CMS so that the average Medicare patient across all conditions has a weight of 1.

[a]Abbreviations: CC - Complicating Conditions, OR - Operating Room, PTCA - Percutaneous Transluminal Coronary Angioplasty.

Table A2 - List of Top DRGs for All Claims

Rank	Number	DRG Name[a]	Weight	Share	Cum. Share
1	121	Circulatory Disorders with AMI and Major Complications, Discharged Alive	1.63	15.1%	15.1%
2	127	Heart Failure and Shock	1.01	8.4%	23.5%
3	116	Other Permanent Cardiac Pacemaker Implant or PTCA with Coronary Artery Stent Implant	2.47	8.0%	31.5%
4	122	Circulatory Disorders with AMI, without Major Complications, Discharged Alive	1.11	7.3%	38.8%
5	123	Circulatory Disorders with AMI, Expired	1.51	3.8%	42.6%
6	132	Atherosclerosis with CC	0.67	2.8%	45.4%
7	107	Coronary Bypass with Cardiac Catheterization	5.46	2.7%	48.1%
8	462	Rehabilitation	1.36	2.7%	50.8%
9	89	Simple Pneumonia and Pleurisy, Age > 17, with CC	1.09	2.5%	53.3%
10	14	Specific Cerebrovascular Disorders Except TIA	1.19	1.9%	55.2%
11	88	Chronic Obstructive Pulmonary Disease	0.94	1.8%	57.0%
12	144	Other Circulatory System Diagnoses with CC	1.15	1.5%	58.5%
13	174	Gastrointestinal Hemorrhage with CC	1.00	1.2%	59.7%
14	112	Percutaneous Cardiovascular Procedures	1.92	1.2%	60.9%
15	124	Circulatory Disorders Except AMI, with Cardiac Cath and Complex Diagnosis	1.40	1.2%	62.1%
16	138	Cardiac Arrhythmia and Conduction Disorders with CC	0.82	1.2%	63.3%
17	143	Chest Pain	0.53	1.2%	64.5%
18	296	Nutritional and Miscelaneous Metabolic Disorders, Age > 17, with CC	0.86	1.2%	65.7%
19	109	Coronary Bypass without PTCA or Cardiac Catheterization	4.04	1.1%	66.8%
20	182	Esophagitis, Gastroenteritis, and Miscelaneous Digestive Disorders, Age > 17, with CC	0.78	1.1%	67.9%

Notes: "Rank" refers to the share of patients with the DRG; "Number" refers to CMS's assigned number for that DRG; "Weight" is a CMS-assigned value that is designed to be proportional to the average cost of treatment and is used to determine reimbursement - the weights are set by CMS so that the average Medicare patient across all conditions has a weight of 1.

[a]Abbreviations: CC - Complicating Conditions, OR - Operating Room, PTCA - Percutaneous Transluminal Coronary Angioplasty, TIA - Transient Ischemic Attack.

Table A3 - Sensitivity of Results to EB Adjustment

	(1)	(2)
EB Adjustment:	Yes	No
Parameter		
μ	0.446	0.446
	(0.00511)	(0.00511)
Static Allocation	2.418	0.440
	(0.0889)	(0.0182)
Dynamic Allocation		
Exit Regression	-0.0329	-0.0138
	(0.00935)	(0.00347)
Growth Regression	0.133	0.0373
	(0.0225)	(0.00759)
Dispersion		
90:10	0.442	0.751
	(0.0112)	(0.0136)
75:25	0.233	0.395
	(0.00590)	(0.00714)
Standard Deviation	0.173	0.293
	(0.00438)	(0.00530)

Notes: Column (1) is baseline specification. Column (2) shows results without the empirical Bayes adjustment. Standard errors are bootstrapped with 300 replications and are clustered at the market level.

Table A4 - Tests of Robustness of Allocation Results

Risk Adjustment:	(1) Baseline	(2) Smaller Market	(3) Nearest Hospital
Static Allocation	2.418	2.816	0.449
	(0.0889)	(0.152)	(0.0685)
Dynamic Allocation			
Exit Regression	-0.0329	-0.0675	0.00407
	(0.00935)	(0.0189)	(0.00889)
Growth Regression	0.133	0.161	0.0219
	(0.0225)	(0.0446)	(0.0220)

Notes: The allocation results are produced by estimating the specifications given in the notes to Table 2.4. Column (1) repeats the baseline full risk adjustment results. Column (2) reports the results from running the same specification with the market defined as an HSA (Hospital Service Area; HSAs partition the baseline set of markets into approximately 10 times as many markets). Since the coefficients are identified by market-years with multiple hospitals, this reduces the effective number of observations by about half. Column (3) reports the baseline results but counterfactually calculates hospital size, growth, and exit by assigning all patients to the nearest hospital in their market, rather than the hospital at which they were actually treated. Standard errors are bootstrapped with 300 replications and are clustered at the market level.

Table A5 - Allocation Metrics By Hospital- and Market-Level Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	ln(Pats)	Growth	ln(Pats)	Growth	ln(Pats)	Growth	ln(Pats)	Growth	ln(Pats)	Growth
ln(Productivity)	2.418	0.133	2.371	0.143	2.280	0.174	3.644	0.171	3.104	0.202
	(0.0889)	(0.0225)	(0.0931)	(0.0236)	(0.119)	(0.0286)	(0.363)	(0.0883)	(0.357)	(0.0627)
× Government-Run Hospital			-0.341	-0.0802						
			(0.143)	(0.0369)						
× ln(Min Distance to Nearest Hospital)					-0.150	-0.0392				
					(0.0537)	(0.0118)				
× Share College+ in Market							-5.372	-0.166		
							(1.467)	(0.357)		
× ln(Pop/KM^2) in Market									-0.157	-0.0159
									(0.0763)	(0.0127)
Government-Run Hospital			-0.511	-0.0358						
			(0.0398)	(0.00816)						
ln(Min Distance to Nearest Hospital)					-0.273	-0.0186				
					(0.0155)	(0.00258)				
Observations	55,540	52,777	55,540	52,777	55,540	52,777	55,540	52,777	55,540	52,777

Notes: Columns (1) and (2) replicate our baseline static and dynamic allocation results from Table 2.4, column 1. Column (1) shows the static allocation relationship between a hosital-year's log(patients) and productivity within a market-year (see equation 1). Column (2) shows the dynamic allocation relationship (within a market-year) between a hospital's one year percent growth and its base year productivity (see equation 3). In the remaining columns these analyses are augmented to include the specified interactions with market- and hospital-level variables (as well as the main effect of these variables as indicated). Standard errors are bootstrapped with 300 replications and are clustered at the market level.

Government-Run is defined using the hospital control field in the CMS Provider of Services file. Min Distance is the distance between the hospital and the nearest hospital to it that treated an AMI patient in that year. Share College+ is defined as the share of the population in the hospital's market that had at least a bachelor's degree in the 2000 Census. Pop/KM^2 is the population density in the hospital's market according to the 2000 Census.

Table A6 - Productivity Dispersion across hospitals.

	(1)	(2)	(3)
Risk Adjustment:	All	Age/Race/Sex	None
90-10	0.442	0.469	0.521
	(0.0112)	(0.0117)	(0.0126)
75-25	0.233	0.247	0.274
	(0.00590)	(0.00614)	(0.00666)
Standard Deviation	0.173	0.183	0.203
	(0.00438)	(0.00455)	(0.00493)

Notes: Productivity is estimated based on the corresponding specification in Table 2.2. Dispersion measures in productivity are constructed nationally each year, and then averaged across years. The top row reports difference in productivity between the 90th percentile hospital and the 10th percentile hospital; the next row reports the difference in productivity between the 75th percentile and the 25th percentile hospital; the bottom row reports the estimated standard deviation of the productivity distribution. Standard errors are bootstrapped with 300 replications and are clustered at the market level.

Table A7 - Allocation Metrics: Concrete vs Hospitals

	Concrete			Hospitals		
Risk Adjustment:	Estimate	DV Mean	Sample (Approx)	Estimate	DV Mean	Sample
Static Allocation	0.299		5,500 plant-years	2.166	3.585	33,155 hospital-years
	(0.076)			(0.094)		
Dynamic Allocation						
Exit Regression	-0.066	0.20	12,400 plant-years	-0.147	0.17	25,359 hospital-years
	(0.018)			(0.028)		
Growth Regression	0.080	-0.075	2,600 plant-years	0.480	-0.62	18,569 hospital-years
	(0.069)			(0.069)		

Notes: Estimates for concrete are based on data from the quinquennial Census of Manufactures from 1972-1992. Estimates for hospitals are based on Medicare AMI patients from 1993-2007 and use our baseline specification (see Table 2.2, column 1). Standard errors are robust analytic (Concrete) or bootstrapped with 300 replications and clustered at the market level (Hospitals). See text for further details on metrics and data (described in more detail in Appendix D).

Table A8 - CCP

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dataset	Medicare Claims 1993-2007			Medicare Claims 1994			CCP 1994-1995		
Risk Adjustment	Baseline	Age/Race/Sex	None	Baseline	Age/Race/Sex	None	Entire Chart	Age/Race/Sex	None
Parameter									
μ	0.446	0.481	0.589	0.456	0.482	0.600	0.282	0.412	0.530
	(0.00511)	(0.00523)	(0.00552)	(0.00857)	(0.00869)	(0.00875)	(0.0074)	(0.0087)	(0.0091)
Static Allocation	2.418	2.496	2.618	2.560	2.540	2.332	1.942	2.104	2.118
	(0.0889)	(0.0851)	(0.0779)	(0.268)	(0.249)	(0.181)	(0.3362)	(0.3284)	(0.2910)
Dispersion									
90:10	0.442	0.469	0.521	0.447	0.465	0.523	0.366	0.401	0.424
	(0.0112)	(0.0117)	(0.0126)	(0.0221)	(0.0219)	(0.0221)	(0.0247)	(0.0267)	(0.0266)
75:25	0.233	0.247	0.274	0.235	0.245	0.275	0.193	0.211	0.223
	(0.00590)	(0.00614)	(0.00666)	(0.0116)	(0.0115)	(0.0116)	(0.0130)	(0.0141)	(0.0140)
Standard Deviation	0.173	0.183	0.203	0.174	0.181	0.204	0.143	0.156	0.166
	(0.00438)	(0.00455)	(0.00493)	(0.00861)	(0.00853)	(0.00861)	(0.0096)	(0.0104)	(0.0104)
Patients		3,530,401			244,070			136,434	
Hospitals		5,346			4,349			3,829	
Hospital-Years		55,540			4,349			3,829	

Notes: Columns 1-3 reproduce our main results from Tables 2.2, 2.4 and 2.6. Columns 4-6 perform the same analysis on a single year of our data (1994), and columns 7-9 show the analysis on the 1994-1995 CCP sample. The CCP sample is smaller than the year of Medicare claims because it only collected data for each region of the country for 8 months and because it excluded patients whose charts had been incorrectly coded as showing evidence of AMI. The CCP results using age/race/sex adjustment (column 8) look similar to our results for one year of data using age/race/sex adjustment (column 5). (We are unable to replicate our baseline set of covariates in the CCP data due to some differences in variable availability). In the CCP, we find that relative to age, race, and sex risk adjustment (column 8), using all information that was abstracted from the patient chart (column 7) slightly weakens the static allocation relationship and slightly reduces dispersion. Standard errors are bootstrapped with 300 replications and are clustered at the market level

Table A9 - Comparison of Input Measures

	(1)	(2)	(3)	(4)	(5)	(6)
Input Measure:	Baseline	Procedures	Fitted Chg	Spending	Baseline	Base+Part B
Sample:	Full	Full	Full	Full	With Part B Data	
Parameter						
μ	0.446		0.714	0.395	0.369	0.399
	(0.00511)		(0.00652)	(0.00508)	(0.00699)	(0.00715)
Static Allocation	2.418	1.497	0.972	1.749	2.326	2.232
	(0.0889)	(0.0879)	(0.0996)	(0.0834)	(0.233)	(0.232)
Dynamic Allocation						
Exit Regression	-0.0329	-0.0199	-0.00661	-0.0245	-0.0330	-0.0347
	(0.00935)	(0.0106)	(0.0106)	(0.00943)	(0.0450)	(0.0476)
Growth Regression	0.133	0.0611	-0.00515	0.0762	0.220	0.211
	(0.0225)	(0.0258)	(0.0263)	(0.0230)	(0.0762)	(0.0798)
Dispersion						
90:10	0.442	0.431	0.428	0.453	0.353	0.343
	(0.0112)	(0.00891)	(0.00908)	(0.0104)	(0.0229)	(0.0227)
75:25	0.233	0.227	0.225	0.239	0.186	0.180
	(0.00590)	(0.00469)	(0.00478)	(0.00545)	(0.0121)	(0.0120)
Standard Deviation	0.173	0.168	0.167	0.177	0.138	0.134
	(0.00438)	(0.00348)	(0.00354)	(0.00404)	(0.00895)	(0.00887)
Patients / 1000	3,530	3,530	3,530	3,525	271.3	271.3
Hospital-Years	55,540	55,540	55,540	55,529	15,039	15,039
Hospitals	5,346	5,346	5,346	5,346	3,092	3,092

Notes: Column (1) is baseline specification. All other columns use alternative input measures (described in more detail in Appendices B and E). Column 5 and 6 are limited to the sub-sample of approximately 30 percent of hospital-years for which we observe Part B physician and outpatient data for at least five AMI patients in that hospital-year; in column 6 our baseline input measure (which uses only Part A inputs) is expanded to include Part B inputs; see text for more details. Standard errors are bootstrapped with 300 replications and are clustered at the market level.

Table A10 - Comparison of Results with Varying Survival and Input Horizons

	(1)	(2)	(3)	(4)	(5)	(6)
Survival Horizon:	1 Year	1 Year	1 Year	30 Days	1 Year	5 Years
Input Window:	30 Days	1 Year	7 Days	30 Days	30 Days	30 Days
Sample Thru:	2007	2007	2007	2007	2003	2003
Parameter						
μ	0.446	0.790	0.172	0.292	0.451	0.585
	(0.00511)	(0.00504)	(0.00959)	(0.00243)	(0.00544)	(0.00791)
Static Allocation	2.418	2.694	2.421	3.992	2.347	2.047
	(0.0889)	(0.0955)	(0.0906)	(0.146)	(0.0938)	(0.0811)
Dynamic Allocation						
Exit Regression	-0.0329	-0.0317	-0.0372	-0.0660	-0.0221	-0.0201
	(0.00935)	(0.00969)	(0.00918)	(0.0173)	(0.0105)	(0.00815)
Growth Regression	0.133	0.138	0.147	0.213	0.101	0.101
	(0.0225)	(0.0230)	(0.0220)	(0.0409)	(0.0251)	(0.0189)
Dispersion						
90:10	0.442	0.422	0.450	0.224	0.446	0.583
	(0.0112)	(0.00981)	(0.0117)	(0.00626)	(0.0119)	(0.0146)
75:25	0.233	0.222	0.237	0.118	0.235	0.307
	(0.00590)	(0.00516)	(0.00617)	(0.00330)	(0.00628)	(0.00770)
Standard Deviation	0.173	0.164	0.175	0.0874	0.174	0.227
	(0.00438)	(0.00383)	(0.00457)	(0.00244)	(0.00465)	(0.00571)
Patients / 1000	3,530	3,530	3,530	3,530	2,702	2,702
Hospitals	5,346	5,346	5,346	5,346	5,180	5,180

Notes: Column (1) is baseline specification. In other columns the time horizon in which we measure survival and/or inputs is modified as indicated in the column headings. Standard errors are bootstrapped with 300 replications and are clustered at the market level.

Table A11 - Sensitivity of Results to μ

	(1)	(2)	(3)	(4)
Source of μ:	Estimated		Imposed	
Value of μ:	0.446	0.1	0.3	0.9
Static Allocation	2.418	2.358	2.399	2.278
	(0.0889)	(0.0883)	(0.0884)	(0.0835)
Dynamic Allocation				
Exit Regression	-0.0329	-0.0361	-0.0343	-0.0263
	(0.00935)	(0.00923)	(0.00930)	(0.00891)
Growth Regression	0.133	0.144	0.138	0.107
	(0.0225)	(0.0220)	(0.0223)	(0.0215)
Dispersion				
90:10	0.442	0.449	0.445	0.457
	(0.0112)	(0.0116)	(0.0114)	(0.0104)
75:25	0.233	0.237	0.234	0.241
	(0.00590)	(0.00611)	(0.00599)	(0.00549)
Standard Deviation	0.173	0.175	0.173	0.178
	(0.00438)	(0.00453)	(0.00444)	(0.00407)

Notes: Column (1) shows results based on estimation of our baseline specification (Table 2.2, column 1). In the other columns μ is imposed rather than estimated. Standard errors are bootstrapped with 300 replications and are clustered at the market level.

References (Additional for Appendices)

Card, David, Carlos Dobkin, and Nicole Maestas. 2009. "Does Medicare Save Lives?." *The Quarterly Journal of Economics*, 124(2): 597–636.

Centers for Medicare and Medicaid Services. 2011. "Hospital Acute Inpatient Services Payment System." URL: http://www.medpac.gov/documents/MedPAC_Payment_Basics_11_hospital.pdf, accessed on July 10, 2012.

Clemens, Jeffrey, and Joshua D. Gottlieb. 2014. "Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?" *The American Economic Review*, 104(4): 1320–1349.

Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Heidi Allen, and Katherine Baicker. 2012. "The Oregon Health Insurance Experiment: Evidence from the First Year." *The Quarterly Journal of Economics*, 127(3): 1057–1106.

Hopenhayn, Hugo A. 1992. "Entry, Exit, and firm Dynamics in Long Run Equilibrium." *Econometrica*, 60(5): pp. 1127–1150.

MedPAC. 2010a. "Outpatient Hospital Services Payment System." URL: http://www.medpac.gov/documents/MedPAC_Payment_Basics_10_OPD.pdf, accessed on July 10, 2012.

MedPAC. 2010b. "Physician Services Payment System." URL: http://www.medpac.gov/documents/MedPAC_Payment_basics_10_Physician.pdf, accessed on July 10, 2012.

MedPAC. 2011. "Hospital Acute Inpatient Services Payment System." URL: http://www.medpac.gov/documents/MedPAC_Payment_Basics_11_hospital.pdf, accessed on February 9, 2012.

Sloan, Frank Allen. 2000. "Not-for-profit ownership and hospital behavior." In *Handbook of Health Economics*. eds. A. J. Culyer, and J. P. Newhouse, 1: Elsevier, , 1st edition, Chap. 21 1141–1174.