Impacts of performance pay for hospitals: The Readmissions Reduction Program *

Atul Gupta†

Job Market Paper

This version: November 7, 2016

For the latest version, please go here

Abstract

Policy makers are increasingly tying payments for health care providers to their performance on quality measures, though there is little empirical evidence to guide the design of such incentives. I deploy administrative Medicare claims data to study a large federal program which penalizes hospitals with high rates of repeat hospitalizations ("readmissions"). I exploit the introduction of the penalty and policy-driven variation in penalty across hospitals to identify the effect of the program on hospital admission and treatment decisions, and on patient health. The program is associated with a 5% decrease in readmission accompanied by a 3% reduction in thirty day mortality. I quantify the role of two mechanisms - improvement in treatment quality and changes in admitting behavior - and find that quality improvement can explain 55-60% of the aggregate decrease. The change in admitting behavior seems driven by the penalty since there is a substantial decrease in admission rate for returning patients that could potentially incur a penalty but no such effect for those that will not. It plays a quantitatively important role and I find suggestive evidence of harm to affected patients.

*I would like to thank Nicholas Bloom, Mark Duggan, Liran Einav and Matt Gentzkow for their mentorship and advice. This paper has also benefited from invaluable comments and suggestions by Jay Bhattacharya, Tim Bresnahan, David Chan, Raj Chetty, Josh Gottlieb, Caroline Hoxby, Petra Persson, Tomas Phillipson, Pooja Garg, Evan Mast and several seminar participants at Stanford, Berkeley and University of Chicago. I am grateful to Betty Henderson-Sparks and Jon Teague of the California Office of Statewide Planning & Health Development for their assistance in providing the Hospital/ER discharge data. I gratefully acknowledge support by the Stanford Institute for Economic Policy Research through the Ely Graduate Fellowship. All remaining errors are my own.

†Department of Economics, Stanford University. 579 Serra Mall, Stanford CA. Email: atulg@stanford.edu.

1 Introduction

Beginning with the Affordable Care Act (ACA), tying payments to quality of care has become a centerpiece of US health care policy.[1] Performance pay contracts help payers (insurers or government) to focus provider effort on improving quality of care. However, there could be unintended effects, most notably the multi-tasking concern (Holmstrom and Milgrom, 1991) whereby providers may divert resources away from non-targeted tasks to focus on the 'rewarded' task, or distortion of treatment decisions resulting in negative effects for patients. Empirical evidence on the effects of such incentives in health care remains limited (Mullen et al., 2010).

This paper provides empirical evidence on the effects of performance pay by exploiting plausibly exogenous variation in incentives created by a national quality incentive scheme for hospitals. I exploit the introduction of the "Hospital Readmissions Reduction Program" (HRRP) — a component of the ACA — which provides differential financial incentives to hospitals to decrease their readmission rates for Medicare patients. A readmission is a repeat hospitalization that begins within thirty days of discharge from an earlier hospital stay. Although the penalty is computed based on a limited set of conditions, it is applied to all the hospital's Medicare revenue, not just that from the penalized conditions. In 2014–15, CMS expected to recover $ 430 million in penalties from hospitals.[2]

To set the penalty applicable to a particular year, the Center for Medicare and Medicaid Services (CMS) calculates readmission rates for all hospitals and penalizes hospitals whose readmission rate is greater than the national average. Other hospitals receive no penalty or reward. The program is structured such that a hospital incurs a (nearly) constant penalty per readmission, provided it exceeds the threshold value. Every 1% increase in the hospital's readmission rate increases its penalty by (approximately) 1% of the revenue received for the penalized condition. Conceptually, Medicare is clawing back revenue paid to the hospital in proportion to its poor performance. Readmissions can occur at any hospital and still count toward the penalty, but the initial hospital is always held accountable.

Hospitals could respond to the penalty in two ways. First — as policymakers intended — hospitals could improve quality of care. This could be implemented by devoting more resources (for example, hiring more case managers to follow-up patients post-discharge) or by improving produc-

[1] For example, in Jan 2015 the federal government declared a target to tie 85% of Medicare hospital payments to quality of care by the end of 2016. Press release available at http://tinyurl.com/qfmzg5j.

[2] See http://khn.org/news/medicare-readmissions-penalties-2015.

tivity i.e. producing better outcomes with the same resources (for example, improving compliance with best clinical practices). Second, hospitals can also change the composition of patients they admit. Patients with the penalized conditions typically arrive at the hospital Emergency Department (ED) and hospitals have considerable discretion on whether to admit them or treat them in the ED as outpatients. The penalty could affect hospital admission decisions for marginal patients, particularly those that had their last stay at the same hospital within thirty days and will incur the penalty if admitted or those making an initial visit but seem like they are likely to need readmission soon. I first establish that the penalty is associated with a decrease in overall readmissions and then exploit the institutional setting to disentangle contributions of the two mechanisms.

The main source of data is administrative claims records for the universe of Medicare fee-for-service patients. I deploy data for the period 2006–14, which spans approximately five years before and three years after the penalty was announced. The data provides a rich level of detail on each health care interaction for each Medicare beneficiary in addition to patient demographics. The unit of analysis is the initial admission that is subject to the penalty. I follow patients after they discharge from the initial case and construct various measures of health care utilization including readmission.

The research design exploits two sources of policy-driven variation. First, the penalty affects hospitals differentially and creates cross-sectional variation. Second, introduction of the penalty introduces temporal variation within-hospital. Hospitals are not randomly assigned to the penalty and the panel is crucial in eliminating constant unobserved differences across hospitals. Hence this setting lends itself to a differences-in-differences research design. An important feature of the program is that a hospital's present readmission rate determines its penalty in the future. Since the program incentivizes improvement only to avoid the penalty, the hospital's response is driven by its expectation of exceeding the threshold rate. This implies that treatment status is fuzzy in this setting and does not simply equal penalty status.

Hospitals with a recent history of very low readmission rates have virtually no chance of exceeding the threshold and hence perceive no penalty per readmission. At the opposite end, hospitals with very high recent readmission rates are almost assured of exceeding the threshold and perceive the full penalty per readmission. Hospitals close to the threshold are uncertain but have greater incentive than the first group of hospitals. This highlights the intensive margin of the variation in incentive — hospitals with the worst prior performance perceive greater penalty per readmission and have more at stake than hospitals just greater than average.

To capture both extensive and intensive margins of the penalty incentive, I construct a measure of hospitals' expectation of exceeding the threshold value in each year based on their observed readmission rate in the previous years. OLS using this constructed measure is potentially biased, with mean reversion being a particular concern. I circumvent this problem by using an instrumental variable approach, which also mitigates concerns due to measurement error in constructing hospital beliefs. The approach uses predetermined hospital characteristics from a much earlier period (subsequently omitted) to instrument for the hospital's belief on exceeding the penalty threshold. I use two alternate instruments to implement this strategy and find similar results.

The baseline instrument is the hospital's expected readmission rate in 2006–07 (the first year of my data), predicted based on patient risk factors. The second instrument is the hospital's readmission rate predicted only by patient demographics (race and income). Patient demographics are relatively stable over time and strongly predict readmission but are not included in CMS' risk adjustment algorithm, an anomaly pointed out recently by Barnett et al. (2015) who conclude that "hospitals with high readmission rates may be penalized to a large extent based on the patients they serve". Hence hospitals with a greater share of minority patients have high ex-ante penalty potential and a differentially higher incentive to respond to the penalty. In both cases, the key identifying assumption is that in absence of the penalty, hospitals with high vs. low expected readmission rates in 2006–07 would evolve along parallel trends. To explore the validity of this assumption I plot fully non-parametric estimated effects each year on all key outcomes.

The baseline IV estimate indicates that moving a hospital from the 25th to the 75th percentile likelihood of being in the penalty range is associated with a decrease of 1.9 percentage point (9%) in its readmission rate over 2012–14. This is an economically large and statistically significant effect. To place this in context, it implies nearly 70,000 avoided hospitalizations for Medicare patients each year, saving approximately \$620 million.[3]

Several key facts point to the causal interpretation of these results. First, there are no differential pre-trends across hospitals with different level of penalty incentive. Second, the timing of the decrease coincides with the introduction of the penalty. Third, I find small and statistically insignificant effects on the readmission rate over 31–60 days, which was not penalized. Although not a strict placebo test, it reassures us that the estimated effect on thirty day readmissions is not driven by macro

[3] In 2011 there were approximately 860,000 initial Medicare cases for the penalized conditions. An average decrease of 1% implies 8.6k fewer readmissions in these three conditions. Extrapolating to all Medicare patients and accounting for the lower baseline readmission rate among other conditions implies a decrease of 69,000 readmissions. The average readmission cost Medicare approximately \$8,000 in 2007, equivalent to \$9,200 in 2016\$.

trends or other reforms in health care.

Applying the same research design, I quantify the role of the two mechanisms, beginning with improvements in treatment quality. I examine a number of metrics, but the key measure is short term mortality. I find a 0.4 percentage point decrease in thirty day mortality. To interpret the magnitude, I use estimates provided by Doyle et al. (2015) on the marginal cost of improving mortality and find that Medicare would have to spend approximately 2.5% ($280) more per patient to achieve equivalent gains in mortality.[4] Assuming that patients do not change how sick they must feel before deciding to return to a hospital to seek care, improvement in treatment quality should translate into a decrease in probability of patients returning to a hospital. Indeed, I find a decrease in probability of patient return which can account for 55–60% of the estimated decrease in readmissions.

I then examine the impact on hospitals' admission behavior as hypothesized above. I find evidence of a large decrease in the probability of hospitals readmitting their own patients when they return to the ED within thirty days, but no corresponding effect on readmission for patients returning to different hospitals. Note that the former group could potentially incur a readmission penalty whereas the latter will not. This channel is quantitatively important and accounts for the remaining decrease in overall readmission. Further, I find patients returning to the same hospital are more likely to make another ED visit within the next 15 days than patients who returned to a different hospital. This suggests there is some harm to patients, although no detectable effects on mortality. I also find hospitals are less likely to admit patients when they make a first visit, however the results indicate (i) this is a minor force and (ii) no evidence of selection on observable risk.

This paper relates to two strands of existing literature. There is a large empirical literature on performance pay. An exhaustive literature review[5] is beyond the scope of this paper, but a key gap is evidence on the health care sector, particularly on quality incentives. Rosenthal, Frank and co-authors examine a quality incentive scheme in California for physicians that targeted multiple clinical process metrics (Rosenthal et al., 2005; Mullen et al., 2010) and find little or no effects. In parallel work, Mellor et al. (2016) also study HRRP, albeit only for Virginia hospitals. They use different data sources and empirical approach, and do not quantify the role of different mechanisms. Norton et

[4] Doyle et al. (2015) find that 1 percent increase in spending reduces 1 year mortality by 0.19 percentage point (Table 3). However, their estimate is based on specific non-deferrable conditions which include Heart Attack and Pneumonia but exclude Heart Failure. Assuming that mortality gains at 90 days persist till one year, I find a 0.45 percentage point improvement in mortality for these two conditions. The average initial case across these two conditions costs Medicare approximately $12,000. Simple calculations arrive at the figure mentioned in the text.

[5] Podgursky and Springer (2007) provide a comprehensive review of the empirical research on the effects of performance pay in education (primarily K-12 schools). Lazear (2000) reviews the theoretical and empirical results on performance pay in the context of firms and labor contracts.

al. (2016) is closest to my paper and study hospital responses to another national quality incentive program using a different empirical approach. They are limited by the lack of micro-data and are unable to decompose aggregate effects into different channels.

A large literature has examined the effects of price regulation in health care. Cutler (1995), Ellis and McGuire (1996) and Acemoglu and Finkelstein (2008) study hospital or physician responses to substituting cost-based reimbursement with prospective payments and find a decrease in supply of care. Dumont et al. (2008) examine the effects of dampening the fee-for-service incentive with base salaries and find further decrease in supply of care. Duggan (2000), Dafny (2005) and Alexander (2015) study gaming responses to changes in pricing for specific patients and medical conditions. This paper attempts to go one step further by quantifying different mechanisms. Further, the setting in this paper is different since providers are hit with an uncertain price change by the regulator.

The rest of the paper proceeds as follows. Section 2 describes key features of the readmissions reduction program and penalty structure. Section 3 describes the data sources, sample construction and summary statistics. Section 4 describes the empirical challenges and motivates the research design. Section 5 presents the main results and quantifies the role of different mechanisms. Finally, section 6 concludes and discusses some implications from this analysis beyond this setting.

2 Setting

2.1 Medicare spending and reform

Medicare is a federal public insurance program, mainly providing health coverage for individuals aged 65 and above. It covers most types of health care services for beneficiaries. In 2011, 75% of beneficiaries were enrolled in Traditional Medicare (henceforth, TM) where they can freely choose their providers. A newer component called Medicare Advantage is administered through private insurers and is small but growing quickly. In 2011, the federal government spent $ 285 billion on TM, of which about 50% was to cover hospital care (MEDPAC, 2013). The level and growth in spending on Medicare and particularly on TM has been a persistent policy concern. The ACA introduced three performance pay programs for hospitals - Hospital Acquired Infection (HAC) program, Hospital Value Based Payments program (HVBP) and the Hospital Readmissions Reduction Program (HRRP), the focus of this paper. The three programs target different quality metrics — all only for

TM patients — and, if successful, will decrease spending as well.[6]

Over the last few years, health care spending has been below the projections made when the ACA was passed. In particular, Medicare spending in 2014 was about 10% (or $ 60 billion) below the figure projected in 2010 (McMorrow and Holahan, 2016). The unexpected decrease is driven by spending per beneficiary rather than due to a decrease in beneficiaries. The budget sequestration of 2011 imposed mandatory price cuts on Medicare and is one source of unexpected saving. It is likely that the performance pay initiatives discussed above have also contributed, though their effects are not well understood.

2.2 The program

The Hospital Readmissions Reduction Program imposes a penalty on hospitals based on their performance on readmissions, specifically re-hospitalizations that occur within thirty days of discharge from a previous hospital stay. Although the ACA was enacted in early 2010, the law did not specify the penalty rules and allowed CMS substantial discretion in designing the magnitude and scope of the penalty. CMS officially announced these details in the federal register of August 2011. The penalty was first applied on admissions starting in October 2012, however since it is computed based on past performance, hospitals had an incentive to react immediately. In the first two years (2012–13 and 2013–14) the penalty was based on performance in three conditions — Heart attack, Heart failure and Pneumonia. However, over time it is expected to become more widely applicable.[7]

Consider a hospital h at the end of year t. For each penalized condition, CMS calculates the risk adjusted readmission rate r_h over the three year period $(t-2, t)$. This is a two step process where CMS first computes the raw readmission rate i.e. the proportion of (say, Pneumonia) cases over this three year period that resulted in readmissions(s) within thirty days.[8] It then adjusts this value to account for patient risk factors.[9] Hospital h's performance is compared to the national mean value $\bar{r}$

[6] Blumenthal et al. (2015) review the reforms introduced by the ACA including the Medicaid expansion and discuss preliminary evidence (based on government analysis) of its impacts.

[7] CMS has already added Chronic Obstructive Pulmonary Disorder (COPD) and Hip/Knee replacement surgeries starting October 2014 and announced the inclusion of Coronary Artery Bypass Graft (CABG) surgeries starting in October 2016.

[8] CMS does not differentiate between one and multiple readmissions within a thirty day period. Hence it is more appropriate to think of the target metric as the probability of readmission.

[9] In practice CMS uses heirarchical logistic regression analysis to compute r_h by estimating case level regressions with an indicator for readmission as the dependent variable and a vector of approximately 40 patient risk factors as the independent variables. Risk factors vary slightly by condition but typically include gender, age over 65 and indicators for diagnosis of various conditions (Eg: septicemia, cancer, diabetes, kidney disease, liver disease, heart disease, physical disability, dementia and many others) within the past one year. Appendix A presents full details of the risk adjustment algorithm.

(normalized to one) across all hospitals for the particular condition. Hospital h is penalized in year $t+2$ for Pneumonia if $r_h > \bar{r} = 1$, else there is no penalty or bonus payment. CMS considers $(r_h - 1)$ as the hospital's excess readmission rate and this drives variation in the penalty rate across hospitals.

The penalty rate also depends on other factors like Medicare inpatient revenue from Pneumonia (or the penalized condition) and for the entire hospital. To convert the penalty rate into dollars, CMS multiplies it by the total Medicare revenue received by the hospital in year $t+2$. In practice, the penalty is deducted per case in year $t+2$ rather than lump sum at the end of the year. The total dollar penalty for the hospital is simply the sum across conditions. The non-linearity in the formula ensures that hospitals do not receive a bonus for one condition which compensates poor performance in another condition. A detailed description of the penalty formula is available in Appendix B.

Part of the variation in penalty dollars is mechanically driven by hospital size — larger hospitals will be hit with a greater penalty than smaller hospitals. The same hospital will be hit with a bigger penalty if it has a larger cardiology department and serves more patients. To isolate the marginal incentive due to readmission performance, I normalize by the Medicare revenue received for the penalized condition. The solid line in figure 1 depicts this kinked relationship between penalty per dollar reimbursement in year $t+2$, p_h and the risk adjusted readmission rate r_h in $(t-2, t)$. The actual penalty rates for hospitals are superimposed in Figure 1 using circles and line up fairly well along the solid line. This shows that the variation in penalty is driven mainly by past performance on readmission.[10]

An important implication of this penalty structure is that the penalty per readmission is essentially constant. If $r_h > \bar{r}$, then each readmission incurs a constant penalty recovered in year $t+2$. This marginal 'cost' per readmission only changes based on whether the hospital exceeds the threshold value, $\bar{r}$, but it does not matter how close or far away the hospital is from $\bar{r}$. The penalty formula is structured such that a one percent increase in the hospital' readmission rate beyond the threshold increases the penalty by one percent of the revenue received for the penalized condition. For example, if a hospital receives \$ 2 million from Medicare for Pneumonia in a year, then a 1% increase in its readmission rate beyond the threshold will result in an increase of \$20,000 in penalty. Conceptually, Medicare is clawing back revenue paid to the hospital for the penalized condition in the same

[10]The third input factor into the formula is a scaling factor that increases (decreases) the penalty if total Medicare revenue at the hospital has increased (decreased) in $t+2$ relative to the annual average value during $t-2, t$. Empirically this value is approximately one within hospitals over time.

proportion as the hospital's poor performance.

Figure 1 also shows the intensive margin in the penalty incentive. Hospitals at the extreme right expect to pay a greater aggregate penalty relative to those just to the right of the kink. Although the marginal incentive is constant, hospitals at the extreme right have a greater aggregate value at stake.

2.3 Timing

The timing of the penalty requires attention since it has important implications for the empirical strategy. The discussion above was from the perspective of a hospital at the end of year t when it has been informed of its penalty rate in the next cycle. However, during year t when the hospital is making admission and treatment decisions for each patient it does not know the end-of-year penalty threshold $\bar{r}$ nor its own readmission rate r_h. An added source of uncertainty is that these are risk-adjusted values which cannot be observed or backed out by an individual hospital. This implies that "treatment" status is fuzzy — hospitals that are eventually not penalized could also respond because they expected a penalty ex-ante. Since the penalty per readmission remains constant, response to the program is sensitive to the hospital's belief on how likely it is to exceed $\bar{r}$, not by how much, though the two are correlated.

An additional implication is that any improvements a hospital makes in year t will bear fruit only in year $t+2$ and beyond. For example, improvements that the hospital makes in 2012 will potentially reduce its penalty burden in 2014 and beyond. Hence, forward looking hospitals will want to make improvements in 2012 even though the penalty is introduced only in 2013.

2.4 Readmission as a quality measure

It is reasonable to wonder why readmission was chosen as the quality measure for this penalty scheme. The struggle to find a good metric for quality of care is not a new one (McClellan and Staiger, 1999). Both short-term mortality and readmission have been used extensively as examples of adverse medical events as well as quality metrics (Cutler, 1995; Currie and Gruber, 1996; Duggan, 2000). Both are noisy indicators of care quality since they are affected by factors outside the hospital's control. Readmission is an appealing proxy for quality since it can be easily calculated, is a common outcome for a wide class of patients and conditions (unlike mortality) and is plausibly correlated with treatment quality in the initial episode of care. The key limitation of using readmission is that its link to patient welfare is not clear. Unlike death, a readmission is not always bad for the patient, hence it is difficult to make a comment on the impact on patients only based on a decrease in readmission.

Two factors could be responsible for the choice of readmission as the key quality metric. First, readmissions are costly for Medicare. The Medicare Payments Advisory Commission (MEDPAC) estimated in its June 2007 report that readmissions cost approximately \$15 billion out of \$105 billion total spending on hospital services (MEDPAC, 2007). In the same report they also estimate that 80% of the readmission spending was on preventable readmissions.[11] There is some debate in the health literature about the share of preventable readmissions with estimates ranging from 10% to 50% (Axon and Williams, 2011), but there is consensus that some proportion could be prevented with better quality care.

Second, several small sample randomized control trials have demonstrated that readmissions can be decreased by implementing specific low-cost quality improvement interventions. For example, decreasing the rate of hospital-acquired infections, incorporating best practice guidelines into clinical care (PHC4, 1996; Hannan et al., 2003), better drug reconciliation checks (Coleman et al., 2005), pre-discharge counseling (Naylor et al., 1999) and improving care coordination with primary care physicians (Kripalani et al., 2007).

Hence while readmission is a noisy indicator of quality, it is sensitive to specific low-cost quality improvements and offers the potential for substantial savings. MEDPAC clearly indicated these goals when they recommended a performance pay scheme to Congress to decrease readmission rates (MEDPAC, 2007).

3 Data sources and initial evidence

3.1 Sample construction

The primary data used for analysis is claim-level data on the universe of Medicare fee-for-service beneficiaries from July 2006 – June 2014. I limit the sample to acute care hospitals excluding various types of specialized hospitals which are not subject to the penalty. Appendix A provides more details on the data sources and sample selection.

The unit of analysis is an "index admission" i.e. the hospitalization to an acute care hospital associated with one of the three penalized conditions subject to the readmissions penalty. Each such

[11] MEDPAC hired 3M to identify preventable readmissions using specialized software and 2005 claims data. They estimated that approximately 75% of the readmissions and 80% of the spending was on preventable readmission.

admission is associated with the hospital at which it occurred, demographics and utilization history of the patient as well as subsequent health care utilization by the patient, including readmissions. Overall, as shown in Panel A of Table 1, the sample contains almost 7 million index admissions, of which 3.1 million associated with heart failure, 2.5 million with pneumonia, and 1.2 with heart attacks. These were identified from an initial sample of nearly 50 million hospital stays over this period across all conditions. There are 3,250 hospitals in the sample. Appendix table A.1 presents information on size, revenue and readmission rates for hospitals categorized by different types of owners.

Using the information on subsequent health care utilization, I construct two key outcomes associated with each index admission. One is whether the discharged patient had a subsequent hospitalization within 30 days, namely readmission. The second is subsequent Emergency Department (ED) visit by the patient, which may or may not result in a readmission. This allows me to disentangle patients' decision to seek care from the hospital's decision to readmit the patient. I also examine the hospital's admission decision at the index admission stage using data on ED visits which would have been classified as index cases if they had resulted in admission. Note that while the index cases are necessarily related to one of the three penalized conditions, the return or readmission cases can be due to any condition.

3.2 Descriptive statistics and pattern over time

Table 1 Panel B presents descriptive statistics on key measures of patient outcomes. Probability of admission when a patient arrives in the index case is 89%, varying from near certainty of admission in case of heart attacks to 84–88% for the other two conditions. Patients not admitted receive outpatient care in the ED.[12] Probability of readmission within thirty days is the basis for the penalty and is approximately 20% with some variation across conditions. As shown in the table, the readmission rate arises from approximately 25% of the discharged patients who return to the ED within 30 days, of whom 75–90% get (re)admitted.

These patterns are also depicted in figure 4 which presents a stylized hospital cycle for a patient having one of the penalized conditions as primary diagnosis. The numbers mentioned in the figure

[12] A limitation of claims data is that it captures the primary diagnosis coded by the hospital ex-post and billed to Medicare. Hence it is possible that several other patients arrived at the ED thinking they had Pneumonia, but tested negative and were sent home. These patients will not enter my sample since their primary diagnosis was coded as some other condition. There is the possibility of strategic relabeling by hospitals, which I discuss in section 5.5 but do not incorporate in the main analysis.

represent actual proportions observed in the data during 2009–11. For every hundred patients admitted by the hospital, 112 patients arrived at the ED. The remaining were deemed not sick enough to merit inpatient treatment and were discharged from the ED as outpatients. Since these conditions are highly acute, in-hospital mortality is substantial at 5%. The remaining 95 patients are discharged, typically after 5–6 days. The program considers this to be the set of 'index' patients i.e. it will track if these patients subsequently get readmitted within thirty days. About 25 of these patients feel sick enough to return to some hospital within thirty days, of which 19 are readmitted — resulting in a readmission rate of approximately 20% (19/95) per the program.

Table 1 Panel C presents descriptive statistics on two additional measures of quality of care. One is short-term mortality: the mean mortality rate at 90 days is quite similar across the three conditions and is approximately 20%. The second metric is 'process of care' scores released by CMS on its hospital compare website.[13] Table 1 panel C presents descriptive statistics on the raw scores across hospitals for each condition. Hospitals have high compliance with these metrics on average, with significant dispersion. I standardize the scores before using them, the details are available in Appendix A. One caveat with using these scores in this setting is that they pertain to all patients and not just Medicare patients.

Figure 2 presents the trend in readmission rate aggregated across all three conditions for all hospitals in the sample. Readmissions were essentially flat before 2011 and a sharp decline begins in 2012, even before the first penalty was levied (in 2013). This is consistent with the explanation that forward looking hospitals will respond to the penalty immediately.

4 Research Design

The introduction of the program creates cross-sectional variation in marginal penalty incentive across hospitals and within-hospital variation over time. This setting lends itself naturally to a differences-in-differences research design to quantify hospital responses to the penalty. There are three empirical challenges to identifying causal effects of the penalty that the proposed design attempts to overcome. First, hospitals are not randomly assigned to the penalty. Since penalty is assigned based on past readmission performance which is correlated with a number of hospital characteristics, pe-

[13]These are also known as "Timely and effective care" measures and the data is available for download at https://data.medicare.gov/data/hospital-compare?sort=relevance&tag=timely%20and%20effective%20care. CMS tracks hospital compliance with a set of care measures for each condition. Each measure relates to a specific clinical intervention and has been accepted as a benchmark of good practice in medicine (Williams et al., 2005; Chandra et al., 2016). For example, the measures for heart attack include the proportion of patients given Aspirin on arrival or at discharge.

nalized and non-penalized hospitals are observably different. Accordingly I rely on within-hospital estimates to difference out unobserved and observable time invariant factors.

Second, as discussed in section 2, treatment status is fuzzy in this setting since hospitals will react to their ex-ante expectation of exceeding the penalty threshold $\bar{r}$ rather than actual penalty status. Hence simply comparing penalized and non-penalized hospitals will underestimate hospital response. Admission and treatment decisions for each patient i are driven by hospital's expectation of exceeding the end-of-year cutoff, conditional on their information set at the start of year t. The linear equation below represents a static version of this economic model

$$Y_{iht} = \alpha_h + \delta_t + \beta \cdot \mathbb{E}\left[\mathbf{1}(r_{ht-2,t} > \bar{r}_{t-2,t}) | I_{t_0,t-1}\right] \cdot \mathbf{1}(t \geq 2012) + X'_{iht}\gamma + \epsilon_{iht} \tag{1}$$

where Y could be any outcome of interest, but it is intuitive to think of readmission as the outcome for this discussion. α_h represents time invariant hospital 'quality' and δ_t indicates a constant shock affecting all hospitals in year t. The third term represents the hospital's expectation of its readmission rate r_h exceeding the national mean $\bar{r}$, based on its information set I at the start of the year.[14] Since the penalty rules were released in August 2011, I assume hospitals first start responding to the penalty in 2011–12. It is possible that some hospitals started responding immediately post-ACA which was enacted in 2010, in which case this approach will yield an under-estimate. Such anticipatory responses have been documented in the context of other large payment reforms (Alpert, 2016). ϵ_{iht} represents all omitted factors that could affect readmission after controlling for patient risk factors X. β therefore quantifies the hospital's response to the penalty incentive.

The third challenge is to construct a measure of hospitals' beliefs $\mathbb{E}\left[\mathbf{1}(r_{ht-2,t} > \bar{r}_{t-2,t}) | I_{t_0,t-1}\right]$ which are unobserved. I make two simplifying assumptions to construct an empirical analog. First, I assume that hospitals base their expectation on knowledge of their observed readmission rate in years $(t-3, t-1)$ and the market average. For example, in 2012 hospitals predict their probability of being penalized in the next cycle using their realized readmission rate in 2009–11. Hospitals are aware of their past performance on readmission in relation to the market since CMS has been releasing raw readmission rates on its Hospital Compare website for these conditions at least since 2009. Second, I assume that hospitals are 'right' on average i.e. their expectations match the realized penalty assignment in the future.

[14] This expectation term is applied to the marginal penalty per readmission, $\bar{R}_h$ to arrive at the cost per readmission in dollars. Due to the penalty formula, $\bar{R}_h$ is the mean reimbursement per index case for the particular condition and there is little variation across hospitals in this value.

$$\mathbf{1}(\text{Penalty}_{ht+1} = 1) = f(\bar{Y}_{ht-3,t-1}) + \xi_{ht}$$
$$P_{ht+1} = \hat{f}(\bar{Y}_{ht-3,t-1})$$

Accordingly I construct the measure P_{ht+1} as a non-parametric local linear fit of penalty status in year $t+1$ on raw readmission rate, $\bar{Y}$ in years $(t-3, t-1)$. Figure 3 plots P_{ht+1} in case of Pneumonia for all hospitals when $t = 2012$ on the Y-axis, against the raw readmission rate $\bar{Y}$ over 2009–11 on the X-axis. The probability of being penalized increases non-linearly with a hospital's readmission rate, characterized by a sharp increase around the overall mean value. The figure also showcases how the sharp discontinuity in penalty incentive discussed in section 2 is smoothed out. Table 1 Panel D presents the inter-quartile range of the ex-ante likelihood of penalty in the first year of the program (approximately 0.9).

This continuously varying measure of hospital belief encapsulates both the extensive and intensive margins of the penalty incentive. Hospitals at the extreme right in figure 3 are certain they will be penalized and have the most incentive to respond. Conversely, hospitals around the mean readmission rate are quite uncertain and their expected loss per readmission is accordingly lower. Hence they have less incentive to improve. Hospitals at the extreme left — but not just below the mean — truly have no marginal incentive to improve.

Estimating equation 1 via OLS using P_{ht+1} as the key explanatory variable introduces endogeneity since it is based on a lagged value of the readmission rate. For example P_{h2013} is based on performance in 2009–11 which is likely correlated with ϵ_{ht}.

$$P_{h2013} = \hat{f}(\bar{Y}_{h09-11}), \quad cov(\bar{Y}_{h09-11}, \epsilon_{ht}) \neq 0$$

This endogeneity concern includes, but is not limited to, mean reversion.[15] This type of dynamic model has been extensively analyzed (Anderson and Hsiao, 1981; Amemiya and MaCurdy, 1986; Arellano and Bond, 1991) and one solution to obtain a consistent estimate of β is using lagged or 'predetermined' characteristics of hospital h as instruments for P_{ht+1} (Arellano and Bover, 1995;

[15] If the high readmission rate over 2009–11 was a temporary aberration then a penalized hospital could plausibly return to a lower 'long-run' value 2012 onwards, even in the absence of a penalty. Appendix figure A.1 presents the trends in mean readmission rate for equal sized groups of hospitals at low, medium and high values of P_{h2013}. The time series indicates that hospitals with greater (lower) values in 2009–11 revert sharply to a lower (greater) value in 2012. Chay et al. (2005) present similar evidence from a similar setting albeit in the context of a school performance enhancement initiative. They are able to overcome this issue empirically by exploiting a sharp discontinuity determining treatment.

Gruber and Saez, 2002; Acemoglu and Finkelstein, 2008) after applying a first difference or within-group transformation. A lagged value Z_h from an earlier period is a valid instrument under the assumption $\mathbb{E}(\epsilon_{ht}\epsilon_{hs}) = 0$ for $t \neq s$ i.e. the unobserved time varying error term is not serially correlated.[16] The IV approach also mitigates concerns of measurement error in constructing hospital expectations.

I construct an instrumental variable (Z_h) based on predetermined characteristics of hospitals in two different ways to implement this strategy. I estimate within-hospital models of the following form.

$$P_{ht+1} \cdot \mathbf{1}(T=1) = \pi_h + \pi_t + \lambda_1 \cdot Z_h \cdot \mathbf{1}(T=1) + X'_{iht}\gamma_1 + \xi_{iht}$$
$$Y_{iht} = \alpha_h + \delta_t + \beta_1 \cdot P_{ht+1} \cdot \mathbf{1}(T=1) + X'_{iht}\gamma_2 + \epsilon_{iht} \tag{2}$$

Where the indicator $T = 1$ when $t \geq 2012$. The baseline instrument is a predicted readmission rate, $\hat{Y}_h$ using data from 2006–07. This is the earliest year for which I have data available and I predict this variable using patient risk factors.[17] I subsequently omit this period from the estimation sample. The identifying assumption is hospitals with low vs. high values of baseline readmission rates would evolve along parallel trends in absence of the penalty.

I also deploy an alternate instrument which exploits a key feature of the institutional setting to use variation across hospitals only in patient demographics (race and income). CMS does not include demographics in its risk adjustment algorithm. These two factors strongly predict readmission rates but were excluded deliberately by CMS due to social justice concerns. The agency was comfortable risk-adjusting for sicker patients, but not for poorer patients. Barnett et al. (2015) show that this exclusion creates a quasi-permanent handicap for hospitals located in poor, high minority share neighborhoods.[18] Hence hospitals with greater share of minority/poor patients are differentially likely to be penalized by CMS and have greater incentive to respond.

I construct the readmission rate $\hat{r}_h$ predicted by demographics of hospital patient mix in 2007–08

[16]This assumption can be relaxed to allow the error term to have a MA(q) structure with $q < T - 1$ i.e. q must be less than the length of the panel. In my setting since the first evaluation period begins in 2009 and the data begins in 2007, $q \leq 1$.

[17]Specifically, I use exactly the same patient risk factors as used by CMS in their risk adjustment algorithm, This includes gender, age and a vector of indicators for approximately forty co-morbidities that were present at the index admission or diagnosed during the one year period prior to the index admission.

[18]They perform a detailed analysis on the role of socio-economic characteristics in determining readmissions and find that excluded patient factors can explain up to 50% of the residual variation in readmission rates across hospitals after applying the CMS algorithm. They use a wealth of information about patients that I do not have access to. Specifically, in addition to race and Medicaid eligibility they use marital status, education level, household income and employment status. They find large and statistically significant differences in the race mix and Medicaid eligibility among patients in the lowest and highest quintile of hospitals (by risk adjusted readmission rate)

after risk adjusting for patient risk factors. Hence it only relies on variation in patient demographics, which are plausibly invariant over time.[19] Appendix table A.2 presents estimated coefficients from the predictive regressions that generate $\hat{r}_h$. The identification assumption underlying this strategy is that in absence of the penalty, the difference in outcomes between hospitals with high and low minority shares would remain constant.

5 Results

5.1 Expected responses

Theory suggests that hospitals could respond to the penalty through two mechanisms. First, they can make improvements in treatment quality. The hospital will now be willing to invest greater effort and cost per patient to prevent readmission and avoid the penalty. This can be achieved either by reallocating resources toward patients entering with the penalized conditions (i.e. the multi-tasking concern, for example reallocating nurses away from other departments), simply incurring greater cost per penalized patient (for example, hiring additional nurses, case managers or even a dedicated readmissions manager) or improving productivity i.e. holding resources constant, they may produce greater quality (for example, improving compliance with best clinical practices or better drug reconciliation checks and discharge planning). I test for improvements in treatment quality but have less to say about the role of underlying channels due to data limitations.

Second, hospitals may try to change the composition of their patient mix to decrease their penalty burden. While there are several possible ways to achieve this, the main mechanism I examine is changing admission decision behavior at the Emergency Department — both when patients arrive for the initial condition as well as when they return to seek readmission. The penalty clearly incentivizes hospitals to readmit fewer patients when they return. Hospitals exercise considerable discretion in admitting patients versus treating them as outpatients and the penalty may motivate them to hold marginally sick patients in the ED as opposed to admitting them. At the initial admission stage, hospitals may try to select a healthier patient mix to decrease their expected readmission risk.

This section presents both OLS and IV results for various outcomes related to these responses, beginning with testing for the aggregate effect on readmissions. I discuss the other possible mecha-

[19] Specifically, I first compute $r_{ih} = 1(Y_{ih} = 1) - X_{ih}^{c'}\gamma^c$ for each patient i where γ^c is the risk adjustment parameter made publicly available by CMS. I then compute $r_h = \frac{1}{I_h}\sum_{i \in I_h} r_{ih}$ and use it as the dependent variable in a regression with race and income being the explanatory variables (X_h^d) as $r_h = \boldsymbol{X}_h^{d'} \cdot \gamma^d + \zeta_h$. The error term ζ_h here represents the residual variation we wish to eliminate and hence I generate the predicted value $\hat{r}_h$ as $\hat{r}_h = \boldsymbol{X}_h^{d'} \cdot \hat{\gamma}^d$

nisms and associated caveats in section 5.5.

5.2 Impact on readmissions

To explore the validity of the identifying assumption underlying the IV estimates, I first present coefficients obtained by estimating the fully non-parametric equation 3 shown below.

$$Y_{iht} = \alpha_h + \delta_t + \sum_{s=2009}^{2014} \beta_s \cdot \mathbf{1}(d_{z_h} = 1) \cdot \mathbf{1}(t = s) + \epsilon_{iht} \tag{3}$$

d_{z_h} is an indicator set to one if hospital h is in the top third of all hospitals, ranked by the instrument z_h. Recall from the discussion in section 2 that hospitals with the highest readmission performance in the past perceive the greatest marginal incentive to improve. This exercise tests if hospitals with the greatest incentive responded differentially to the penalty.

Figure 5a uses the baseline instrument $\hat{Y}_{h07}$ and plots results for two outcomes, probability of readmission in 0–30 days (targeted by the penalty) and in 31–60 days, which was not penalized. All values are relative to 2008 which is normalized to one. There are three takeaways. First, the plot suggests there are no pre-existing differential trends between hospitals with low and high values of $\hat{Y}_{h07}$ which is reassuring since it relates to the identifying assumption. Second, there is a statistically significant and economically meaningful (approximately 2 percentage point) decrease in the penalized metric after the penalty was announced in 2011. There is suggestive evidence of a drop in 2011 itself which is plausible due to anticipatory responses by hospitals after the enactment of the ACA. Third, there is no corresponding response on 31–60 day readmissions, which were not penalized, ruling out the possibility of some industry-wide trend producing the deviation in 0–30 day readmission.

Figure 5b presents estimates obtained from the following equation

$$Y^j_{iht} = \alpha_h + \delta_t + \beta^j \cdot \mathbf{1}(d_{z_h} = 1) \cdot \mathbf{1}(T = 1) + \epsilon_{iht} \tag{4}$$

where instead of documenting the effect on readmission rate for each year, I quantify the effect on readmissions (β^j) in each ten day interval j (0–10,11–20,...51–60) in the first two months after discharge from the index case. Y^j_{iht} is an indicator that takes value 1 if the patient was readmitted to some hospital in interval j, else it is zero.

This helps examine the effects at a more granular time duration. Since the mean readmission rate across time intervals varies from 9% (0–10 days) to 2% (51–60 days) I normalize the estimated coefficients by the respective pre-penalty mean readmission rate. There is a marked decrease in the estimated effect between 21–30 and 31–40 days i.e. on either side of the arbitrary thirty day penalty window, although the difference is not not statistically significant.

Table 2 presents the OLS (Panel A) and IV (Panels B and C) estimates corresponding to the specifications described in Section 4. Recall that these specification do not use the dummy d_z but the actual readmission rate $\hat{Y}_{h07}$ or $\hat{r}_{h07}$ depending on the instrument. In each panel I present estimates separately for probability of readmission in 0–30 and 31–60 days (similar to the non-parametric evidence). I also present the first stage coefficient on the instrument and F-test statistic from the regression with 0–30 day readmission as the outcome. Both instruments are highly predictive with comfortably large F-test values. Columns 3–5 provide the estimates separately for each condition while the last column presents a weighted average of the estimates across conditions.

The OLS estimates are substantially larger than the IV estimates, which is consistent with the concern that mean reversion can potentially exaggerate the penalty response on readmission and related outcomes like probability of return. Second, as in figure 5a, there is little or no effect on readmissions in the 31–60 day interval (and usually statistically insignificant) while a large and statistically significant decrease in 0–30 day readmissions.

Panels B and C presents estimates from the IV specification in equation 2. Panel B uses the lagged readmission rate $\hat{Y}_{h07}$ as the instrument. To interpret magnitudes, note that the difference in probability of being penalized between the 25th and 75th percentile hospital is about 0.9 and hence the effect of moving a hospital from the 25th to the 75th percentile is approximately 1.9 (0.021*0.9) percentage point (pp). The base readmission rate before the penalty was 19.6% and hence this is equivalent to a 9% decrease relative to the pre-penalty mean value. The estimate in Panel C matches reasonably well, though smaller in magnitude. The difference stems mainly due to the estimated effect for Heart Failure, where the demographics based instrument consistently finds little or no effects. I defer all discussion on heterogeneity across conditions to section 5.5.

The estimated effect on readmissions implies a large decrease in hospitalizations overall. In 2011 there were approximately 860,000 index cases across the three penalized conditions. An average decrease of 1 percentage point in readmission implies 8,600 fewer hospitalizations *in these three conditions* alone. Extrapolating to all Medicare patients and accounting for the lower base readmission rate in remaining conditions, this implies approximately 70,000 fewer hospitalizations each year. At an average cost of $9,000 per case[20], the annual gross saving to Medicare is $ 620 million.

[20] MEDPAC (2007) report that the average readmission cost approximately $8,000 in 2007. This is equivalent to $9,200 in 2016.

Extrapolating to other insurers (Medicare Advantage) would naturally increase the estimated savings further.

There are few benchmarks to guide us in interpreting the magnitude of the obtained effect on readmission. Hansen et al. (2011) conduct a systematic review of studies on interventions to decrease readmissions and find results ranging from a 28 percentage point decrease to a 10 percentage point increase. The key issue is that most studies are on small samples (< 500 patients). Results from four large sample ($n > 5,000$) studies range from 0.7 to a 2.5 percentage point decrease which is comparable to the result in this paper. To be clear, none of these studies involved the use of financial incentives — but the literature indicates this magnitude of decrease is plausible.

5.3 Mechanisms

5.3.1 Improvement in treatment quality

The hospital's objective function has typically been modeled as a combination of profit and concern for patient utility (Clemens and Gottlieb, 2014; Dickstein, 2015; Alexander, 2015). Under this model, optimal treatment quality (including follow-up care post-discharge) will equate the marginal cost of incremental treatment (through its impact on profit) with marginal benefit to the patient. The penalty provides hospitals an incentive to increase the optimal quality level to the extent it decreases the patient's probability of readmission and in turn, the hospital's expected penalty in the future. Hence under the penalty, hospitals will weakly increase their treatment quality on average - and this response will be greater for hospitals who believe they are more likely to be penalized in the future.

Assuming that patients do not change how sick they feel to decide to return to a hospital, an implication of improvement in treatment quality is that fewer patients will return to a hospital post penalty. In this section I provide evidence on quality improvements as well as the change in probability that patients return to seek care.

Probability of return: Figure 6 plots non-parametric estimated effects obtained using equation 3. There are two dependent variables - probability of return to any hospital within 30 days (diamonds) and in 31–60 days (squares). The pattern is qualitatively similar to that for probability of readmission. Hence we take away the same implications as discussed earlier. This pattern is also evident in the simple time series plot of probability of return across all hospitals, presented in appendix figure A.2 Panel A. Note that the magnitude of the estimated decrease in return suggested by the non-parametric estimates is *smaller* than the corresponding decrease in probability of readmission.

Table 3 presents OLS (Panel A) and IV (Panels B and C) estimates obtained using equations 1 and 2 respectively with probability of return as the dependent variable. Results from the preferred specification (Panel B) indicate an aggregate effect of 1.8 percentage points. Interpreting this in the same way as readmissions, it implies that moving a hospital from the 25th to 75th percentile likelihood of being penalized is associated with a 1.5 percentage point decrease. Relative to the mean pre-penalty value of 27 pp, this represents a 5.5% decrease in probability of return. Estimates in Panel C are smaller in magnitude, but qualitatively similar — except in case of Heart Failure.

Compliance with best practices: Treatment quality is multi-dimensional and can be measured through several types of measures. One such metric is hospital compliance with 'best-practice' treatment protocols. As discussed in Section 3, I gather annual hospital scores on process compliance released by CMS and standardize them for use in regression analysis. Appendix table A.3 presents the results from OLS (Panel A) and IV (Panels B and C) with these scores as outcomes. Interpreting the coefficients is easier in this case since the scores are transformed to have mean zero and standard deviation one. Hence moving a hospital from the 25th to the 75th percentile likelihood of being penalized results in a change in process score of approximately 0.1 standard deviations. The magnitude of the estimated effect is large but not always statistically significant, perhaps due to sampling noise (since the scores pertain to all patients not just Medicare patients).

Other channels: I tested and failed to reject null hypotheses of no change in other possible channels of quality improvements (not reported). For example I find hospitals are not keeping patients longer in response to the penalty. I also find no increase in the probability of having a doctor's visit within fifteen days of discharge from the index case, suggesting that increased coordination with primary care physicians is not quantitatively important. It also suggests no substantial increase in treatment costs per case, although this only reflects billed costs.

5.3.2 Changes in admission decisions

One way to influence the readmission rate without improving treatment quality is to systematically change admission thresholds for patients when they arrive at the Emergency department. Hospitals can use this to their advantage both at the time of initial admission as well as when patients return.

Initial admission: Although 90% of Medicare patients with the penalized conditions are admitted when they arrive in the initial case, there is considerable dispersion across hospitals in the admission rate. The penalty decreases financial attractiveness of inpatient treatment relative to outpatient care. In the context of the simple model where hospitals maximize a combination of profit and patient utility, the average probability of admission should decrease due to the penalty and the decrease should be greater (lower) for hospitals facing a greater (lower) penalty rate applicable to admissions in the current year. Further, future penalty considerations provide an incentive to alter the case mix composition to lower ex-ante readmission risk.

Figure 7 plots corresponding non-parametric estimated effects for each year obtained by estimating equation 3 with probability of admission as the dependent variable. It confirms that the decrease in probability of admission is differentially greater for hospitals expecting greater penalty and it begins in 2011. The figure indicates that probability of admission at one-third of hospitals with highest readmission rate in 2007 decreased by about 1 percentage point relative to the remaining hospitals. This confirms the intuition provided by the aggregate time series trend in appendix figure A.2 Panel B.

Table 4 presents OLS (Panel A) and IV (Panels B and C) estimates obtained with probability of admission as the dependent variable. This is the only outcome where instead of exploiting variation in ex-ante probability of being penalized (P_{ht+1}) across hospitals, I use the penalty rate p_{ht} since this is the key factor affecting expected profit from the index case. The baseline estimates (Panel B) indicate that moving a hospital from the 25th to 75th percentile penalty rate is associated with a decrease of 0.7 (0.14*0.05) pp. The estimate in Panel C is larger and implies a decrease of 2 (0.4*0.05) pp. These are statistically significant but small relative to the mean probability of admission of 89 pp in the pre-penalty period.

It is important to understand whether hospitals have changed their admitting behavior to decrease the ex-ante readmission risk of their patient mix. As a first pass to answering this question I test if admitting behavior is correlated with observables. I first estimate a predictive probit model of readmission risk using 2006–07 data on a vector X of risk factors like gender, age and nearly 20 co-morbidities.[21]

$$P(Y_i = 1) = \Phi(X_i \gamma)$$

[21] These co-morbidities are used to construct the Charlson Co-morbidity Index which is frequently used in the health and economics literature to indicate sickness severity and potential for resource usage in hospitals.

Where Φ denotes the normal CDF. I apply the obtained parameter vector $\hat{\gamma}$ to cases in the estimation sample to generate an ex-ante predicted probability of readmission for each case, $\hat{P}(Y_i = 1)$. I use this as an index that flexibly captures readmission risk based on multiple factors. I then estimate a standard difference-in-difference-in-difference (DDD) specification to answer the question whether penalized hospitals have been differentially avoiding riskier patients after the penalty was introduced.

$$\begin{aligned} Y_{iht} = \alpha_h + \delta_t + \lambda \cdot \hat{P}(Y_i = 1) + \beta_1 \cdot z_{h07} \cdot \mathbf{1}(T = 1) + \beta_2 \cdot z_{h07} \cdot \hat{P}(Y_i = 1) \\ + \beta_3 \cdot \hat{P}(Y_i = 1) \cdot \mathbf{1}(T = 1) + \beta_4 \cdot z_{h07} \cdot \hat{P}(Y_i = 1) \cdot \mathbf{1}(T = 1) + \epsilon_{iht} \end{aligned} \tag{5}$$

z_{h07} refers to either of the two instruments discussed in section 4. λ captures the pre-penalty marginal effect on probability of admission of an increase in readmission risk. β_1 captures the differential impact for all patients at higher penalty hospitals in the post-period. β_2 captures differential impact for marginally riskier patients at higher penalty hospitals in the pre-period. β_3 captures the differential impact for riskier patients in the post-period across hospitals. β_4 is the standard DDD estimator and captures the marginal impact on probability of admission for riskier patients at higher penalty hospitals relative to lower penalty hospitals, after the penalty was introduced.

Table 5 presents results from estimating equation 5 using each of the two instruments. Across conditions we see a consistent pattern - the average probability of admission has declined in the post period at penalized hospitals (indicated by β_1) while the marginal effect (β_4) of riskier patients is to *increase* probability of admission. Somewhat surprisingly, this indicates that the decline in average probability of admission is driven entirely by the healthiest patients, rather than the sicker patients (on observable risk). This can be explained by the simple economic model discussed above if sicker patients are more profitable than marginal patients for the hospital, or incur less readmission penalty than marginal patients after accounting for risk adjustment.

Admission when patients return: When patients return to seek care, hospitals again take an admission decision. The penalty creates different incentives for hospitals when they consider their own returning patient vs. a patient discharged by another hospital. An additional readmission of their own patient will potentially add to their penalty in the future depending on how they compare to the national average readmission rate. However, readmitting a patient originally discharged by another hospital will *never* increase their penalty. Hence we should expect a greater decrease in probability of readmission for patients returning to the same hospital relative to those returning to a different hospital. In the evaluation period 2009–11, approximately 80% of patients returned to the same hospital as in the index case. The outcome of interest is an indicator for readmission for the patients returning after being discharged from a hospital for one of the penalized conditions.

Figure 8 graphically depicts the research design by plotting estimates from equation 3. The non-parametric evidence confirms that while the probability of readmission drops substantially for patients returning to the same hospital, the decrease is much smaller and not statistically significant for patients returning to a different hospital. In fact there is a slight increase in the probability of readmission for patients returning to other hospitals in 2012 and 2013. Appendix figure A.2 Panel C presents the corresponding time series pattern in aggregate probability of readmission for all returning patients. It was already declining in the period 2008–11, but there is a noticeable acceleration in the decline post 2011.

Table 6 presents the OLS (Panel A) and IV (Panels B and C) estimates. The regression estimates confirm the intuition provided by the non-parametric estimated effects. There is a large and statistically significant decrease in probability of readmission for patients returning to the same hospital while a smaller and in most cases statistically insignificant effect for patients returning to a different hospital. Both IV estimates are reassuringly similar.

The baseline results suggest that moving a hospital from the 25th to the 75th percentile probability of being penalized is associated with approximately 3.2 pp (3.5*0.9; 4%) decrease in probability of readmission for their own patients. The corresponding value for patients returning to a different hospital is zero. Panel C estimates are similar on patients returning to the same hospital but suggest that patients returning to a different hospital may be *more likely* to be readmitted, though the estimates are not statistically significant.

How could this change in admitting behavior be implemented in practice? A possible mechanism is the increased use of observation status for returning patients, which allows hospitals to monitor patients on-site for up to two days without formally admitting them. There has been a general trend of increase in the use of observation status for Medicare patients by hospitals and it has received media attention, but it could also be linked to the readmission penalty.[22] To test this hypothesis I exploit the same thought experiment as before of comparing the experience of patients returning to the same vs. different hospitals. Figure 9 and table 7 respectively present the non-parametric

[22] Several news articles have noted the general increase in use of observation status, see http://www.reuters.com/article/us-column-miller-medicare-idUSKCN11Z1DU. One recent article linked it to readmissions: http://www.wsj.com/articles/medicare-rules-reshape-hospital-admissions-1449024342.

estimated effects (equation 3) and regression results (equation 2). The baseline results indicate that hospitals have steadily increased the use of observation status for their own patients but there is no consistent pattern for patients returning to other hospitals.

5.3.3 Quantitative role of different mechanisms

I perform simple back-of-the-envelope calculations using the baseline (Panel B) estimates presented in tables 2, 3 and 6 to learn more about the quantitative role of treatment quality and selective readmission responses. Table 8 summarizes these calculations. Panel A repeats the pre-penalty mean values presented in Table 1 of probability of return, probability of readmission conditional on return and the unconditional probability of readmission (which is a product of the two). These values serve as useful benchmarks to compare the penalty response.

Panel B presents estimates of probability of readmission post-penalty under different scenarios. These estimates are calculated using the coefficient values as-is without applying any specific interpretation. Since the goal is to attribute share of total decrease in readmission to the two mechanisms, the percentage point decrease is not relevant here.

First I present the estimate assuming hospitals respond both by improving treatment quality as well as selectively readmitting patients, which is what I observe empirically. The probability of return decreases from 0.26 to 0.24 while the probability of readmission conditional on return decreases from .75 to 0.72.[23] As a result, the unconditional probability of readmission decreases from 0.196 to 0.175. To clarify, this is not the independently estimated effect on the probability of readmission — but it matches nearly exactly that baseline coefficient in Table 2 of -0.021. This indicates that these two mechanisms collectively account for the estimated effect on overall readmission.

The next scenario allows only a decrease in probability of return holding the conditional probability of readmission at the pre-penalty value. This can be generated using two different assumptions. First, suppose that the improvement in treatment quality affects all patients equally. Then the patients who do not return will be the ex-ante healthiest patients. These patients were also less likely to be readmitted conditional on return, hence the remaining patient mix will have a higher readmission rate relative to the average. The unconditional probability of readmission would then be 0.184 instead of 0.175 discussed above. An alternate assumption is less intuitive, but allows for the possibility that patients who do not return under penalty to have the average probability of readmission. The

[23]Here I have to weight the estimates for patients returning to the same (80%) and different (20%) hospitals.

unconditional probability of readmission would then be a marginally lower 0.183. In either case, improvement in treatment quality can account for 55–60% of the aggregate decrease in readmission.

If I allow hospitals to only respond by selectively readmitting patients and hold the probability of return constant at the pre-penalty value then the probability of readmission would be 0.188 instead of 0.175.

5.4 Patient health

5.4.1 All patients

Table 9 presents the estimated effect on different measures of short term mortality - at seven, thirty and ninety days post discharge respectively. I construct indicators for these mortality outcomes for all patients admitted with one of the penalized conditions as well as arrivals at the ED who were not admitted. The ED patients are included in the analysis to avoid conditioning on the hospital's admission decision which has already been shown to be one of the response margins. Table 9 presents OLS (Panel A) and IV (Panels B and C) results and follows the usual format as with the previous tables.

The general pattern of estimated coefficients suggests that mortality has declined, regardless of which time duration we choose. The baseline estimates (Panel B) indicate that 30 day mortality has declined by 0.4 percentage point and this implies that moving a hospital from the 25th to the 75th percentile is associated with a 0.35 pp (0.4*0.9, 3%) decrease in mortality. The corresponding decrease in 7 day and 90 day mortality is similar at 0.3 (4%) and 0.6 (3%) pp respectively. Note that the estimated effects are significant only at the 10% level. Figure 10 presents estimated non-parametric effects for each year on 90 day mortality using equation 3.

A different Medicare performance pay scheme (Value Based Payments, VBP) was introduced at the same time as HRRP and independently incentivizes hospitals to decrease their risk adjusted mortality in the same three conditions. It is possible that the estimated effects on mortality partly reflect improvements made to benefit under VBP, however it is unlikely to be the driving factor for two reasons. First, VBP provides very low powered, diffused incentives relative to HRRP. The penalty magnitude is similar to that of HRRP in dollar terms, but hospitals are graded on relative performance in nearly thirty quality and efficiency metrics. One of these is mortality, which was not given any weight until 2014 and even then accounted for 25% of the score (Norton et al., 2016). Second, in results not summarized here I find mortality gains among Chronic Obstructive Pulmonary

Disorder (COPD) patients, a fourth condition penalized by HRRP 2014 onwards and not included in VBP at all.

Previous studies have found that the use of intensive treatments (for example, use of cardiac catheterization) for heart attacks is associated with a decrease of 8–12 percentage points in 30 day mortality (McClellan et al., 1994; McClellan and Newhouse, 1997). In comparison I find a 1–2 percentage point decrease in 30 day mortality for heart attack patients. Hence the readmission penalty has delivered mortality gain 12–25% as large as the impact of using a new therapy technology without additional cost to Medicare. In a recent paper Doyle et al. (2015) show that Medicare pays hospitals approximately 10% more to obtain a 2 percentage point improvement in one year mortality. Applying these estimates out-of-sample in my setting implies that Medicare would have to pay 2.5%[24] more per patient to obtain the estimated mortality gain. This is equivalent to approximately $280 per patient in any of these penalized conditions.

5.4.2 Returning patients

One concern with changing the standard of care (i.e. substituting inpatient with outpatient care) for some returning patients is that their health may suffer relative to if they were treated as inpatients. I follow these patients and construct an indicator for an adverse outcome within 15 days — either death or return to the ED. I chose a relatively short time window because it makes a causal link between the adverse outcome and the ED's readmission decision more plausible.

Figure 11a presents non-parametric estimates of the effect on probability of ED visit within 15 days of returning post discharge for each year, and figure 11b presents corresponding estimates on mortality within 15 days of returning to a hospital. As with the previous figures, these estimates were obtained using equation 3. The figure shows an increase in probability of ED use for patients returning to the same hospital, but not for those that return to a different hospital. However, the mortality results indicate no effect on either set of patients.

These figures suggest there are no negative mortality effects for the marginal patient affected by the readmission decision, but she is more likely to make an ED visit soon after being denied readmission at the hospital. I take this as suggestive evidence of some harm to patients affected by the change in readmission behavior, though no detectable effect on mortality.

[24] A caveat here is that their estimated effect is on one year mortality which I apply to the gain in 90 day mortality in my setting. This assumes that the gains persist as-is until 1 year. I find the mortality gain is relatively stable between 30 and 90 days and so this seems reasonable. Their estimates are based on non-deferrable conditions which include Heart Attack and Pneumonia but exclude Heart Failure. The average index case costs Medicare $12,000.

5.5 Discussion

A consistent pattern across all results is that the decrease in readmission is lower for Heart Failure relative to the other two conditions *and* a substantially lower share of it is accounted for by improvement in treatment quality. The decrease in probability of readmission is approximately a third that for the other two conditions (1 pp vs. $\approx$ 3 pp), however the decrease in probability of return is less than a quarter that for the other two conditions (0.6 pp vs. $\approx$ 2.7 pp). As a result only 40% of the decrease in readmission can be accounted for by quality improvement. Results using the alternate instrument find that the probability of return may have *increased* for Heart Failure patients i.e. they may be worse off.

I also find that process compliance decreased for penalized hospitals in case of Heart Failure patients. This may well be driven by sampling error or the experience of non-Medicare patients (recall that process of care scores are based on a random sample of all patients), but it could indicate that hospitals have not been able to improve treatment quality in response to the penalty.

Heart Failure is a chronic condition that sometimes leads to hospitalization due to exacerbations, but is mainly managed through consultations with primary physicians or cardiologists in their offices. Heart Attack and Pneumonia are highly acute conditions that are more likely to be managed through hospitalization. This fundamental difference between the conditions could explain the diminished role of quality improvement in case of heart failure. Hospitals may feel more comfortable relying on increased coordination with primary physicians in case of Heart Failure relative to the other two conditions.

It is useful to discuss an important limitation of the empirical strategy. The analysis assumes that the sickness distribution of patients with the penalized conditions arriving at a hospital is unaffected by introduction of the penalty. This can be violated by demand and supply side responses not considered here. First, information about penalty status could result in greater demand for lower readmission hospitals or sorting of riskier patients toward higher quality hospitals. The specifications account for observable patient risk factors — and indicate this is not the case — but if the sorting is based on unobservable risk factors, it could exaggerate the improvement in quality due to the penalty. Existing evidence on whether demand responds to quality reporting is mixed (Kolstad, 2013) but suggests that demand response is much smaller in case of patients admitted through the ED (Chandra et al., 2016). Since about 80% of the patients in the three penalized conditions are initially admitted via the ED, demand response is perhaps of lesser importance in this setting.

Second, the analysis currently ignores the possibility of other hospital responses to reduce their exposure to Medicare or at least to the penalized conditions. This could be implemented by strategic relabeling of patients to appear as if they came with a related-but-not-penalized condition, or by actively targeting other insurers or conditions. The latter channel seems more important in the longer term. Both channels would result in a differentially greater decrease in patient arrivals at penalized hospitals after the introduction of the penalty. In analysis not summarized here, I find no evidence of a differential decrease in the number of patient arrivals in Heart Attack and Pneumonia at penalized hospitals, and a decrease in Heart Failure cases. This is consistent with the general pattern of results where hospitals seem more willing to substitute inpatient care for outpatient care in case of Heart Failure, but not the other two conditions.

6 Conclusion

This paper presents new empirical evidence on the effects of introducing performance pay for hospitals to incentivize quality of care. I exploit the introduction of a large national performance pay scheme called the Hospital Readmissions Reduction Program. The program provides differential penalty incentives across hospitals based on their predetermined characteristics. Motivated by institutional details, the empirical analysis relies on an instrument variable strategy where the instruments exploit variation in lagged readmission performance or in patient demographics across hospitals.

I find that moving a hospital from the 25th to 75th percentile likelihood of being penalized is associated with a decrease of 1.9 percentage point (9%) in readmission rate over 2012–14. The decrease in readmission is accompanied by a (marginally significant) decrease in short-term mortality, an unambiguous indicator of improvement in patient health. I also examine the role of two mechanisms — improvement in quality of care and changes in admission decisions by hospitals. I find that improvement in treatment quality accounts for 55–60% of the aggregate decrease. Changes in the readmission decision account for the remainder of the decrease, hence it is quantitatively important and I find suggestive evidence of harm to affected patients.

Three factors contributed to the success of this particular scheme which may not apply to other performance pay initiatives. Firstly, readmission as a quality metric proved to be well-chosen in the sense that it is correlated with treatment quality. If the chosen metric is not sensitive to quality improvements (eg: patient satisfaction scores) then the penalty simply transfers risk to hospitals and may result in unintended consequences including hospitals trying to diminish the importance of the insurer imposing the penalty. Second, the penalty conditions chosen are highly acute with substantial short term mortality. The seriousness of patient situations would naturally deter manipulation responses. This may not be equally true when performance pay is extended to non-acute or elective hospitalizations, for example hip replacements. Third, it is possible that the magnitude of gaming observed is lower than what hospital administrators would like, due to agency issues with physicians who actually make the admission and treatment decisions and are typically self-employed. If performance pay targets physicians directly — as a recently introduced law does[25] — then physicians may engage in more manipulation than found here.

I think of this study as a first step towards building a comprehensive understanding of performance pay in health care. In addition to the limitations discussed previously, there are other facets not explored here. For example, did the quality improvements for Medicare spillover to non-Medicare patients? This could be an important source of welfare gain or loss depending on the net effects and is ultimately an empirical question. Second, perhaps a longer term consequence is that such penalty schemes could motivate marginal hospitals to exit Medicare or at least the penalized conditions and this could inadvertently hurt patients. More empirical work is needed on these issues to inform design of performance pay in health care.

[25]The Medicare Access and CHIP Reauthorization Act of 2015 (MACRA) provides quality incentives to physicians and nurses directly by way of bonus or penalty payments related to their Medicare Part B billing.

References

Acemoglu, Daron and Amy Finkelstein, "Input and Technology Choices in Regulated Industries: Evidence from the Health Care Sector," *Journal of Political Economy*, 2008, *116* (5), 837–880.

Alexander, Diane, "How do Doctors Respond to Incentives? Unintended Consequences of Paying Doctors to Reduce Costs," Technical Report 2015.

Alpert, Abby, "The anticipatory effects of Medicare Part D on drug utilization," *Journal of Health Economics*, 2016, *49*, 28 – 45.

Amemiya, Takeshi and Thomas E MaCurdy, "Instrumental-variable estimation of an error-components model," *Econometrica*, 1986, pp. 869–880.

Anderson, Theodore Wilbur and Cheng Hsiao, "Estimation of dynamic models with error components," *Journal of the American statistical Association*, 1981, *76* (375), 598–606.

Arellano, Manuel and Olympia Bover, "Another look at the instrumental variable estimation of error-components models," *Journal of econometrics*, 1995, *68* (1), 29–51.

_ **and Stephen Bond**, "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *The review of economic studies*, 1991, *58* (2), 277–297.

Axon, R Neal and Mark V Williams, "Hospital readmission as an accountability measure," *Journal of the American Medical Association*, 2011, *305* (5), 504–505.

Barnett, Michael L, John Hsu, and J Michael McWilliams, "Patient characteristics and differences in hospital readmission rates," *JAMA Internal Medicine*, 2015, *175* (11), 1803–1812.

Blumenthal, David, Melinda Abrams, and Rachel Nuzum, "The affordable care act at 5 years," *New England Journal of Medicine*, 2015, *372* (25), 2451–2458.

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson, "Healthcare exceptionalism? Performance and allocation in the US healthcare sector," *American Economic Review*, 2016.

Chay, Kenneth Y, Patrick J McEwan, and Miguel Urquiola, "The Central Role of Noise in Evaluating Interventions That Use Test Scores to Rank Schools," *American Economic Review*, 2005.

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

Coleman, Eric A, Jodi D Smith, Devbani Raha, and Sung joon Min, "Posthospital medication discrepancies: prevalence and contributing factors," *Archives of internal medicine*, 2005, *165* (16), 1842–1847.

Currie, Janet and Jonathan Gruber, "Saving Babies: The Efficacy and Cost of Recent Changes in the Medicaid Eligibility of Pregnant Women," *Journal of Political Economy*, 1996, *104* (6), 1263–1296.

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

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

Dickstein, Michael, "Physician vs. patient incentives in prescription drug choice," 2015.

Doyle, Joseph, John Graves, Jonathan Gruber, and Samuel Kleiner, "Measuring returns to hospital care: Evidence from ambulance referral patterns," *The journal of political economy*, 2015, *123* (1), 170.

Duggan, Mark G., "Hospital Ownership and Public Medical Spending," *The Quarterly Journal of Economics*, 2000, *115* (4), 1343–1373.

Dumont, Etienne, Bernard Fortin, Nicolas Jacquemet, and Bruce Shearer, "Physicians' multitasking and incentives: Empirical evidence from a natural experiment," *Journal of Health Economics*, 2008, *27* (6), 1436–1450.

Ellis, Randall P and Thomas G McGuire, "Hospital response to prospective payment: moral hazard, selection, and practice-style effects," *Journal of health economics*, 1996, *15* (3), 257–277.

Gruber, Jon and Emmanuel Saez, "The elasticity of taxable income: evidence and implications," *Journal of public Economics*, 2002, *84* (1), 1–32.

Hannan, Edward L, Michael J Racz, Gary Walford, Thomas J Ryan, O Wayne Isom, Edward Bennett, and Robert H Jones, "Predictors of readmission for complications of coronary artery bypass graft surgery," *Jama*, 2003, *Journal of the American Medical Association* (6), 773–780.

Hansen, Luke O., Robert S. Young, Keiki Hinami, Alicia Leung, and Mark V. Williams, "Interventions to Reduce 30-Day Rehospitalization: A Systematic Review," *Annals of Internal Medicine*, 2011, *155* (8), 520–528.

Holmstrom, Bengt and Paul Milgrom, "Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design," *Journal of Law, Economics, & Organization*, 1991, *7*, 24–52.

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

Kripalani, Sunil, Frank LeFevre, Christopher O Phillips, Mark V Williams, Preetha Basaviah, and David W Baker, "Deficits in communication and information transfer between hospital-based and primary care physicians: implications for patient safety and continuity of care," *Journal of the American Medical Association*, 2007, *297* (8), 831–841.

Lazear, Edward P, "Performance Pay and Productivity," *The American Economic Review*, 2000, *90* (0), 1346–1361.

McClellan, Mark and Douglas Staiger, "The quality of health care providers," Technical Report, National bureau of economic research 1999.

_ **and Joseph P Newhouse**, "The marginal cost-effectiveness of medical technology: a panel instrumental-variables approach," *Journal of Econometrics*, 1997, *77* (1), 39–64.

_ **, Barbara J McNeil, and Joseph P Newhouse**, "Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality?: analysis using instrumental variables," *Journal of American Medical Association*, 1994, *272* (11), 859–866.

McMorrow, Stacey and John Holahan, "The widespread slowdown in health spending growth: Implications for future spending projections and the cost of the Affordable Care Act," Technical Report, Robert Wood Johnson Foundation 2016.

MEDPAC, "Report to the Congress: Promoting greater efficiency in Medicare," Technical Report 2007.

_ , "A data book: Health care spending and the Medicare program," Technical Report 2013.

Mellor, Jennifer, Michael Daly, and Molly Smith, "Does It Pay to Penalize Hospitals for Excess Readmissions? Intended and Unintended Consequences of Medicare's Hospital Readmissions Reductions Program," *Health Economics*, 2016.

Mullen, Kathleen J, Richard G Frank, and Meredith B Rosenthal, "Can you get what you pay for? Pay-for-performance and the quality of healthcare providers," *The Rand journal of economics*, 2010, *41* (1), 64–91.

Naylor, Mary D, Dorothy Brooten, Roberta Campbell, Barbara S Jacobsen, Mathy D Mezey, Mark V Pauly, and J Sanford Schwartz, "Comprehensive discharge planning and home follow-up of hospitalized elders: a randomized clinical trial," *Journal of the American Medical Association*, 1999, *281* (7), 613–620.

Norton, Edward C, Jun Li, Anup Das, and Lena M Chen, "Moneyball in Medicare," Technical Report, National Bureau of Economic Research 2016.

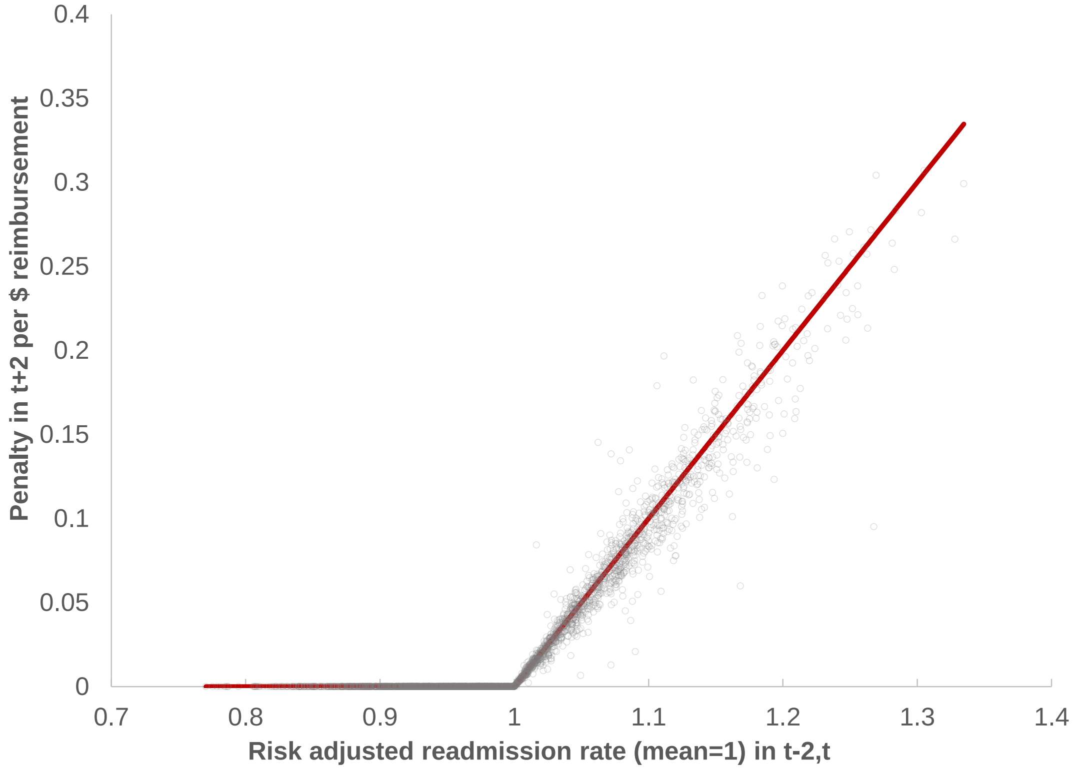
PHC4, *Pennsylvania's Guide to Coronary Artery Bypass Graft Surgery, 1994-1995*, Pennsylvania Health Care Cost Containment Council, 1996.

Podgursky, Michael J and Matthew G Springer, "Teacher performance pay: A review," *Journal of Policy Analysis and Management*, 2007, *26* (4), 909–949.

Rosenthal, Meredith B, Richard G Frank, Zhonghe Li, and Arnold M Epstein, "Early experience with pay-for-performance: from concept to practice," *Journal of American Medical Association*, 2005, *294* (14), 1788–1793.

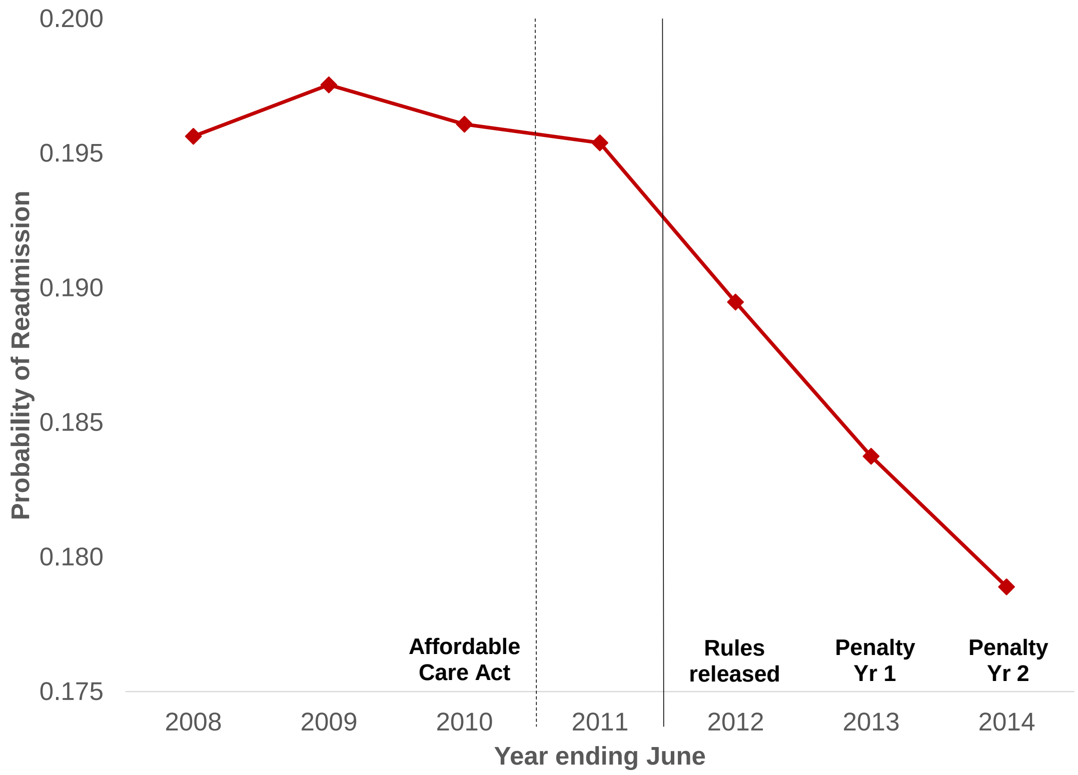
Williams, Scott C, Stephen P Schmaltz, David J Morton, Richard G Koss, and Jerod M Loeb, "Quality of care in US hospitals as reflected by standardized measures, 2002–2004," *New England Journal of Medicine*, 2005, *353* (3), 255–264.

Figure 1: Readmission rate and penalty incentive



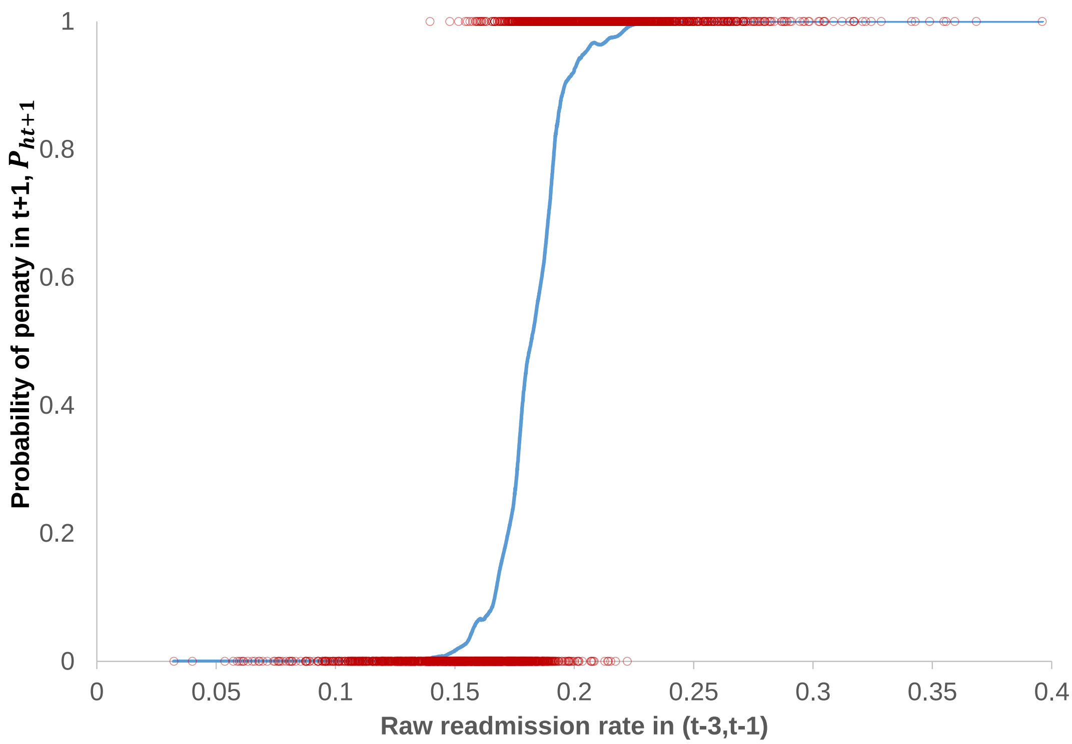
Note: Each observation is a hospital. The figure makes a conceptual point by depicting the linear relationship (solid line) between risk adjusted readmission rate over years $(t-2, t)$ on the X-axis and penalty rate in year $t+2$ on the Y-axis. It also superimposes the actual penalty rate (circles) imposed in 2013 for Pneumonia against the risk adjusted readmission rate over 2009-11. The risk adjusted readmission rate is normalized by CMS so that the mean equals one.

Figure 2: Trend in Probability of readmission



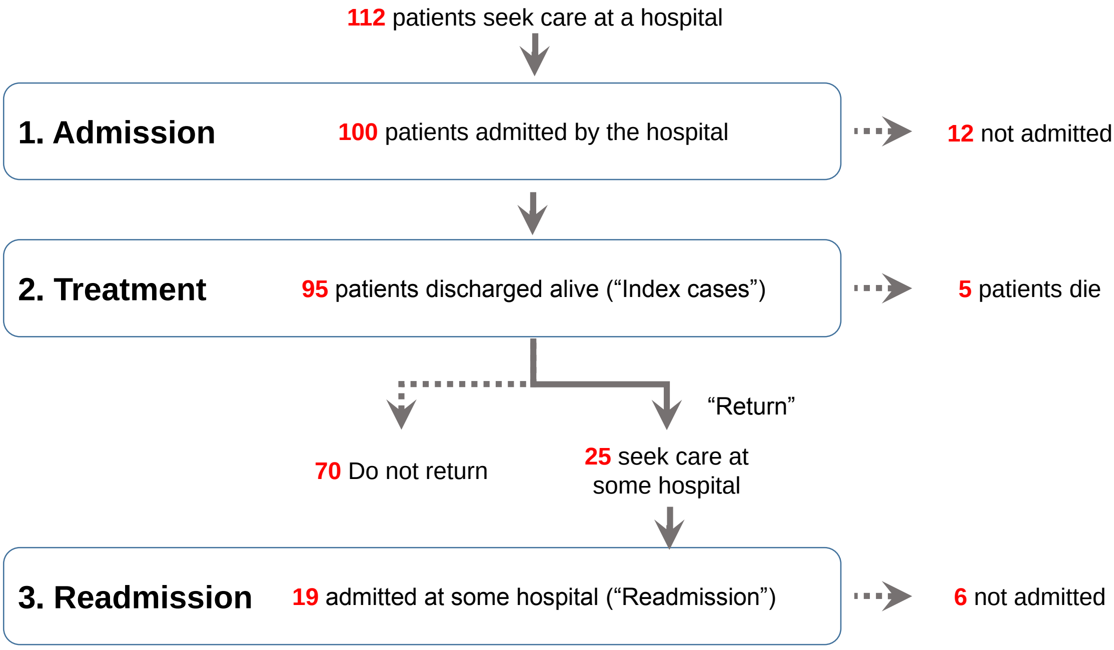
Note: This figure plots the time series of thirty-day average probability of readmission across Heart Attack, Heart Failure and Pneumonia. The Affordable Care Act was enacted in 2010 and formally introduced the Hospitals Readmissions Reduction Program. Details like penalized conditions and penalty formula were announced in August 2011. Penalty was first imposed in 2013, based on performance over 2009-11. Similarly, penalty in 2014 was set based on performance in 2010-12.

Figure 3: Measure of hospital's expectation of future penalty



Note: This figure illustrates the construction of a measure of a forward looking hospital's expectation of being penalized in $t + 1$ that determine it's decisions in t. The solid line indicates the non-parametric local linear fit of penalty status in year $t + 1$ predicted by raw readmission rate in $(t - 3, t - 1)$. The figure plots the special case of penalty status in 2013 against readmission rate in 2009-11. It also plots the actual penalty status in Year 1 for each hospital (circles). This particular figure plots data for Pneumonia, however the pattern is strikingly similar across conditions.

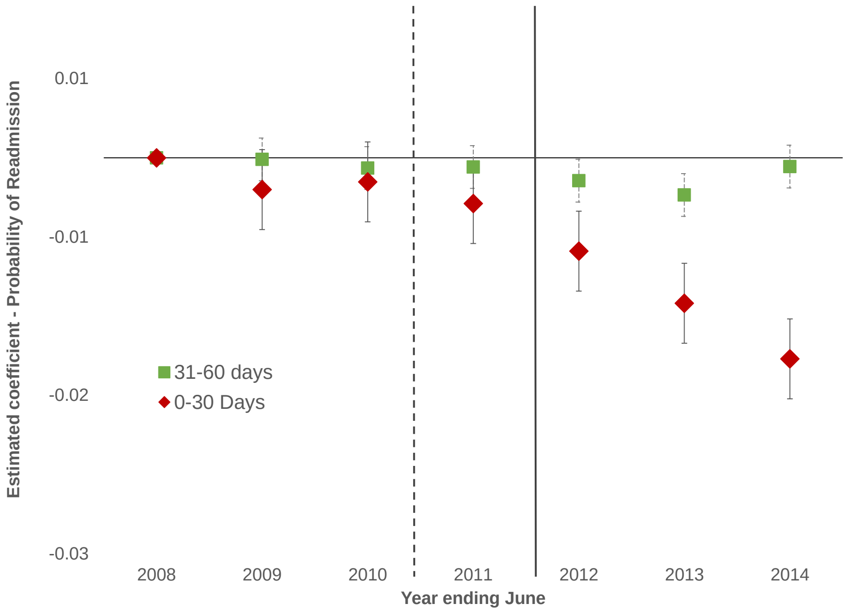
Figure 4: Stylized readmission cycle



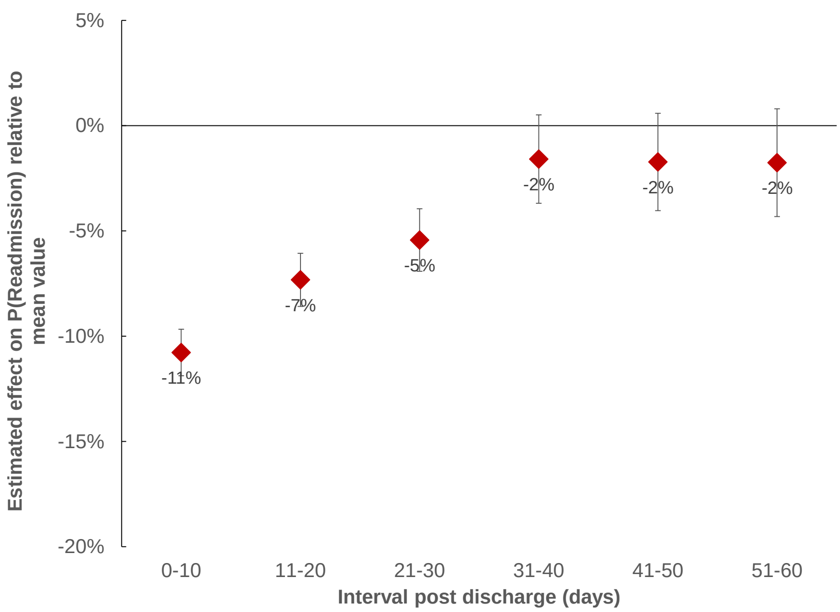
Note: This figure presents a stylized flowchart of the steps a patient undergoes during the course of an admission and readmission. The numbers are normalized such that 100 patients are admitted to the hospital initially. The proportions match the actual proportions for Medicare hospital cases with Heart Attack, Heart Failure and Pneumonia in the 2007-11 period i.e. before the Readmissions reduction program was implemented. As the figure indicates, index cases correspond to patients discharged alive at the initial admission. Technically, patients discharged against medical advice or transferred out are also not considered index cases but occur very rarely ($< 1\%$). The return and readmission can be to any hospital, not just the original discharging hospital.

Figure 5: Impact on probability of readmission

(a) By Year

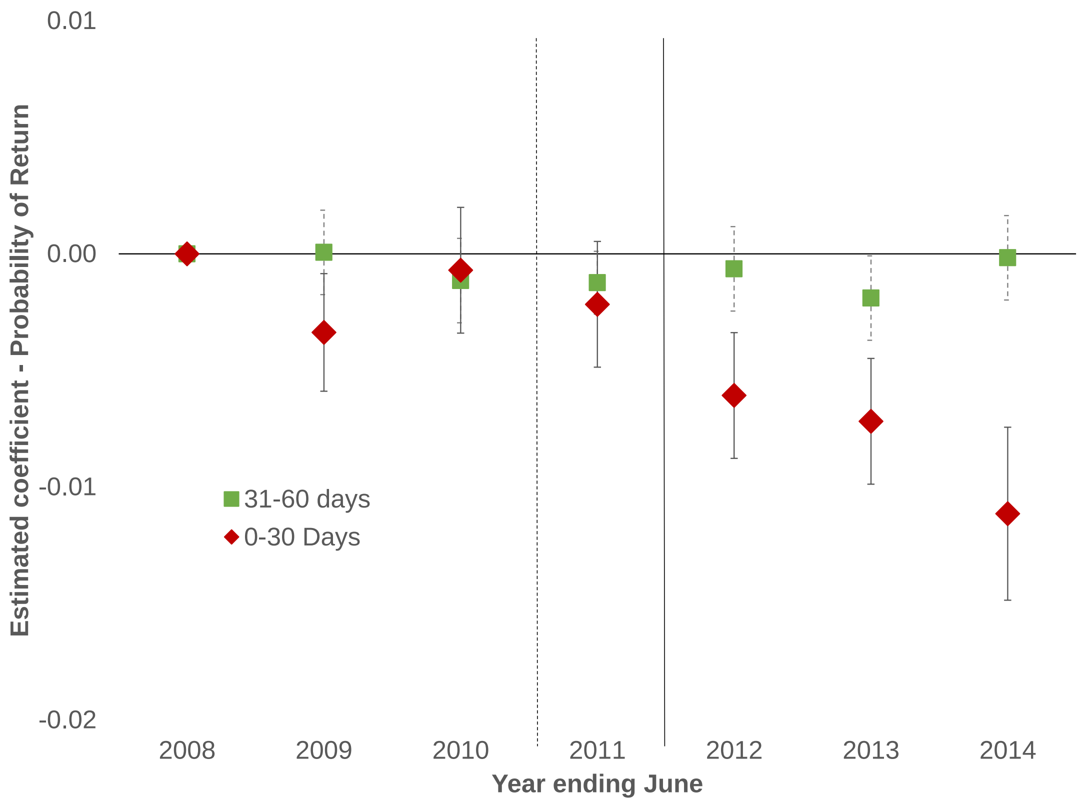


(b) By ten day Interval



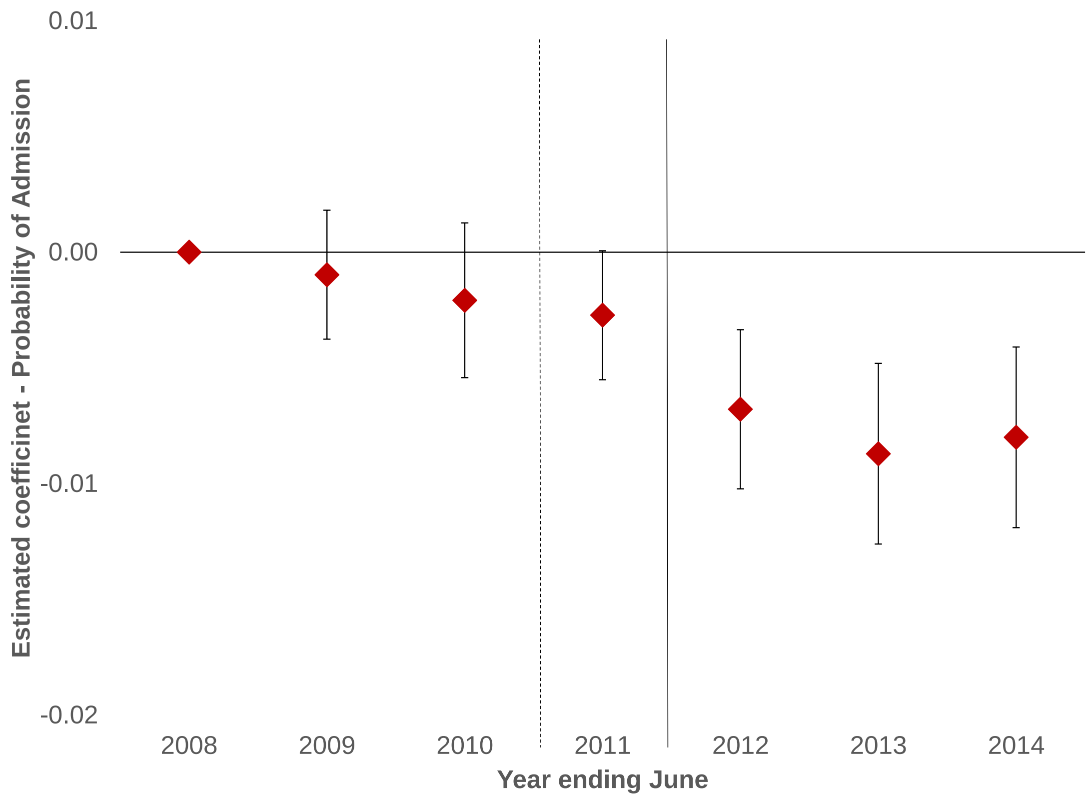
Note: This figure presents estimated coefficients from a fully non-parametric specification on probability of readmission in each year from 2008-14 (Panel A) and in each ten day interval post discharge from zero through sixty days (Panel B). In Panel A, 2008 is the omitted year. The dependent variables are probability of readmission to any hospital in 0-30 (diamond) and 31-60 (square) day windows respectively. Coefficients were estimated independently for each condition using equation 3 (Panel A) or 4 (Panel B) and then a weighted average value was calculated. Standard errors are clustered by hospital. Error bars indicate 5% confidence intervals. To interpret the coefficients, note that the mean probability of readmission in the pre-period is 0.2 and 0.06 in 0-30 and 31-60 days respectively. In Panel B, coefficient values are normalized by the mean value of readmission for the particular interval since there is wide variation in the average probability of readmission by interval.

Figure 6: Probability of return



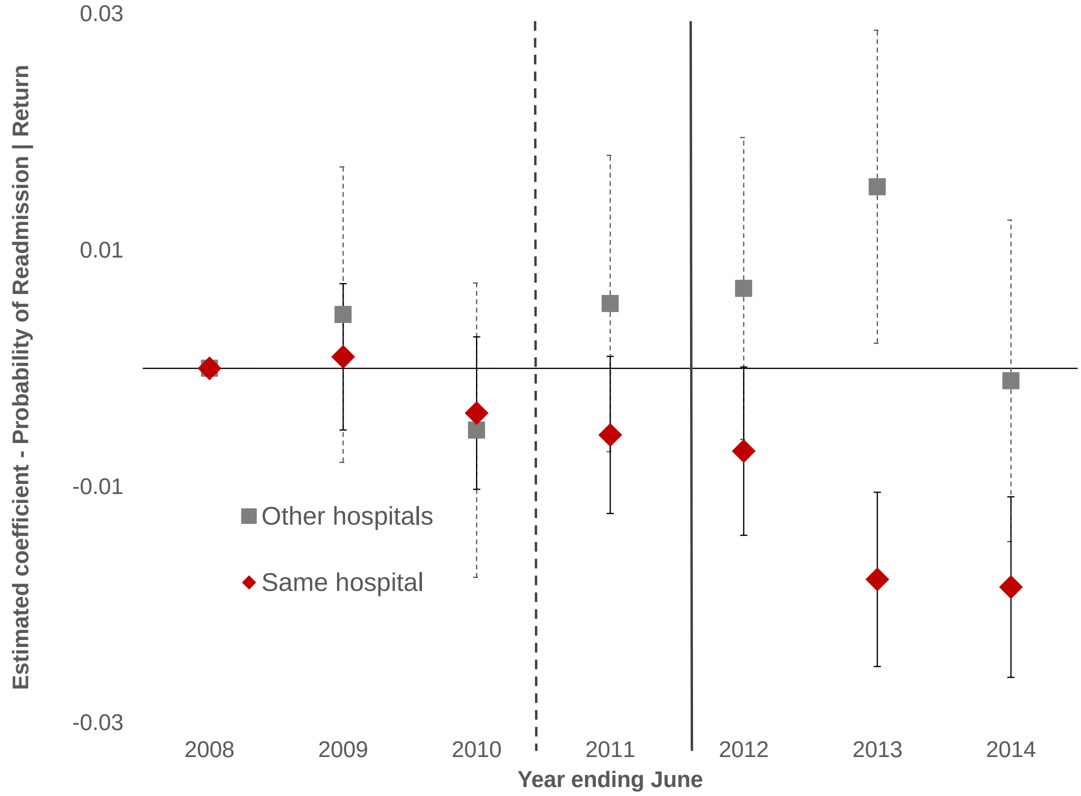
Note: This figure presents the aggregate estimated coefficients across conditions by year with 2008 being omitted. Dependent variable is probability of readmission to any hospital. Coefficients were estimated using equation 3. Standard errors are clustered by hospital. Error bars indicate 5% confidence intervals. To interpret the coefficients, note that the mean probability of return (pre-HRRP) in 0-30 and 31-60 days is 0.27 and 0.09 respectively and mean probability of being penalized across conditions is approximately 0.5. Moving a hospital from the 25th to the 75th percentile of readmission rate increases the probability of being penalized by 0.9.

Figure 7: Probability of admission



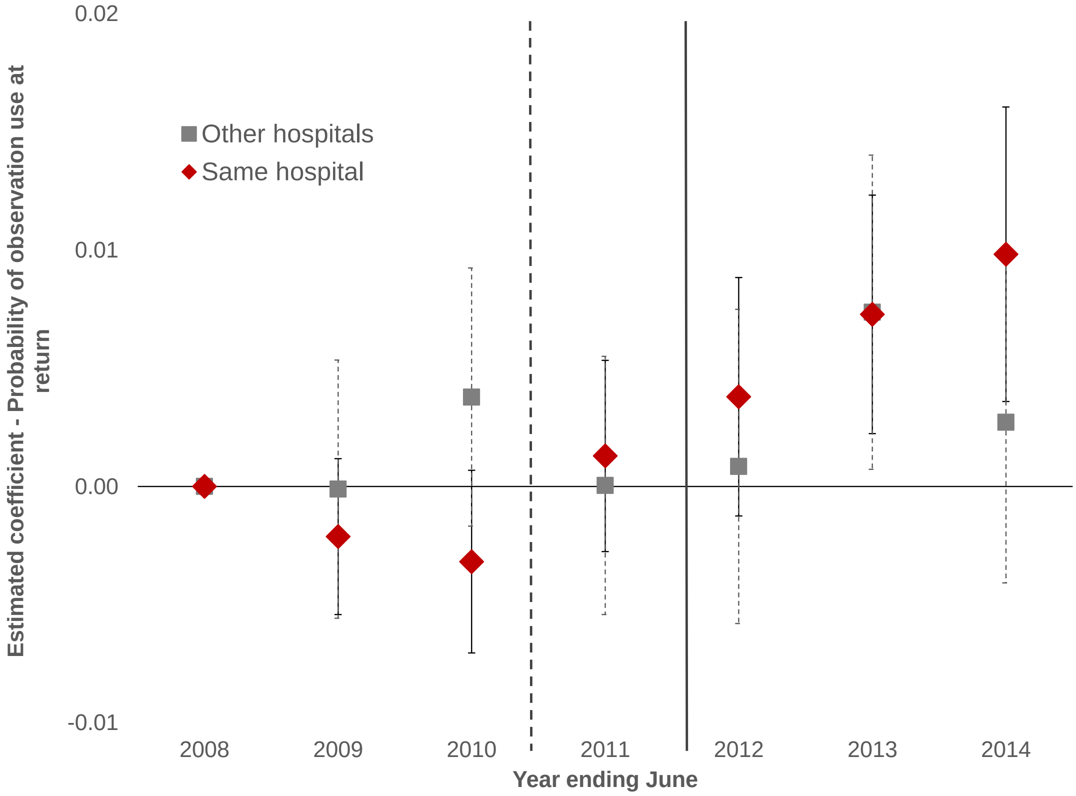
Note: This figure presents the aggregate estimated coefficients across conditions by year with 2008 being omitted. Dependent variable is probability of initial admission. Coefficients were estimated using equation 3. Standard errors are clustered by hospital. Error bars indicate 5% confidence intervals. To interpret the coefficients, note that the mean probability of admission in the pre-period is 0.89 and mean probability of being penalized across conditions is approximately 0.5. Moving a hospital from the 25th to the 75th percentile of readmission rate increases the probability of being penalized by 0.9.

Figure 8: Probability of readmission conditional on return



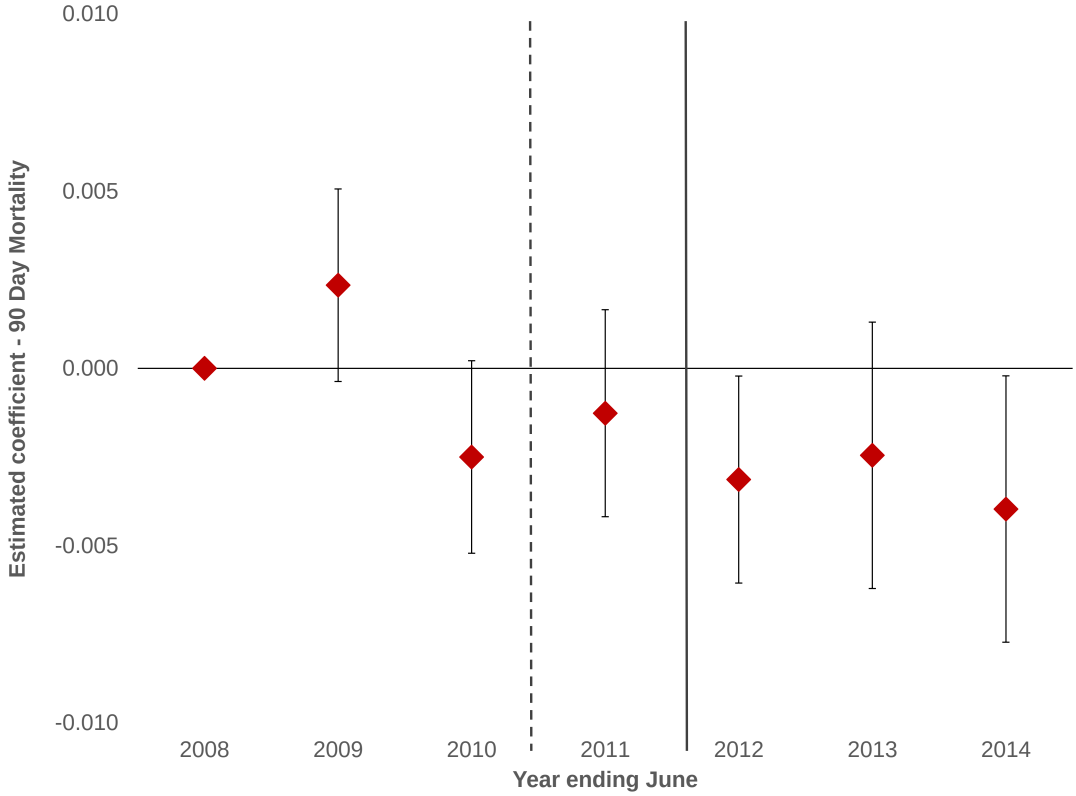
Note: This figure presents the aggregate estimated coefficients across conditions by year with 2008 being omitted. Dependent variable is probability of readmission to any hospital. Coefficients were estimated using equation 3. Standard errors are clustered by hospital. Error bars indicate 5% confidence intervals. To interpret the coefficients, note that the mean probability of readmission conditioning on return (Pre-HRRP) is 0.69 and 0.74 for own returning patients and other returning patients respectively and mean probability of being penalized across conditions is approximately 0.5. Moving a hospital from the 25th to the 75th percentile of readmission rate increases the probability of being penalized by 0.9.

Figure 9: Use of observation status



Note: This figure presents the aggregate estimated coefficients across conditions by year with 2008 being omitted. Dependent variable is probability of use of observation status upon return for patients returning to the same and other hospitals respectively. Coefficients were estimated using equation 3. Standard errors are clustered by hospital. Error bars indicate 5% confidence intervals. To interpret the coefficients, note that the mean probability of using observation status at return (Pre-HRRP) is 0.06 and 0.04 at the same and other hospitals respectively. Mean probability of being penalized across conditions is approximately 0.5. Moving a hospital from the 25th to the 75th percentile of readmission rate increases the probability of being penalized by 0.9.

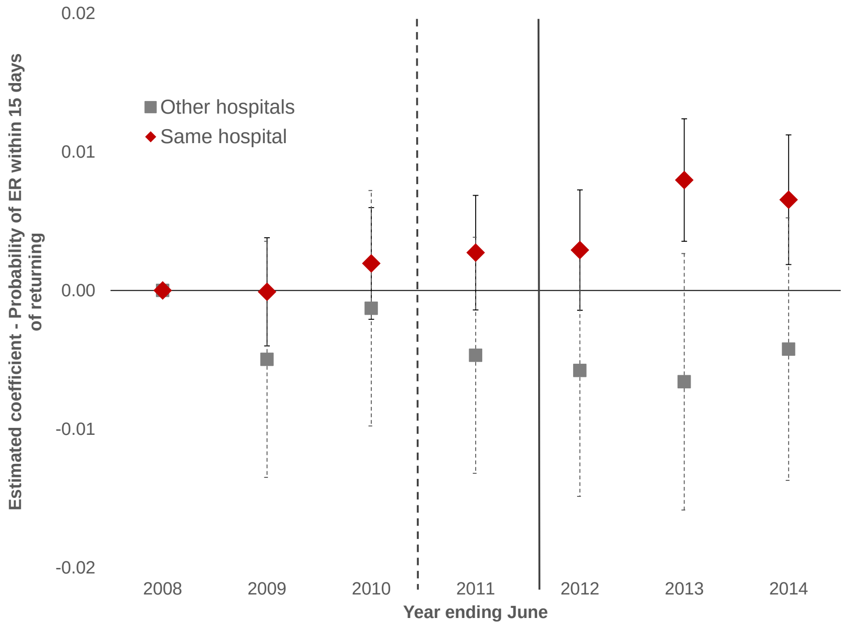
Figure 10: 90 Day mortality



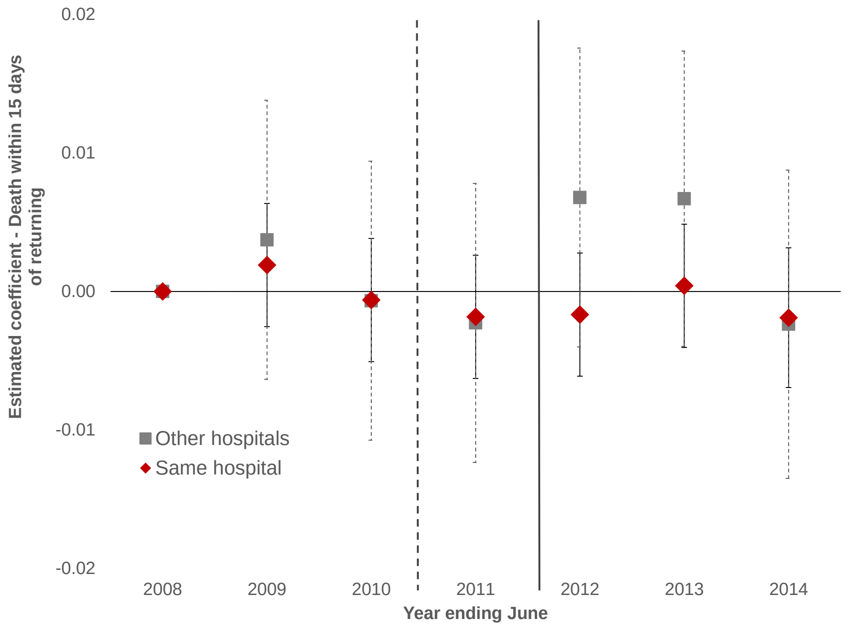
Note: This figure presents the aggregate estimated coefficients across conditions by year with 2008 being omitted. Dependent variable is probability of death within 90 days of discharging from index case. Coefficients were estimated using equation 3. Standard errors are clustered by hospital. Error bars indicate 5% confidence intervals. To interpret the coefficients, note that the mean probability of mortality at 90 days (pre-HRRP) is 0.19 and mean probability of being penalized across conditions is approximately 0.5. Moving a hospital from the 25th to the 75th percentile of readmission rate increases the probability of being penalized by 0.9.

Figure 11: Adverse outcomes for returning patients

(a) ER visit within 15 days of return



(b) Mortality within 15 days of return



Note: This figure presents the aggregate estimated coefficients across conditions by year with 2008 being omitted. Dependent variable is probability of ER visit (Panel A) and Death (Panel B) within 15 days of returning to the ER after the index stay. Coefficients were estimated using equation 3. Standard errors are clustered by hospital. Error bars indicate 5% confidence intervals. To interpret the coefficients, note that the mean probability of outcomes in the pre-period are 0.08 and 0.15 respectively and mean probability of being penalized across conditions is approximately 0.5. Moving a hospital from the 25th to the 75th percentile of readmission rate increases the probability of being penalized by 0.9.

Table 1: Descriptive Statistics

	Heart Attack	Heart Failure	Pneumonia
Panel A: Number of observations			
Index cases	1,234,894	3,140,914	2,499,537
Panel B: Key outcomes (09-11)			
P(Readmission)	0.181	0.220	0.173
	(0.057)	(0.035)	(0.035)
P(Admission)	0.990	0.880	0.840
	(0.02)	(0.089)	(0.089)
P(Return)	0.254	0.288	0.241
	(0.055)	(0.036)	(0.037)
P(Readmission | Return)	0.645	0.692	0.657
	(0.107)	(0.083)	(0.09)
Panel C: Quality (09-11)			
Mortality			
In-hospital	0.076	0.030	0.043
	(0.025)	(0.012)	(0.018)
30 Day	0.143	0.107	0.119
	(0.042)	(0.023)	(0.027)
90 Day	0.193	0.194	0.187
	(0.055)	(0.031)	(.034)
Process of Care	94.214	88.682	91.026
	(6.782)	(14.603)	(7.926)
Panel D: Penalty incentive IQR			
Penalty likelihood, P_{h2013}	0.889	0.969	0.938
Penalty rate, p_{h2013}	0.055	0.049	0.051

Note: This table presents descriptive statistics from the main data sample used in the analysis. Panel A presents the number of index cases and returns within 30 days over 2007-14. Index case is the unit of analysis in the paper and refers to the initial episode of care associated with one of these conditions that is subject to the readmission penalty. Returns pertains to all patients that were readmitted or returned to the Emergency Department but not readmitted, within 30 days of discharge from the index case. Panel B presents mean and standard deviations across hospitals over 2009-11 for the key outcome variables pertaining to readmission. Panel C presents mean and standard deviations across hospitals over 2009-11 for quality metrics - mortality at different durations and process of care metrics released by CMS. Panel D aims to give a sense of the magnitude of the penalty incentive by presenting the inter-quartile range across hospitals for the the ex-ante likelihood of being penalized in the first year, P_{h2013} and penalty rate p_{h2013} applicable to admissions in the first year of the program.

Table 2: Impact on probability of Readmission

		Heart Attack	Heart Failure	Pneumonia	Aggregate
Panel A: OLS					
$P_{ht+1} \cdot 1(T=1)$	0-30 Days	-0.0430***	-0.0275***	-0.0345***	-0.033
		(0.002)	(0.001)	(0.001)	
	31-60 Days	0.0016	0.0024***	-0.0004	0.001
		(0.001)	(0.001)	-0.001	
Panel B: IV (Baseline)					
$P_{ht+1} \cdot 1(T=1)$	0-30 Days	-0.0349***	-0.0100***	-0.0281***	-0.021
		(0.006)	(0.003)	(0.003)	
	31-60 Days	0.0021	0.0000	-0.0100***	-0.003
		(0.003)	(0.002)	(0.002)	
First stage					
$\hat{Y}_{h07} \cdot 1(T=1)$		5.5116***	11.3697***	9.8632***	
		(0.289)	(0.373)	(0.357)	
F-Stat		91.4	172.5	179.74	
Panel C: IV (Demographics)					
$P_{ht+1} \cdot 1(T=1)$	0-30 Days	-0.0294***	0.0040	-0.0172***	-0.010
		(0.008)	(0.007)	(0.007)	
	31-60 Days	-0.0038	0.0004	-0.0099**	-0.004
		(0.005)	(0.004)	(0.005)	
First stage					
$\hat{r}_{h07} \cdot 1(T=1)$		4.9264***	5.622***	6.3246***	
		(0.498)	(0.539)	(0.521)	
F-Stat		66.56	118.91	137.7	
Observations		900,399	2,276,911	1,778,537	-
Y Mean (0-30 days)		0.17	0.22	0.18	0.20
Y Mean (31 - 60 days)		0.06	0.07	0.07	0.07

Note: This table presents regression coefficients obtained by estimating Equations 1 (Panel A) and 2 (Panels B and C) respectively on case level data as described in Section 4. The dependent variable is probability of readmission. Panel B presents results using predicted readmission rate $\hat{y}_{h07}$ as the instrument while Panel C presents corresponding results using risk adjusted readmission rate predicted by demographics, $\hat{r}_{h07}$. These panels also present the first stage estimates and corresponding F-test statistics. Estimates are obtained for each condition (Heart Attack, Heart Failure and Pneumonia) independently. Column 6 presents the weighted average estimate across conditions. Weights are in proportion to number of patients - approximately as 18%, 46% and 36% respectively. P_{ht+1} denotes the ex-ante probability of hospital h being penalized in year $t+1$ calculated in year t. $T=1$ denotes the period post the implementation of the penalty i.e. 2012 onwards. All models include a vector case-mix risk factors, hospital and year fixed effects. Standard errors are clustered by hospital. To interpret coefficients note that mean probability of being penalized across conditions is approximately 0.5. Moving a hospital from the 25th to the 75th percentile of readmission rate increases the probability of being penalized by 0.9.

Table 3: Impact on probability of Return

		Heart Attack	Heart Failure	Pneumonia	Aggregate
Panel A: OLS					
$P_{ht+1} \cdot 1(T=1)$	0-30 Days	-0.0368***	-0.0211***	-0.0328***	-0.028
		(0.002)	(0.002)	(0.002)	
	31-60 Days	0.0043***	0.0037***	0.0001	0.003
		(0.001)	(0.001)	(0.001)	
Panel B: IV (Baseline)					
$P_{ht+1} \cdot 1(T=1)$	0-30 Days	-0.0286***	-0.0061*	-0.0268***	-0.018
		(0.006)	(0.003)	(0.004)	
	31-60 Days	0.0044	0.0016	-0.0053**	0.000
		(0.003)	(0.002)	(0.002)	
Panel C: IV (Demographics)					
$P_{ht+1} \cdot 1(T=1)$	0-30 Days	-0.0236***	0.0153**	-0.0102	-0.001
		(0.009)	(0.008)	(0.008)	
	31-60 Days	-0.0034	-0.0022	-0.0093*	-0.005
		(0.006)	(0.004)	(0.005)	
Observations		900,399	2,276,911	1,778,537	-
Y Mean (0-30 days)		0.25	0.28	0.24	0.26
Y Mean (31 - 60 days)		0.08	0.09	0.09	0.09

Note: This table presents regression coefficients obtained by estimating Equations 1 (Panel A) and 2 (Panels B and C) respectively as described in Section 4 on case level data. The dependent variable is an indicator for return. Panel B presents results using predicted readmission rate $\hat{y}_{h07}$ as the instrument while Panel C presents corresponding results using risk adjusted readmission rate predicted by demographics, $\hat{r}_{h07}$. Estimates are obtained for each condition (Heart Attack, Heart Failure and Pneumonia) independently. Column 6 presents the weighted average estimate across conditions. Weights are in proportion to number of patients - approximately as 18%, 46% and 36% respectively. P_{ht+1} denotes the ex-ante probability of hospital h being penalized in year $t+1$ calculated in year t. $T=1$ denotes the period post the implementation of the penalty i.e. 2012 onwards. All models control for time varying hospital case-mix characteristics, hospital and year fixed effects. Standard errors are clustered by hospital. To interpret coefficients note that mean probability of being penalized across conditions is approximately 0.5. Moving a hospital from the 25th to the 75th percentile of readmission rate increases the probability of being penalized by 0.9.

Table 4: Impact on probability of initial admission

	Heart Attack	Heart Failure	Pneumonia	Aggregate
Panel A: OLS				
$p_{ht} \cdot 1(T = 1)$	-0.0018	-0.0777***	-0.0281	-0.047
	(0.005)	(0.017)	(0.020)	
Panel B: IV (Baseline)				
$p_{ht} \cdot 1(T = 1)$	-0.1106***	-0.1872***	-0.0877	-0.137
	(0.038)	(0.048)	(0.057)	
Panel C: IV (Demographics)				
$p_{ht} \cdot 1(T = 1)$	-0.1862***	-0.4836***	-0.3939***	-0.401
	(0.061)	(0.098)	(0.107)	
Observations	994,504	2,727,984	2,272,736	-
Y Mean	0.99	0.88	0.84	0.88
p_{ht} I.Q.R.	0.06	0.05	0.05	0.05

Note: This table presents regression coefficients obtained by estimating Equations 1 (Panel A) and 2 (Panels B and C) respectively as described in Section 4 on case level data. The dependent variable is probability of initial admission. Panel B presents results using predicted readmission rate $\hat{y}_{h07}$ as the instrument while Panel C presents corresponding results using risk-adjusted rate predicted by demographics, $\hat{r}_{h07}$. Estimates are obtained for each condition (Heart Attack, Heart Failure and Pneumonia) independently. Column 5 presents the weighted average estimate across conditions. Weights are in proportion to number of (denominator) patients - approximately as 16%, 46% and 38% respectively. P_{ht+1} denotes the ex-ante probability of hospital h being penalized in year $t+1$ calculated in year t. $T = 1$ denotes the period post the implementation of the penalty i.e. 2012 onwards. All models control for a vector of case-mix risk factors, hospital and year fixed effects. Standard errors are clustered by hospital. To interpret coefficients note that mean probability of being penalized across conditions is approximately 0.5. Moving a hospital from the 25th to the 75th percentile of readmission rate increases the probability of being penalized by 0.9.

Table 5: Selection based on predicted readmission risk

	Heart Attack		Heart Failure		Pneumonia	
	(1)	(2)	(3)	(4)	(5)	(6)
$z_{h07} \cdot 1(T = 1)$	-0.0643	-0.1564**	-0.4748***	-1.5904***	-0.2508	-0.9487***
	(0.042)	(0.071)	(0.181)	(0.208)	(0.194)	(0.225)
$z_{h07} \cdot \hat{P}(r = 1)$	2.5710***	3.7601***	-16.2676***	7.4727***	-17.1592***	-12.1333***
	(0.312)	(0.543)	(1.274)	(1.660)	(1.422)	(1.716)
$1(T = 1) \cdot \hat{P}(r = 1)$	0.0329***	0.0230***	0.0892***	0.0520***	-0.0136	-0.0240
	(0.005)	(0.005)	(0.016)	(0.016)	(0.020)	(0.020)
$z_{h07} \cdot 1(T = 1) \cdot \hat{P}(r = 1)$	-0.1263	0.2167	1.6160**	5.9903***	1.5800	4.6757***
	(0.213)	(0.380)	(0.783)	(0.896)	(1.007)	(1.156)
Observations	994,504	994,504	2,727,984	2,727,984	2,272,736	2,272,736
Y Mean	0.99	0.99	0.879	0.879	0.837	0.837
X IQR	0.11	0.08	0.07	0.06	0.08	0.07

Note: This table presents regression coefficients obtained by estimating Equation 5 using either the predicted readmission rate $\hat{y}_{h07}$ (Columns 1,3 and 5) or the risk adjusted rate predicted by demographics, $\hat{r}_{h07}$ (Columns 2,4 and 6) as the source of variation across hospitals. Each observation is a case of potential admission i.e. including arrivals at ED that were not admitted. Estimates are obtained for each condition (Heart Attack, Heart Failure and Pneumonia) independently. $T = 1$ denotes the period post the implementation of the penalty i.e. 2012 onwards. $\hat{P}(r = 1)$ is a predicted probability of readmission for the patient based on parameters estimated on 2007 data. The last row presents the DDD estimator. All models also include $\hat{P}(r = 1)$, hospital and year fixed effects. Standard errors are clustered by hospital. To interpret coefficients note that the inter-quartile range of z across conditions is approximately 0.08.

Table 6: Impact on probability of Readmission for returning patients

			Heart Attack	Heart Failure	Pneumonia	Aggregate
Panel A: OLS						
$P_{ht+1} \cdot 1(T=1)$	Own patients	0-30 Days	-0.0538***	-0.0371***	-0.0416***	-0.042
			(0.006)	(0.003)	(0.004)	
	Other patients	0-30 Days	-0.0333***	-0.0307***	-0.0191**	-0.027
			(0.010)	(0.006)	(0.008)	
Panel B: IV (Baseline)						
$P_{ht+1} \cdot 1(T=1)$	Own patients	0-30 Days	-0.0384***	-0.0225***	-0.0488***	-0.035
			(0.014)	(0.007)	(0.009)	
	Other patients	0-30 Days	0.0082	-0.0158	0.0226	0.002
			(0.017)	(0.012)	(0.018)	
Panel C: IV (Demographics)						
$P_{ht+1} \cdot 1(T=1)$	Own patients	0-30 Days	-0.0336	-0.0374**	-0.0595***	-0.045
			(0.021)	(0.016)	(0.019)	
	Other patients	0-30 Days	0.0534	-0.0142	0.0210	0.011
			(0.037)	(0.025)	(0.036)	
Own patients:						
Observations			173,114	588,394	392,173	-
Y Mean			0.68	0.72	0.67	0.69
Other patients:						
Observations			54,917	134,184	80,602	-
Y Mean			0.80	0.75	0.71	0.74

Note: This table presents regression coefficients obtained by estimating Equations 1 (Panel A) and 2 (Panels B and C) respectively as described in Section 4 on case level data. The dependent variable is an indicator for readmission conditioning on return for two different groups of patients - those returning to the same hospital ("Own", approximately 80%) and those returning to a different hospital ("Other"). Estimation is done from the perspective of the initial hospital. Panel B presents results using predicted readmission rate $\hat{y}_{h07}$ as the instrument while Panel C presents corresponding results using risk adjusted readmission rate predicted by demographics, $\hat{r}_{h07}$. Estimates are obtained for each condition (Heart Attack, Heart Failure and Pneumonia) independently. Column 6 presents the weighted average estimate across conditions. Weights are in proportion to number of patients - approximately as 18%, 46% and 36% respectively. P_{ht+1} denotes the ex-ante probability of hospital h being penalized in year $t+1$ calculated in year t. $T = 1$ denotes the period post the implementation of the penalty i.e. 2012 onwards. All models control for case level risk factors, hospital and year fixed effects. Standard errors are clustered by hospital. To interpret coefficients note that mean probability of being penalized across conditions is approximately 0.5. Moving a hospital from the 25th to the 75th percentile of readmission rate increases the probability of being penalized by 0.9.

Table 7: Impact on use of observation status

			Heart Attack	Heart Failure	Pneumonia	Aggregate
Panel A: OLS						
$P_{ht+1} \cdot 1(T=1)$	Own patients	0-30 Days	0.0198***	0.0109***	0.0140***	0.014
			(0.005)	(0.003)	(0.003)	
	Other patients	0-30 Days	-0.0157	0.0111***	0.0089**	0.005
			(0.013)	(0.003)	(0.004)	
Panel B: IV (Baseline)						
$P_{ht+1} \cdot 1(T=1)$	Own patients	0-30 Days	0.0170*	0.0203***	0.0211***	0.020
			(0.010)	(0.006)	(0.007)	
	Other patients	0-30 Days	-0.0157	0.0070	0.0191**	0.007
			(0.013)	(0.007)	(0.010)	
Panel C: IV (Demographics)						
$P_{ht+1} \cdot 1(T=1)$	Own patients	0-30 Days	-0.0011	-0.0079	0.0045	-0.002
			(0.017)	(0.013)	(0.012)	
	Other patients	0-30 Days	-0.0007	0.0024	0.0217	0.009
			(0.021)	(0.015)	(0.019)	
Own patients:						
Observations			174,725	589,661	393,872	-
Y Mean			0.08	0.06	0.06	0.06
Other patients:						
Observations			55,796	135,067	81,589	-
Y Mean			0.04	0.04	0.04	0.04

Note: This table presents regression coefficients obtained by estimating Equations 1 (Panel A) and 2 (Panels B and C) respectively as described in Section 4 on case level data. The dependent variable is an indicator for use of observation status. Panel B presents results using predicted readmission rate $\hat{y}_{h07}$ as the instrument while Panel C presents corresponding results using risk adjusted readmission rate predicted by demographics, $\hat{r}_{h07}$. Estimates are obtained for each condition (Heart Attack, Heart Failure and Pneumonia) independently. Column 6 presents the weighted average estimate across conditions. Weights are in proportion to number of patients - approximately as 18%, 46% and 36% respectively. P_{ht+1} denotes the ex-ante probability of hospital h being penalized in year $t+1$ calculated in year t. $T = 1$ denotes the period post the implementation of the penalty i.e. 2012 onwards. All models control for case-mix characteristics, hospital and year fixed effects. Standard errors are clustered by hospital. To interpret coefficients note that mean probability of being penalized across conditions is approximately 0.5. Moving a hospital from the 25th to the 75th percentile of readmission rate increases the probability of being penalized by 0.9.

Table 8: Contribution of different channels

	Heart Attack	Heart Failure	Pneumonia	Aggregate
Panel A: Mean values pre-penalty				
P(Return)	0.25	0.28	0.24	0.261
P(Readmit | Return)	0.707	0.778	0.736	0.75
P(Readmission)	0.17	0.22	0.18	0.20
Panel B: Response scenarios				
1. Both channels				
P(Return)	0.22	0.28	0.21	0.243
P(Readmit | Return)	0.68	0.76	0.70	0.72
P(Readmission)	0.147	0.210	0.148	0.175
2. Only Treatment quality				
Healthiest patients				
P(Readmission)	0.156	0.217	0.157	0.184
Share of total decrease	69%	36%	67%	57%
Average patients				
P(Readmission)	0.154	0.216	0.156	0.183
Share of total decrease	76%	43%	71%	62%
3. Only Selective readmission				
P(Readmission)	0.167	0.215	0.167	0.188
Share of total decrease	27%	58%	32%	39%

Note: The table presents the contribution of two channels (return, selective readmission) that produce the total decrease in probability of readmission for each of the three penalized conditions and in aggregate. The initial admission channel is not considered here. Panel A presents the mean value over 2007-11 for probability of return, readmission conditional on return and of (unconditional) readmission. Panel B presents the estimated probability of readmission incorporating the estimated effects of one or both channels. It also presents the share of total decrease that is due to the each channel. These estimates are obtained by simple back-of-the-envelope calculation using the obtained estimates for each of these outcomes.

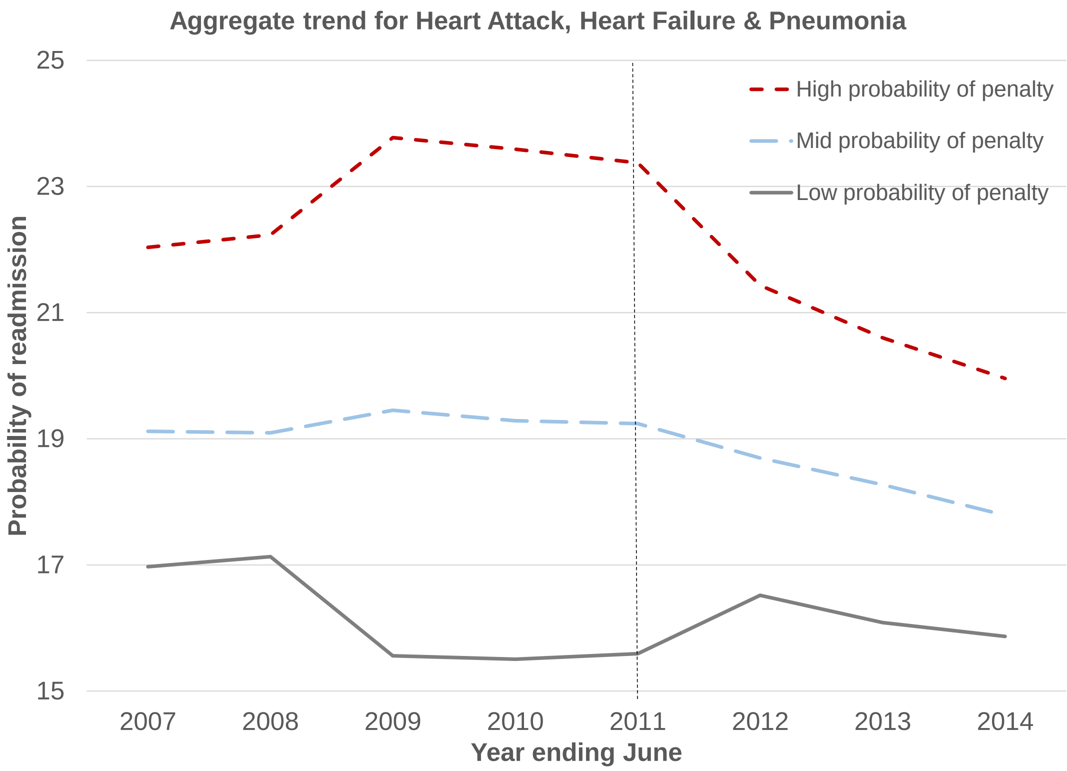
Table 9: Impact on mortality

		Heart Attack	Heart Failure	Pneumonia	Aggregate
Panel A: OLS					
$P_{ht+1} \cdot 1(T=1)$	7 Day	-0.0003	-0.0008	-0.0002	0.000
		(0.002)	(0.001)	(0.001)	
	30 Day	-0.0033	-0.0026**	-0.0052***	-0.004
		(0.002)	(0.001)	(0.001)	
	90 Day	-0.0096***	-0.0045***	-0.0089***	-0.007
		(0.003)	(0.001)	(0.002)	
Panel B: IV (Baseline)					
$P_{ht+1} \cdot 1(T=1)$	7 Day	-0.0076*	-0.0025*	-0.0002	-0.003
		(0.004)	(0.001)	(0.002)	
	30 Day	-0.0093*	-0.0035*	-0.0021	-0.004
		(0.005)	(0.002)	(0.003)	
	90 Day	-0.0191***	-0.0014	-0.0042	-0.006
		(0.006)	(0.003)	(0.003)	
Panel C: IV (Demographics)					
$P_{ht+1} \cdot 1(T=1)$	7 Day	-0.0181**	-0.0087**	-0.0026	-0.008
		(0.008)	(0.003)	(0.004)	
	30 Day	-0.0217**	-0.0052	0.0005	-0.006
		(0.009)	(0.005)	(0.006)	
	90 Day	-0.0323***	0.0022	0.0055	-0.003
		(0.010)	(0.006)	(0.007)	
Observations		994,504	2,727,898	2,273,045	-
Y Mean (7 Day)		0.11	0.06	0.07	0.07
Y Mean (30 Day)		0.14	0.11	0.12	0.12
Y Mean (90 Day)		0.19	0.19	0.19	0.19

Note: This table presents regression coefficients obtained by estimating Equations 1 (Panel A) and 2 (Panels B and C) respectively as described in Section 4 on case level data. The dependent variable is an indicator for death within 7,30 or 90 days post discharge from the initial admission. Panel B presents results using predicted readmission rate $\hat{y}_{h07}$ as the instrument while Panel C presents corresponding results using risk adjusted rate predicted by demographics, $\hat{r}_{h07}$. Estimates are obtained for each condition (Heart Attack, Heart Failure and Pneumonia) independently. Column 6 presents the weighted average estimate across conditions. Weights are in proportion to number of patients - approximately as 18%, 46% and 36% respectively. P_{ht+1} denotes the ex-ante probability of hospital h being penalized in year $t+1$ calculated in year t. $T = 1$ denotes the period post the implementation of the penalty i.e. 2012 onwards. All models control for time varying hospital case-mix characteristics, hospital and year fixed effects. Standard errors are clustered by hospital. To interpret coefficients note that mean probability of being penalized across conditions is approximately 0.5. Moving a hospital from the 25th to the 75th percentile of readmission rate increases the probability of being penalized by 0.9.

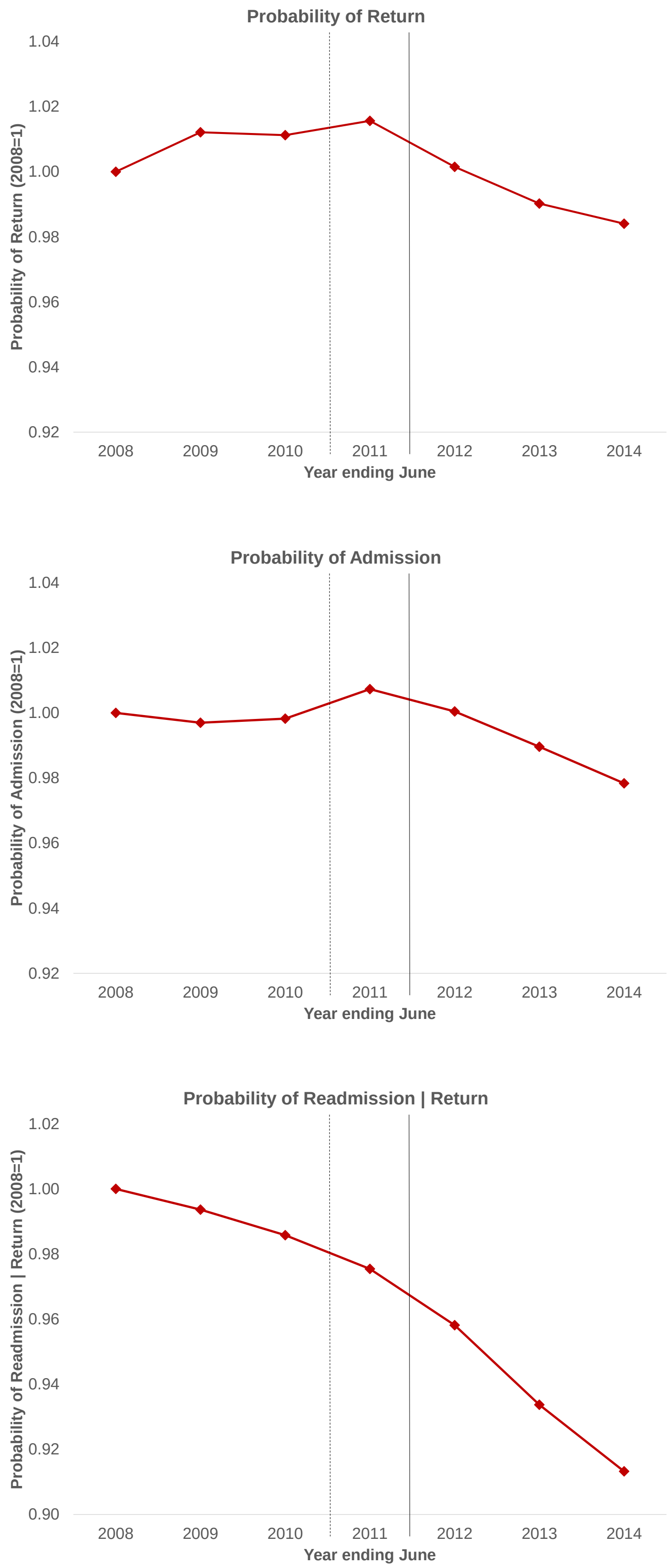
Appendix figures and tables

Figure A.1: Trend in probability of readmission



Note: This figure plots the raw time series of probability of readmission across the three penalized conditions. Hospitals are divided into three equal sized buckets based on their probability of being penalized in 2012 based on readmission performance over 2009-11. The vertical black line indicates that HRRP rules were announced at the end of 2011. Note the sharp mean reversion trend for both high and low probability hospitals.

Figure A.2: Trends in other outcomes



Note: This figure plots the time series of mean probability of return, initial admission and readmission conditional on return, for Heart Attack, Heart Failure and Pneumonia. The values are normalized against the value in 2008 for each time series. The Affordable Care Act was enacted in 2010 and penalty details were announced in August 2011.

Table A.1: Hospitals

Owner type	# hospitals	Medicare revenue ($ Mn)	Medicare patients/year	Beds	Index patients/year	P(Readmission) in 30 days	Penalty rate		
							Heart Attack	Heart Failure	Pneumonia
For-profit	746	18	2,238	155	278	0.19	0.04	0.04	0.04
City/County	478	19	2,203	181	274	0.20	0.03	0.03	0.03
Non-profit	1,978	34	3,940	248	423	0.19	0.04	0.03	0.03
State	51	40	3,062	381	291	0.20	0.05	0.03	0.04
Aggregate	3,253	28	3,281	219	366	0.19	0.03	0.03	0.03

Note: This table presents basic descriptive statistics on the hospitals that appear in the Medicare claims analysis sample, by type of ownership. All values pertain to the pre-penalty period i.e. July 2006-June 2011. Medicare revenue and number of patients are mean values for a particular category of hospitals per year. The probability of readmission is averaged across the three target conditions. The penalty rate corresponds to the object p_{h2013} discussed in the text and can be interpreted as the proportion of annual condition revenue that will be clawed back as penalty in the first year of the program. It is averaged over all hospitals, including the approximately 50% hospitals with no penalty.

Table A.2: Constructing predicted readmission rate, $\hat{r}_h$

Parameters	Heart Attack	Heart Failure	Pneumonia
Frac. Dual	-0.0489	0.0916	-0.1082*
	(0.110)	(0.066)	(0.057)
(Frac. Dual)^2	0.0382	0.0051	0.0800***
	(0.055)	(0.026)	(0.026)
Frac. White	-0.2340	-0.0202	-0.0441
	(0.152)	(0.078)	(0.065)
(Frac. White)^2	0.1519*	0.0385	-0.0032
	(0.082)	(0.037)	(0.034)
Frac. Black	0.0238	0.0476	-0.0186
	(0.103)	(0.063)	(0.058)
(Frac. Black)^2	-0.1484*	-0.0306	-0.0105
	(0.088)	(0.039)	(0.040)
Frac. Hispanic	-0.0306	-0.0770	-0.0627
	(0.102)	(0.049)	(0.048)
(Frac. Hisp.)^2	0.0832	0.1180	0.1132
	(0.207)	(0.097)	(0.092)
Frac. Native	0.1873**	0.0167	0.0170
	(0.084)	(0.036)	(0.025)
Frac. Dual · White	0.2857***	0.0610	0.1695***
	(0.106)	(0.062)	(0.051)
Frac. Dual · Black	0.2539**	-0.0008	0.1504**
	(0.118)	(0.066)	(0.061)
Observations	2,918	2,991	3,006
Adj R-squared	0.238	0.320	0.308

Note: This table presents regression coefficients estimated using OLS. The dependent variable is the risk adjusted readmission rate, denoted as r_h in the text. "Frac." stands for fraction, for example Frac. dual indicates the fraction of patients in the potential admissions of a hospital in 2007-08 that were Medicaid eligible. All regressions include HRR fixed effects. These coefficients are then used to predict the instrument $\hat{r}_h$ as described in section 4.

Table A.3: Impact on Process of care measures

	Heart Attack	Heart Failure	Pneumonia	Aggregate
Panel A: OLS				
$P_{ht+1} \cdot 1(T=1)$	0.0390*	0.0183	0.107***	0.054
	(0.0233)	(0.0177)	(0.0249)	
Panel B: IV (Baseline)				
$P_{ht+1} \cdot 1(T=1)$	0.0863	-0.118**	0.138**	0.012
	(0.0620)	(0.0468)	(0.0658)	
Panel C: IV (Demographics)				
$P_{ht+1} \cdot 1(T=1)$	0.192*	-0.322***	0.0863	-0.078
	(0.113)	(0.110)	(0.142)	
Observations	14,213	22,441	22,546	-

Note: This table presents regression coefficients obtained by estimating Equations 1 (Panel A) and 2 (Panels B and C) respectively as described in Section 4 on hospital-year level data. The dependent variable is hospital score on process of care measures, standardized to have mean zero and S.D. one in each year for each condition. Panel B presents results using predicted readmission rate $\hat{y}_{h07}$ as the instrument while Panel C presents corresponding results using risk adjusted readmission rate predicted by demographics, $\hat{r}_{h07}$. Estimates are obtained for each condition (Heart Attack, Heart Failure and Pneumonia) independently. Column 6 presents the weighted average estimate across conditions. Weights are in proportion to number of patients - approximately as 18%, 46% and 36% respectively. P_{ht+1} denotes the ex-ante probability of hospital h being penalized in year $t+1$ calculated in year t. $T=1$ denotes the period post the implementation of the penalty i.e. 2012 onwards. All models control for time varying hospital case-mix characteristics, hospital and year fixed effects. Standard errors are clustered by hospital. To interpret coefficients note that moving a hospital from the 25th to the 75th percentile of readmission rate increases the probability of being penalized by 0.9.

Appendices

A Data construction

This section describes details of the procedure used to construct the sample from raw Medicare claims data, as well as the construction of key variables used in the analysis. I obtained access to Medicare claims records for calendar years 2006-2014 for the universe of fee-for-service Medicare patients. Since CMS uses July-June periods in its analysis to penalize hospitals, I also organize my analysis around this cycle rather than calendar years. Hence I have access to eight complete years from July 2006 through June 2014.

The Medicare claims data is organized around health care interactions so that each observation represents an inpatient stay, outpatient visit, emergency department visit, doctor's visit, etc. In each observation I observe a rich vector of patient co-morbidities, the principal diagnosis for the visit or stay, what procedures were performed and the actual dates of the event. I can follow patients over time as well as identify the provider (physician and/or hospital) for each health care interaction. Separately, for each patient I observe limited demographic information (gender, age and race), Medicaid eligibility and mortality status at least one year following June 2014.

A.1 Sample selection

The penalty applies to all general acute care hospitals that accept Medicare patients and are paid under the inpatient prospective payment system (IPPS).[26] CMS excludes a special class of rural hospitals called "Critical Access Hospitals" from all performance pay schemes.[27] Correspondingly I also exclude them from analysis. These are also not useful as controls since they are very small (<25 beds) and isolated (at least 35 miles to the next hospital) by definition.

I apply two minor restrictions to decide which hospitals are included in my analysis. First, I exclude Veterans Affairs hospitals. These are federally owned and operated hospitals that were initially exempt from the penalty but subsequently included in 2014. Second, consistent with CMS as well as prior studies that have used readmission as a quality measure (Chandra et al., 2016) I exclude small hospitals (less than 50 admissions over 2009-11) from the main analysis.[28] My final sample contains approximately 3,250 out of the 3,334 acute care hospitals that participate in IPPS (MEDPAC, 2013). Table A.1 presents more details on the characteristics of hospitals in the sample.

In addition to care at the hospital, patients receive care in the community through their primary care physician (PCP) and specialists. This includes diagnostic imaging, radiology and consultation visits. This data is organized in a different fashion and requires some modifications to be incorporated into the analysis.[29] This so-called "Part B" (named after Medicare Part B which provides the insurance coverage for these services) data is useful in testing for increase in coordination between hospitals and physicians.

I make two sample restrictions to focus on the relevant observations. First the 'type of service' should be coded as medical care, consultation, diagnostic imaging, lab work or therapeutic radiology.

[26] This implies that other types of hospitals (psychiatric, rehabilitation and long-term care) or states exempted from IPPS (Maryland) are explicitly excluded. In addition, I limit attention to hospitals within the continental US (excluding Alaska, Hawaii and territories).

[27] MEDPAC (2013) reports that there were 1,307 CAH hospitals in 2011 and although they account for 28% of all hospitals they served only 3.6% of Medicare hospital discharges.

[28] Readmission rates tend to be very noisy for small hospitals, particularly with less than a hundred admissions. CMS uses a lower cutoff of 25 admissions. The results are not sensitive to the exact cutoff value used ranging from 25 to 100.

[29] These claims are recorded in so-called "Carrier files" or Part B files. Each observation is not a separate episode of care but a separate service within each episode. For example, if you go to the doctor and get a blood test and an MRI then it will generate three observations - one for the consultation, and two more for the other services. These three observations will be identified by the same claim number. I collapse Part B observations to the claim number level to make it similar in structure as the hospital claims.

This excludes services like ambulance transport. Second, the 'place of service' must be doctor's office, outpatient, hospice, home health or skilled nursing facility. These two restrictions limit the data to 70% of all Part B claims. Further I exclude Part B claims that are concurrent with a hospital stay or outpatient visit and focus on standalone visits to doctor's offices.[30] This ensures the data provides a view on trends in care outside the hospital.

A.2 Key variables

Starting with the raw data files, I construct condition specific cohorts. For example, all cases admitted to hospitals with heart attack as the principal diagnosis during the analysis period form the heart attack cohort. The same patient may be present in two different cohorts if she was admitted separately with each of the conditions as the principal diagnosis. From hereon all procedures are applied in parallel across cohorts.

A.2.1 Index cases and readmissions

I first identify the 'initial' episode of care. As discussed in the paper, these are called 'Index' cases. The key conditions for a hospital stay to be an index case are that it should begin at least thirty days after a prior discharge for the same condition and that the patient should be discharged alive.[31]

There are two special situations that warrant further explanation. First, an index case cannot be a readmission following a hospital stay for the same condition, but it can be a readmission following another condition. For example, if the first hospitalization observed for a patient is for heart failure and ends on September 20^{th} and the next one is for Pneumonia beginning on October 5^{th} then both are considered index cases for the two conditions respectively. The pneumonia admission is also considered a readmission for the heart failure case. If instead of pneumonia, the patient was readmitted for heart failure again, that would not qualify as a separate index case.

Second, multiple readmissions within thirty days do not incur additional penalty. To continue the above example, if the patient is admitted a third time on October 15^{th} it is not counted as an additional readmission. Hence CMS is not penalizing the number of readmissions, but the probability of readmission. Finally, I cannot identify index cases in January 2006 since I do not observe hospitalizations in December 2005. Hence I flag index cases and readmissions starting in February 2006.

Once an index case is identified, I follow the patient for the next 60 days to identify readmission to any hospital for any reason[32] within thirty days of discharge. To be consistent with the penalty rules I replicate CMS procedures exactly.[33]

[30] When a patient is hospitalized it generates two types of claims. An inpatient claim is generated under Part A for the payment made to the hospital. Separately a Part B claim is generated for the payment to the physician (Since they are typically not hospital employees). This is also the case for an ER visit. In order to ensure that I only track standalone doctors' visits I exclude any Part B claims that fall within dates of a preceding inpatient or outpatient visit. For example, consider a patient hospitalized from Jan 1-5 and then receives some follow-up imaging and consultation a week later on Jan 12. The Part B claims files will actually record observations corresponding to the Jan 1-5 period as well. I ignore these observations and only consider the care received on Jan 12.

[31] Two quantitatively minor conditions are that the patient should not have discharged 'against medical advice' or been transferred to another hospital. I deal with transfers as follows - if a patient's mode of arrival is transfer from another hospital then I combine the transfer case with the patient's previous hospitalization and treat them as a single episode. The admission date will pertain to the previous case and the discharge date will pertain to the transfer. The combined case can be an index case, it will be attributed to the second hospital. For conditions other than heart attack transfers are rare, less than 5%. About 10% of heart attack admissions result in transfers. Specific conditions sometimes have additional requirements. For example, heart attack admissions are not considered index if the patient was discharged the same day.

[32] CMS allows very few reasons for a re-hospitalization to be classified as 'planned' and exempt from the penalty. These account for less than 5% of all readmissions.

[33] I obtained the original SAS code used by CMS' contractor to identify index cases and readmissions and used to set the penalty. I then adapted it to my Medicare claims data sample and replicated it. More details on the rules are available at https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/

A.2.2 Risk adjustment

To set penalties for each condition, CMS adjusts the raw readmission rate for differences in hospital case mix. The goal is to control for variation in readmission rates due to observable patient risk factors and focus on the residual, presumably a better reflection of hospital quality. The risk adjustment procedure involves fitting a random effects logit model as shown below.

$$P(Y_{iht} = 1) = \frac{exp(\alpha_h + X'_{ih}\gamma)}{1 + exp(\alpha_h + X'_{ih}\gamma)}$$

where α_h represents the mean zero random effect assumed to be normally distributed with mean zero and variance to be estimated, and X is a vector of indicators for co-morbidities[34] for each patient. The set of co-morbidities used varies by condition and flags present as well as past complications. Y is an indicator set to one if patient i was readmitted within 30 days of discharge. All index cases over the three year evaluation period are included in the sample.

The next step is to predict the intercept (obtained using the posterior distribution of α_h) for each hospital, $\hat{\alpha}_h$ using the model estimated variance of α_h. Two values are computed for each hospital, which the algorithm calls

- Predicted readmission rate

$$\hat{Y}_h = \frac{1}{N_h} \cdot \sum_{i=1}^{N_h} logit(\hat{\alpha}_h + X'_{ih}\beta)$$

- Expected readmission rate

$$\tilde{Y}_h = \frac{1}{N_h} \cdot \sum_{i=1}^{N_h} logit(X'_{ih}\beta)$$

The predicted rate, $\hat{Y}$ incorporates unobserved hospital specific heterogeneity while the "expected" value $\tilde{Y}$ predicts a value assuming the hospital quality was at the mean (zero). Risk adjusted rate r_h for hospital h is then given as

$$r_h = \hat{Y}_h / \tilde{Y}_h$$

Conceptually the risk adjusted rate measures how large or small a hospital's readmission rate should be relative to mean when you focus only on unobserved hospital specific heterogeneity. Appendix B discusses how this value r_h is used to set the hospital's penalty.

A.2.3 Other outcomes

In addition to readmissions, I examine hospital response on admission and readmission decisions at the Emergency Department (ED). This helps disentangle patient decision to seek treatment from the hospital's decision to amit or readmit. Usage of the ED is inferred through the corresponding

`AcuteInpatientPPS/Readmissions-Reduction-Program.html`

[34] For example, the co-morbidities used in case of Heart Attack patients are age above 65, gender, history of PTCA, history of CABG, AMI, other location of myocardial infarction, history of infection, metastatic cancer and leukemia, Cancer, Diabetes Mellitus and complications, protein-calorie malnutrition, disorders of fluid/electrolyte/acid-base, iron deficiency and other anemias and blood disease, Dementia and other specified brain disorders, hemiplegia, paraplegia, paralysis, functional disability, congestive heart failure, Acute coronary syndrome, angina pectoris, old myocardial infarction, coronary atherosclerosis, valvular and rheumatic heart disease, arrhythmias, stroke, cerebrovascular disease, Chronic obstructive pulmonary disorder, asthma, pneumonia, end-stage renal disease, renal failure, other urinary tract disorders, ulcer/skin ulcer. In addition some other complications are also used unless they only occur at the time of the index admission.

revenue code.[35] This allows me to construct an indicator for hospital stays and outpatient care visits that originate in the ED.

I then focus on admission and readmission decisions at the ED. I pool hospital cases with outpatient cases that originated in the ED and would have been classified as index cases had the patient been admitted. This allows me to compute the probability of admission, i.e. proportion of patients who arrived at the hospital and were admitted for inpatient treatment.[36] I refer to this as the sample of 'potential admissions'. I also create an indicator to identify that a patient returned to the hospital ED within thirty days of discharge from the index case to seek care, but was not admitted. As discussed in the paper, I refer to these as returns.

A.2.4 Process of care scores

I use 'process of care' scores released by CMS on its hospital compare website.[37] CMS tracks hospital compliance with a set of best practices for each condition. Each measure relates to some clinical intervention and has been accepted as a benchmark of good practice in medicine. A key limitation of this data is that the scores pertain to all patients and not only Medicare patients. Hence these scores offer a noisy measurement of changes in protocols for Medicare patients and will potentially under-estimate changes in treatment quality.

There is considerable variation in measures recorded for a condition over time. Since I track hospital performance over 2007-2014, I require measures that have been tracked for at least a few years and preferably all years. In all I use 5,3 and 5 measures for Heart Attack, Heart Failure and Pneumonia respectively.

I standardized the scores following Chandra et al. (2016). To reduce the possibility of measurement error I first exclude measures that were computed on less than 25 patients over the three year period of July 2008-June 2011. CMS uses the same cutoff to determine if a hospital should be exempted from the readmission penalty. If a hospital computes a raw score for a performance metric over this period and has less than 25 patients in this period then I drop this metric for this hospital for all years. I normalize raw scores to have weighted mean zero and S.D.=1 within each measure-year across hospitals. So each hospital receives a standardized score for each of its measures each year. I average the standardized scores across measures within the same condition-year for a hospital, weighted by the total number of patients comprising the measure. Following this, I collapse the data to the hospital-condition-year level. Finally I standardize the condition level score for each hospital by year to have mean zero and S.D.=1. Here I use the unweighted mean across hospitals. So each hospital has a standardized score for each condition-year. These are the scores used in the regressions.

B Readmission penalty

This section begins by describing the exact formula used by CMS to determine the readmission penalty for hospitals. It then walks the reader through a simple transformation that looks at the penalty from the perspective of a forward looking hospital. It concludes by discussing some key implications for hospital behavior.

[35] Specifically I use revenue codes 0450, 0451, 0452, 0456, 0459 and 0981 as mentioned in the CMS technical guide to their administrative data files. Available at `https://www.ccwdata.org/web/guest/technical-guidance-documentation`.

[36] Approximately 90% of patients with these three conditions are admitted to the hospital through the Emergency Department. The remaining are admitted under directions of their physician. In all my analysis I do not distinguish between patients based on their mode of entry. I assume that patients admitted under their physician's guidance are infra-marginal and would have been admitted if they arrived at the ED as well.

[37] These are also known as "Timely and effective care" measures and the data is available for download at https://data.medicare.gov/data/hospital-compare?sort=relevance&tag=timely%20and%20effective%20care.

For a hospital h at the end of year t, the penalty rate applicable to year $t+2$ is set as below (suppressing the subscript h for brevity)

$$\Delta_{t+2} = \frac{1}{B_\tau} \cdot \sum_{k=1}^{K} \max\left[0, (r_{k\tau} - 1)\right] \cdot b_{k,\tau}$$

where $\tau \in \{t-2, t\}$ is a three year evaluation period ending in year t. k denotes one of K penalized conditions. This paper studies the program in 2012-14 when $K = 3$. B is the total Medicare and b is total condition k base[38] inpatient operating reimbursement respectively received over the evaluation period. $r_{k\tau}$ denotes the risk adjusted readmission rate calculated by CMS for patients of condition k at hospital h during the evaluation period. It is normalized to have mean one across hospitals. For details on how $r_{k\tau}$ is computed, see appendix A.

This penalty rate is applied to all Medicare reimbursement received in year $t+2$. It is capped at 3% by law. Fewer than 5% of hospitals actually hit the cap in the third year of the penalty (2015) when the cap was first implemented. Hence this rate converts to dollars as follows.

$$\Delta_{t+2}(\$) = B_{t+2} \cdot \left(\frac{1}{B_\tau} \cdot \sum_{k=1}^{K} \max\left[0, (r_{k\tau} - 1)\right] \cdot b_{k,\tau} \right)$$

where B_{t+2} is the total Medicare base operating reimbursement received in year $t+2$. The total dollar value can only be calculated ex-post since it depends on revenue received in year $t+2$. Hence a forward looking hospital deciding whether to invest in improvements needs to rely on an expected value of the penalty. This motivates the following transformation.

1. Total Medicare inpatient revenue appears both in the numerator and denominator of the formula. First, convert the aggregate values over τ to annual mean values and re-write slightly differently

   $$\Delta_{t+2}(\$) = \frac{B_{t+2}}{\bar{B}_\tau} \cdot \sum_{k=1}^{K} \max\left[0, (r_{k\tau} - 1)\right] \cdot \bar{b}_{k,\tau}$$

   Empirically, $B_{t+2} \approx B_t$ i.e. for the average hospital, the Medicare revenue in a given year is very close to that received in the previous years. Hence a forward looking hospital administrator can reasonably predict the penalty burden as

   $$\Delta_{t+2}(\$) \approx \frac{B_t}{\bar{B}_\tau} \cdot \sum_{k=1}^{K} \max\left[0, (r_{k\tau_1} - 1)\right] \cdot \bar{b}_{k\tau_1}$$

   Now all these quantities are known at the end of year t. It further turns out that $B_t/\bar{B}_\tau \approx 1$. This implies that the size of a hospital's total Medicare business does not affect the penalty burden and will be a less important force in determining their response. This is an important implication since an intuitive approach could have been to exploit variation in hospitals' share of revenue received through Medicare.

2. Since the penalty for each condition is bounded to be negative this implies a hospital cannot rely on good performance on one condition to compensate for poor performance on another condition. Further, assuming improvements are costly a hospital's response will be tailored

[38] Base payments do not include reimbursements received for training of graduate students and disproportionate service to poor patients. It is about 80% of total reimbursement for the average hospital. Details available at https://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/Downloads/AcutePaymtSysfctsht.pdf

to its condition specific performance i.e. it may choose not to disturb a department that is performing relatively well and focus its attention on weak departments. It is therefore useful to think about each condition specific penalty separately i.e. consider $\Delta_{k,t+2}$ instead of the Δ_{t+2}.

3. Penalty burden for a given condition increases mechanically with the size of the penalized condition for a hospital. A hospital with a large cardiology practice (greater $\bar{b}_k$) will expect a greater penalty burden relative to another hospital with a small cardiology department. Ex-ante it is not clear whether hospitals will respond to an absolute value (i.e. a million dollar penalty spurs action irrespective of the size of the hospital) or a rate (5% spurs greater action than 1%). I normalize by the size of the condition's revenue and convert the dollar value to a penalty rate, denoting it $p_{k,t+2}$.

$$p_{k,t+2} = \frac{B_t}{\bar{B}_\tau} \max\left[0, (r_{k\tau} - 1)\right]$$

$p_{k,t+2}$ is easy to interpret - it is the proportion of condition k's mean annual revenue in the evaluation period that a hospital expects will be clawed back by CMS as penalty in year $t+2$.

In the paper, the penalty rate p_h refers to the quantity described above. Figure 1 plots the penalty rate applied in 2013 (circles) for Pneumonia against the risk adjusted readmission rate over 2009-11.