Online Appendix: Impacts of Performance Pay for Hospitals: The Readmissions Reduction Program

Atul Gupta*

November 2020

A HRRP penalty

This section describes the exact formula used by CMS to determine the penalty for hospitals. It then walks the reader through a simple transformation to consider the penalty from the perspective of a forward-looking hospital. It concludes by discussing some key implications for hospitals.

For a hospital h at the end of year t, the penalty rate applicable to year $t+2$ is set as follows (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 through the period ending in June 2014, during which K was three. B is the total Medicare and b is total condition k base inpatient operating reimbursement, respectively, received over the evaluation period.[1] $r_{k\tau}$ denotes the risk-adjusted readmission rate calculated by CMS for patients of condition k at hospital h during the evaluation period, τ. It is normalized by the mean value across hospitals. For details on how $r_{k\tau}$ is computed, see appendix B.

This penalty rate is applied to all Medicare reimbursement received in year $t+2$. It is capped at 3% by law; however, fewer than 5% of hospitals reached the cap in the third year of the penalty (2015). The penalty rate translates to a dollar value by applying it to total Medicare inpatient base payments in year $t+2$, B_{t+2}.

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

The total dollar value can only be calculated ex-post since it depends on revenue received in year $t+2$. Hence, when a forward-looking hospital decides whether to

*The Wharton School, University of Pennsylvania, 3641 Locust Walk, 306 Colonial Penn Center, Philadelphia Pa 19104, atulgup@wharton.upenn.edu.

[1]Base payments do not include reimbursements received for training of graduate students and disproportionate service to poor patients. They represent 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`.

invest in improvements, it has to rely on an expected value of the penalty. This motivates the following transformation.

1. Total Medicare inpatient base payment appears in both the numerator and the denominator. First, we convert the aggregate values over τ to annual mean values and re-write the rate 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, 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 to be

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

   All of 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 revenue does not affect its 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 Medicare's share of hospital revenue across hospitals.

2. Since the penalty for each condition is constrained by the formula to be negative, this implies that a hospital cannot rely on good performance on one condition to compensate for poor performance on another. Further, assuming improvements are costly, a hospital's response will be tailored 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 receive a greater penalty burden compared to if it had a small cardiology department, all else equal. It is not clear whether hospitals will respond to an absolute value (a million dollars versus a $200,000 penalty) or a proportional value (5% of total revenue versus 1%). I choose to normalize the absolute value by the size of the condition's revenue ($\bar{b}_{k\tau}$), thus converting it to a proportional value, denoted as $p_{k,t+2}$:

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

   $p_{k,t+2}$ can be interpreted as 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 text, the penalty rate p_h refers to the object $p_{k,t+2}$ described above. Figure 1a plots the penalty rate applied in 2012-13 (circles) for pneumonia against the risk standardized readmission rate over July 2008–June 2011 (referred to as 2009–11).

B Data construction

This section describes details of the procedure used to construct the sample and key variables from raw Medicare claims files. I obtained access to Medicare claims files for calendar years 2006–2014 for the universe of fee-for-service Medicare patients. Since CMS uses July–June periods to compute risk-adjusted readmission rates for the penalty, I also organize my analysis around this cycle rather than by calendar years. I observe eight complete years from July 2006 through June 2014.

The Medicare claims data is organized around health care interactions. Each observation represents an inpatient stay, outpatient visit, emergency department visit, doctor's visit, etc. For each interaction I observe a rich vector of patient co-morbidities, the principal diagnosis for the visit or stay, what procedures were performed, and the dates of service. I can follow patients over time as well as identify the provider (physician and/or hospital) for each 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.

B.1 Sample selection

The penalty applies to all GAC hospitals that accept Medicare patients and are paid under the inpatient prospective payment system (IPPS).[2] CMS excludes critical access hospitals (CAHs) from all performance pay schemes, and so do I. Section B.5 below compares GAC and CAH hospitals, and shows that the two are very different on observed attributes. Another key difference is that CAH are not paid under IPPS, but rather receive cost-plus reimbursements.

I apply three additional restrictions to decide which hospitals to include 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 other studies that have used readmissions as a quality measure (Chandra et al., 2016; Ziedan, 2018), I exclude small hospitals from the main analysis.[3] Third, I limit attention to hospitals within the continental US (excluding Alaska, Hawaii, and all US territories). My final sample contains approximately 3,250 out of the 3,334 GAC hospitals that participate in IPPS (MedPAC, 2013).

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 recorded in "Carrier" or "Part B" files and requires some modifications to be incorporated into the analysis since it has a

[2] This implies that other types of hospitals (psychiatric, rehabilitation, and long-term care) or states exempted from IPPS (Maryland) are excluded.

[3] Readmission rates tend to be very noisy for small hospitals, particularly those with less than fifty admissions. CMS uses a cutoff of 25 admissions, while I prefer to use a cutoff of 50. The results are not sensitive to the exact cutoff value used.

different structure.[4] Part B files are useful in testing for an increase in coordination between hospital and community care post-discharge. 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. This excludes services such as ambulance transport. Second, the "place of service" must be the 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 identify and keep separate the Part B claims concurrent with a hospital stay or outpatient visit.[5] This ensures that I capture all spending on hospital care and interactions outside the hospital.

B.2 20% random Medicare sample

Due to data access limitations, several robustness checks and additional analyses are performed using a 20% random sample of Medicare claims. This sample covers the same period as the 100% files used in the main analysis, and the files are very similar—with two key limitations. First, the reduction in sample size naturally leads to a loss of precision (standard errors are 50–100% larger), and in some cases the loss of precision is enough to affect statistical significance of the IV estimates (e.g., mortality, screening of patients returning within 30 days of discharge). The loss of sample size also affects which hospitals can be retained in the sample. In the original analysis I exclude hospitals with fewer than fifty index cases for a given condition over the three year period 2009–11. With the 20% data, I have to proportionately lower the threshold in order to retain a similar number of hospitals. I use a threshold of fifteen patients for inclusion (an average of five patients per year). The resulting sample retains fewer hospitals than in the original analysis. The drop is about 5% of hospitals in the cases of heart failure and pneumonia. In the case of heart attack, the number of hospitals drops from about 1,850 to 1,550 (about 15%). Accordingly, the coefficients between the full sample and 20% sample differ the most for heart attack patients.

Second, instead of the Standard Analytic File (SAF) on inpatient claims, data on hospital stays are now sourced from the Medicare Provider Analysis and Review (MedPAR) file. These files are equivalent except in one important aspect. The SAF contains revenue codes, which can be used to determine services and drugs provided during the stay. For example, I use these to construct indicators for admission through ED and observation stays. However, MedPAR does not contain revenue codes, but instead has pre-coded indicators for these (and other) services. Unfortunately, the dummy for use of observation status was introduced only in 2011, permitting only one data point prior to the introduction of HRRP.[6] For these two reasons, I prefer to limit

[4]Each observation is not a separate interaction but a separate service within the same interaction. For example, if a patient visited the doctor's office and received a vaccination, this interaction would generate two rows of observations—one for the consultation or physician's professional fees, and one for the vaccination. These two observations would be identified by the same claim number. I collapse Part B observations to the claim number level to make it similar in structure to the hospital claims.

[5]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. This is also the case for an Emergency Department (ED) visit. For example, consider a patient that is hospitalized from January 1–5 and then receives some follow-up imaging and consultation a week later. The Part B file would record two claims, one for the hospital stay, and another for the consultation.

[6]For more details, see https://www.resdac.org/articles/identifying-observation-stays-those-

the use of the 20% sample.

Despite these limitations, a benefit of the auxiliary sample is that it also contains the relevant files on post-acute care use, prescription drugs (Part D), and records charges, which the primary data does not contain. I use charges as a measure of intensity of care. The primary data did not contain these files. I use the same algorithms and code to identify index cases and patient outcomes as I did for the main analyses. The data covers the same 302 Hospital Referral Regions (all except Alaska, Hawaii, Maryland, and US territories) as in the main analysis.

Table A.2 presents coefficients and R-squared values from hospital-level OLS regressions of sample quantities over the benchmark period 2009–11 on corresponding values from the 100% sample used in the main analysis. The regressions confirm that, on average, the number of index cases in the auxiliary sample are 20% as large as in the original sample, with R-squared values close to 1. Hence, the 20% sample replicates variation in patient volume (number of index cases) very well. Table A.2 panel B shows that hospital readmission rates are also aligned with those calculated using the universe of claims. The constant term in both regressions is close to zero and often statistically insignificant. The R-squared value in panel B is lower than in panel A, but remains substantial.

Table A.1 uses the 20% sample to replicate summary statistics on key outcomes presented in Table 1. This demonstrates that the auxiliary sample replicates population statistics on the key measures. Comparing values in the two tables indicates that the sample generates similar mean values for key outcomes. In most cases, the mean values differ by less than 0.01. As expected, the smaller sample size generates greater standard deviations (by up to 2x). The table does not repeat descriptive statistics for variables that were obtained from public sources (process of care scores) or that were retained from the original data (penalty related information).

The table also presents descriptive statistics for patients of two condition cohorts featured in additional analyses. These include 1) patients admitted with gastroenteritis (GI)—a condition unrelated to the targeted conditions and used in prior studies as a comparison group; and 2) patients admitted with conditions not targeted by HRRP and not in Circulatory or Respiratory disease classification groups ("Others"). I employ this cohort as a comparison group using a standard DD research design ignoring penalty status. The table also presents summary statistics on additional variables excluded from Table 1 due to brevity. These help characterize the index stays, readmissions, and spending by Medicare for the 30-day period starting with the index stay.

B.3 Key variables

Starting from the raw claims 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 for inpatient care separately with each of the conditions as the principal diagnosis. All analyses are conducted independently across condition cohorts.

beneficiaries-admitted-hospital.

B.3.1 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 30 days after a prior discharge for the *same* condition and the patient should be discharged from the hospital alive.[7]

Two situations 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 which 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. Instead, if 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 considered as a readmission or index case. Hence, CMS is not penalizing the number of readmissions, but the probability of any readmission. Finally, I cannot identify index cases in January 2006 since I do not observe hospitalizations in December 2005. I therefore 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 readmissions to any hospital for any reason within 30 days of discharge.[8] I tried to replicate CMS procedures exactly in order to identify index cases and readmissions.[9]

B.3.2 Process of care scores

I use 'process of care' scores released by CMS on the Hospital Compare website.[10] CMS tracks hospital compliance with a set of best practice clinical protocols for each condition. A key limitation of this data is that the scores pertain to all patients and not only to Medicare patients. Hence, these scores allow noisy measurement of changes in protocols for Medicare patients and will potentially under-estimate changes in treatment quality. There is considerable variation over time in the measures used for a condition. I prefer to use measures that have been reported both before and after

[7] Two quantitatively minor conditions are that the patient should not have been discharged against medical advice or have 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 then pertains to the previous case and the discharge date pertains to the transfer stay. If the combined case is an index case, it will be attributed to the first hospital. Other than heart attacks, transfers occur in less than 5% of cases. About 10% of heart attack admissions result in transfers. In addition, specific conditions sometimes have additional requirements. For example, heart attack admissions are not considered index if the patient was discharged the same day.

[8] 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.

[9] I obtained SAS code from the team at the Yale School of Public Health that executed this project for CMS. 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/AcuteInpatientPPS/Readmissions-Reduction-Program.html`.

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

2012 and ideally for all years in my sample. Accordingly, I use five, three, and five measures for heart attack, heart failure, and pneumonia, respectively.[11]

I standardize the scores following Chandra et al. (2016). To reduce the possibility of measurement error, I first exclude measures that were computed on fewer than 25 patients over the three-year period of July 2008–June 2011, the same threshold used by CMS to determine if a hospital should be exempted from the readmission penalty. Note that I drop the measure for the hospital for all years. I normalize raw scores to have weighted means of zero and standard deviations of one within each measure-year across hospitals. Each hospital therefore receives a standardized score for each measure in every year. I then obtain the mean of standardized scores across measures within the same condition-year for a hospital, weighted by the total number of patients comprising each measure. This collapses the data to the hospital-condition-year level. Finally, I standardize hospital scores within each condition to have means of zero and standard deviations of one and compute the unweighted means across hospitals. These scores are used in the regressions.

B.3.3 Hospital Value Based Purchasing measures

The Hospital Value Based Purchasing (HVBP) program was also introduced in 2012, i.e., the same time as the program studied in this paper. HVBP provides payment incentives to hospitals based on performance on compliance, quality, spending, and patient satisfaction. While HVBP does not target readmission rates, it does consider mortality rates for patients admitted with the same targeted conditions starting in 2013-14. In order to partition out changes in mortality potentially driven by responses to HVBP, I estimate additional specifications that include HVBP hospital performance scores from 2013-14.

HVBP incentivizes hospitals using a large array of performance measures across different categories (process of care, patient outcomes, efficiency, and patient satisfaction). The mapping from performance metrics to penalty is non-linear and more complicated than in HRRP. Norton et al. (2018) discuss this mapping in detail. I circumvent these complications by focusing on two aggregate performance measures. First, I use the "Total Performance Score" (TPS), which is a composite measure of hospital performance across conditions and measurement categories. Hospitals receive a score from 0–100 that reflects their relative performance. I standardize the TPS so it has a mean of zero and a standard deviation of one across hospitals. This data was obtained from Hospital Compare (`https://data.medicare.gov/data/archives/hospital-compare`). Second, I use the actual payment adjustment factors applied to inpatient reimbursements. These adjustments range from 0.99–1.01. A value above one indicates that the hospital will receive a bonus payment for good performance (unlike HRRP where hospitals do not receive bonus payments). These values were obtained from Table 16B available at `https://tinyurl.com/yadghn8g`.

Table A.10 panel A presents results using these controls. They include the TPS scores applicable for 2013-14. I also tested for the possibility of anticipatory effects by using corresponding scores released by CMS in 2012-13 and obtained similar results.

[11] Heart Attack: use of aspirin on arrival, on discharge, outpatient aspirin, PCI within 90 minutes of arrival, and statin prescribed at discharge. Heart failure: use of ACE inhibitor, discharge instructions, and evaluation of left ventricular systolic function. Pneumonia: appropriate antibiotic prescribed, ED blood culture before first dose, antibiotic administered within 6 hours of arrival, assessed and given influenza vaccination, and assessed and given pneumococcal vaccination.

B.3.4 Hospital characteristics

Hospital characteristics were obtained from two data sources. The primary hospital service type was assigned using information available within the Medicare claims data itself. Based on the Medicare provider number, hospitals were assigned to one of GAC, CAH, long-term care, or psychiatric. All the results presented in the text pertain to GAC hospitals only.

In addition, I used survey data from the American Hospital Association (AHA) to obtain information on hospital ownership, teaching affiliation, system affiliation (i.e., whether it is owned/operated by a chain and the identity of the chain), bed capacity (total staffed beds), utilization (inpatient days), location (city, hospital service area, hospital referral region, etc.), range of post-acute services offered (skilled nursing facility, hospice, home health, long-term care, intermediate care), number of full-time physicians and intensive care specialists, and share of different physician contract types (salary, equity, etc.) Hospitals are assigned values as of 2009, which is the first year with data available for all these variables. If information is not available for 2009, then I use the first year the hospital appears in the survey.

One of the hospital attributes studied in the heterogeneity analysis is whether high-penalty-risk system-owned hospitals show greater response to the penalty. To do so, I first classify hospitals as members of high-penalty-risk systems for each condition or not. I compute risk standardized readmission rates for each hospital in 2007 by replicating the CMS algorithm and assign each hospital to being penalized or not using the program's decision rule.

$$I_{h2007} = 1(r_{h2007} > 1)$$

I then identify if the hospital is owned/operated by a system. I only consider systems that have three or more hospitals treating Medicare patients. Smaller systems are disregarded and their members are considered as standalone. I then compute a leave-one-out proportion q_h^s of penalized hospitals in the system s, weighted by the number of index cases for the condition. This avoids spurious correlation between the hospital's own penalty status and the system's penalty proportion and takes into account volume heterogeneity in the system. The proportion is:

$$q_h^s = \frac{\sum_{l \in s, l \neq h} n_l \cdot I_{l2007}}{n_s - n_h},$$

where n_h and n_s are total index cases for hospital h and system s in 2007, respectively. Standalone hospitals are assigned $q_h = 0$. I then construct an indicator for hospitals in a system with more than 50% of index patients being treated at a hypothetically penalized hospital. These hospitals are considered to belong to an at-risk system. The indicator $I_s = 1(q_h^s > 0.5)$ is used in equation C.1 to characterize hospitals in at-risk systems.

B.3.5 Risk adjustment

To set penalties for each condition, CMS adjusts the observed readmission rate for differences in hospital case mix. The goal is to control for variation in readmission rates due to observed patient risk factors and focus on the residual, which is 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 random effect, assumed to be normally distributed with mean zero, and X is a vector of indicators for co-morbidities for each patient.[12] 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 second step obtains the intercept $\hat{\alpha}_h$ for each hospital, using the model estimated posterior variance of α_h. Two values are computed for each hospital:

- 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 is at the mean (zero). The risk-adjusted rate r_h for hospital h is then

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

Conceptually, the risk-adjusted rate measures how large or small a hospital's readmission rate is relative to the average hospital with the same patient mix. Appendix A discusses how r_h is used to set the hospital's penalty.

B.3.6 Other outcomes

In addition to readmissions, I examine hospital responses in admission and readmission decisions at the ED. This helps disentangle the patient's decision to seek treatment from the hospital's decision to admit. Patients' usage of the ED is inferred through the corresponding revenue codes.[13] This allows me to construct an indicator for hospital stays and outpatient care visits that originate in the ED.

[12] For example, the co-morbidities used in the 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 disease, asthma, pneumonia, end-stage renal disease, renal failure, other urinary tract disorders, and ulcer/skin ulcer. In addition, some other complications are used unless they only occur at the time of the index admission.

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

I have to change the approach of identifying ED use in the 20% sample since I do not have access to the Inpatient SAF. Instead, I have access to the MedPAR file, which does not include revenue codes. It does include an indicator for ED use that can be used directly. However, when I validate this indicator by comparing it with ED use recorded in the associated Part B claims for these patients (which is more reliable since I observe revenue codes in the Part B claims), it tends to flag a lower share of hospital stays as originating in the ED compared to the Part B claims. Hence, I preferr to use the indicator inferred from Part B claims.

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. In other words, these cases satisfy other conditions to be classified as index cases, such as no prior case of the same condition within the previous 30 days, but are not inpatient stays. This allows me to compute the probability of admission, i.e., the proportion of arriving index cases that were admitted for inpatient treatment.[14] I also create an indicator to identify that a patient returned to the ED within 30 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."

B.4 Assigning patients to conditions

I follow CMS to identify patients discharged with conditions targeted by HRRP.[15] CMS uses the principal discharge diagnosis ICD-9 code in order to assign patients to a particular cohort. Since the principal discharge diagnosis is also recorded in outpatient claims, I can also assign patients in the outpatient file to these conditions. This allows me to assign patients who were discharged from the ED to a cohort.

Apart from the three targeted conditions, I also construct cohorts of patients for other conditions. There are two such groups. The first is composed of specific non-targeted conditions to illustrate the gradient in spillover effects across different conditions, depending on how closely they are related to the targeted conditions. These include cardiac dysrythmia (ICD9 dx 427xx), cerebrovascular disease or stroke (ICD9 dx 430xx–438xx), gastroenteritis (GI), and renal failure (ICD9 dx 584xx–586xx). The first two are members of the "circulatory" category and are closely related to heart disease, so I expect to find spillover effects. I also consider COPD, which is a chronic respiratory disease added to HRRP in 2014-15. Hence, while we may expect some effects prior to 2014, they should be smaller in magnitude. GI and renal failure are relatively distant from the targeted conditions and are examples of conditions where I do not expect to find spillover effects. To mitigate the possibility of contamination, I exclude patients who also appear in one of the targeted condition cohorts.

Assigning cases to GI is a bit involved since there is no guidance from CMS on identifying discharge diagnosis codes, and prior studies have used Diagnosis Resource Groups (DRGs) to identify GI cases. For example, Mellor, Daly and Smith (2017) used the MS-DRG codes 329, 330, 331, 377, 378, 379, 391, and 392 to identify GI

[14] Approximately 90% of patients with these three conditions are admitted to the hospital through the ED. The remaining are admitted under directions of their physician. 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.

[15] I used the 'original methodology' reports for heart attack, heart failure, pneumonia, and COPD. The reports are available for download at `https://www.qualitynet.org/inpatient/measures/readmission/methodology`.

cases. However, MS-DRG codes are only assigned for inpatient claims. I would also like to identify GI patients who arrive at the ER but were not admitted for inpatient care. Hence, I use the claims data to identify principal discharge ICD-9 codes that are highly correlated with the above listed DRG codes. I select all ICD-9 codes for which more than 80% of inpatient claims belong to the above MS-DRG list. This algorithm produces the following set of principal diagnosis codes: 1530, 1531, 1532, 1534, 1536, 2113, 5301, 5310, 5314, 5320, 5324, 5378, 535xx, 5589, 562xx, 564xx, 5693, 578xx, 787xx, and 7890. For the years July 2008–June 2011, the benchmark period for the first HRRP penalty, this approach yields a 96% match rate, i.e., 96% of the inpatient claims with these ICD9 codes have DRG codes belonging to the above list. The three most important ICD9 codes by volume are 562 (one quarter of all stays), 578 (20%), and 5589 (8%). These three account for more than half of all GI stays, and their collective match rate is 98%.

I define a second comparison group for additional analysis using a simple exclusion restriction—I retain all patients except those discharged with a principal diagnosis within circulatory and respiratory classification groups, hip/knee replacement surgeries, and sepsis. Heart attack and heart failure belong to the first category, while pneumonia and COPD belong to the second. Hence, excluding these categories mitigates the potential for spillover effects. Hip and knee replacement surgeries were also targeted by HRRP starting in 2015. Finally, sepsis and pneumonia tend to have substantial overlap in patients. I refer to this composite group as 'others.' It is naturally a large cohort and accounts for more than 40% of all Medicare hospital admissions. Intuitively, GI and renal failure belong to this composite group. I use the 'others' cohort as the comparison group in a robustness check using a standard differences-in-differences research design, discussed in Section VI.B.

B.5 Critical Access Hospitals

CAHs have been used by some previous studies on HRRP as a comparison group to GAC hospitals since they were excluded from HRRP's penalty (Ibrahim et al., 2018). However, these hospitals are very different. By definition, they must have 25 or fewer acute care beds and must be located more than 35 miles from the nearest hospital. They therefore tend to be located in rural areas. In addition, they are not paid under the prospective payment system, instead they are paid on a cost-plus basis. Since they have low patient volume, it is difficult if not impossible to credibly estimate risk-adjusted readmission rates for these hospitals. Table A.11 illustrates the magnitude of these differences by presenting descriptive statistics for GACs and CAHs on patient volume. These values were obtained using the 20% Medicare data sample. The first five columns present values based on patient cohorts for specific conditions, and the last column presents aggregate values across all patients.

Row 1 shows that there are about 1,250 CAHs across the country, compared to 3,300 GAC hospitals. Hence, CAHs account for about a quarter (with some variation across conditions) of all hospitals serving Medicare patients. The second row shows that the average CAH admitted only 1, 4, and 6 patients per year in the 20% sample in each of the targeted conditions during the benchmark period of 2009–11. Hence, each hospital-condition-year cell tends to be very small and would lead to noisy estimates. In my analysis, I typically drop hospitals with fewer than 15 patients over this three year period. Since there are fewer heart attack cases relative to the other two targeted

conditions, using a uniform threshold affects more hospitals in the heart attack analysis sample. Row 3 shows that 35% of GAC hospitals in the sample have heart attack volumes below this threshold, but the corresponding proportion for CAHs is 70%. Similarly, I would have to exclude 63% and 43% of CAH hospitals for the other two targeted conditions under this threshold.

One may argue that if I used a difference-in-difference research design with CAH as the comparison group and ignore penalty status, I could use within-market analysis instead of within-hospital and circumvent the issue of small hospital size. Such an approach is feasible but not ideal since I would not eliminate unobserved differences across hospitals, and CAHs are very different indeed. However, even this approach is difficult to implement since CAH's aggregate share of patients is only 8%, as seen in the bottom row. This varies across conditions, from 2% of heart attack cases to 10% of pneumonia cases.

There is also a sharp difference in the geographic distributions of GAC and CAH hospitals. The former are most likely to be found in California, Texas, New York, Florida, and Pennsylvania. These states are among the largest by population and also account for a third of all GAC hospitals. However, these states only have 11% of CAHs. Instead, CAHs are most likely to be found in Iowa, Kansas, Minnesota, Texas, and Nebraska. In fact, there are only four states that have more than 50 of both types of hospitals—Texas, Minnesota, Illinois, and Wisconsin.

C Miscellaneous topics

C.1 Clinical mechanisms to improve quality

This section attempts a comprehensive examination of changes in treatment patterns during the index stay and in the 30 days following discharge, which I consider to be the episode. I examine estimated effects on the use of intensive care, length of stay, compliance with best practices for the index case, use of primary and post-acute care (PAC), and prescription drugs in the 30 days following discharge. I consider aggregate measures of cost as well as specific margins of care such as the use of diagnostic services, medical supplies, and procedures. Table A.5 panels A, B, and C present the corresponding IV estimates obtained using equation 3. These models are estimated using the 20% Medicare sample. The specifications adjust for patient medical history, although the coefficients are similar (as in other cases) if I only adjust for patient demographics. As is common in the literature, I use recorded charges as a proxy for treatment cost and use the two words interchangeably. All charges and spending values are expressed in 2016 dollars. Regressions for spending and length of stay are performed in logs, so the coefficients can be interpreted approximately as percent changes.

Panel A presents estimated effects on total 30-day charges split between charges for readmission(s) (zero in cases with none) and all remaining services. This allows me to test for reductions not from readmissions, which account for 9% of the total episode cost for heart attacks and about 16% for the other two conditions. Further, I limit the sample to patients discharged alive from the initial stay so as to avoid confounding due to changes in in-hospital mortality. I find evidence of a statistically and economically significant decline in readmission charges for heart attack and pneumonia, while in the case of heart failure the effect is much smaller in magnitude and not statistically

significant. The coefficients imply a 12–13 pp decline in readmission charges on average for heart attack and pneumonia. This translates to a 1–2 pp decline in total episode cost, in line with the estimated decline in readmissions reported in Section IV. The coefficients in row 2 of panel A indicate no evidence of a net change in the cost for the remaining services provided to patients during the 30-day episode for any condition.

The aggregate null effect on total episode cost (other than readmissions) may mask changes in specific types of services. Table A.5 panel B examines treatment intensity during the initial hospital stay, including patients who died in the hospital. There is no evidence to indicate an increase in hospital charges or length of stay for any of the conditions, which belies the hypothesis that hospitals may keep patients for a longer duration in order to better stabilize them before discharge. Further, no increase was found in a variety of specific services that could be plausibly hypothesized—laboratory cost, operating room cost, cardiology services, pharmacy cost, and others. These are not presented for brevity but are available on request.

Nevertheless, there are clear indications of increases in intensity for heart attack patients on a number of dimensions. For example, the coefficients indicate an approximately 25% statistically significant increase, conditional on using any supplies, in physician and medical supply costs. This implies an economically significant increase of about 13% on average (since the mean probability of penalty is about 0.5). There is also an implied increase of 2 pp (0.04 x 0.5) in the use of radiology diagnostic imaging. One could characterize these results as a pattern of physicians spending more time with heart attack patients and potentially using more diagnostics and supplies. I also find evidence of greater use of procedures—an implied increase of 2–3 pp in the use of Angioplasty and Coronary Care Units. In contrast, the coefficients indicate a 3% (7%) decline in hospital (physician) charges during the initial pneumonia stay. While these are not always significant at the 5% level, the coefficients for pneumonia are consistently negative across a range of services, which suggests the pattern of decreasing intensity is robust.

Panel C investigates changes in treatment post-discharge from the hospital, for which I once again limit the sample to patients discharged alive from the initial stay. A possible response to the penalty would be to increase patient referrals to facility based PAC (extensive margin) as well as lean on them to increase their treatment intensity (intensive margin). The latter channel is certainly feasible in cases where the PAC is vertically integrated, and a recent study by David, Gupta and Kim (2019) suggests this phenomenon may also extend to standalone PAC providers. Table A.5 panel C tests this hypothesis on both extensive (row 1) and intensive (row 2) margins. Overall, the coefficients suggest that the use of post-acute care has decreased for both heart attack and pneumonia patients, but through different mechanisms. In the case of AMI patients, there is no change on the extensive margin, but there is a 10% decrease in cost on the intensive margin. Pneumonia patients are 2 pp less likely to be sent to PAC, with no statistically significant change on the intensive margin. There is also no support for greater use of primary care in the first week post-discharge. Taken together, this evidence belies the hypothesis that hospitals have increased coordination with community-based providers.

Anecdotal evidence based on discussions with physicians and hospital administrators indicates that hospitals responded to the penalty by strengthening their discharge planning checklists and provided greater assistance to patients in adhering to their drug regimens. I examine this dimension of care by investigating changes in the use of

prescription drugs as recorded in the Part D claims. I find no change on the extensive margin, but there is evidence that patients are using more (or more expensive) drugs —Medicare payments increased by 4–5% for the heart attack and pneumonia cohorts.

The introduction of the penalty may have spurred hospitals to tighten adherence to 'best-practice' treatment protocols. If care protocol standardization helps decrease 'mistakes' that cause readmissions, this may be an important mechanism. As discussed in Section B.3.2, I compile annual hospital scores on process compliance released by CMS and standardize them for use in regression analysis. A higher score signals better performance. Table A.5 panel D presents the IV estimates with these scores as the outcome. Since the scores have been transformed to have mean zero and standard deviation of one, the coefficients are easily interpretable. The magnitude of the estimated effect is large but not always statistically significant, perhaps due to sampling noise since the scores pertain to all cases of the targeted conditions at the hospital—including readmissions and non-Medicare patients. The results also indicate that protocol compliance decreased for heart failure patients while it improved for the other two conditions.

To summarize the results in this section, there is no sharp evidence of an increase in treatment intensity or process compliance and no consistent pattern across conditions. I do find consistent evidence of an increase in intensity and process compliance—although not on aggregate cost—for heart attack patients. The evidence is quite mixed for the other two conditions. There is no statistically significant change in any dimension for heart failure, except for a decline in process compliance. In the case of pneumonia, I find a decrease in intensity of care but improved process compliance. Overall, these results imply that if there are consistent changes in clinical protocols or treatment intensity, they are occurring on margins not captured in claims data.

C.2 Evidence beyond claims data

Joynt, Figueroa and Jha (2016) analyze survey responses from about 950 hospital administrators and report that penalized hospitals were more likely than non-penalized hospitals (71% versus 59%) to respond that HRRP significantly impacted their efforts to reduce readmissions. Section IV.B sheds light on possible changes in clinical mechanisms, but clearly more evidence is needed. Discussions with hospital administrators and physicians indicate that hospitals responded through a combination of targeted initiatives such as more standardized discharge planning, drug reconciliation, and greater post-discharge follow-up by case managers, but did not implement major changes in treatment protocols. In some cases, these initiatives were directed and monitored by a central office specially established by the hospital to respond to value-based payment reforms. David, Gupta and Kim (2019) use operations data from a large home health care firm and show that the firm incurred greater costs for Medicare patients targeted by HRRP, with corresponding reductions in readmissions. This suggests that while hospitals did not increase the use of PAC on the extensive margin, they may have done so effectively on the intensive margin. Taken together, this limited evidence indicates that the clinical responses are difficult to measure solely using claims data and may necessitate the use of clinical or operational data.

C.3 Heterogeneity across hospitals

This section presents evidence on heterogeneity across hospitals in response to the readmissions penalty, holding constant their penalty expectations. I draw on the prior literature for guidance on four specific attributes that could cause hospitals to be differentially sensitive to the penalty or differentially capable of responding to performance pay incentives—ownership (private or government), membership in a larger system that is penalized, patient volume, and spare bed capacity. To quantify heterogeneous responses, I use a triple-difference research design where the third dimension is a predetermined hospital feature of interest (see equation C.1 below). The identification assumption is that hospitals with and without the feature of interest would evolve along parallel trends in the absence of the penalty. These regressions are estimated on hospital-year level data.

$$
\begin{aligned}
Y_{ht} = \alpha_h + \delta_t + \theta_1 \; Z_h \cdot \mathbf{1}(t \geq 2012) \\
+ \quad \theta_2 \; K_h \cdot \mathbf{1}(t \geq 2012) + \theta_3 \; Z_h \cdot K_h \cdot \mathbf{1}(t \geq 2012) + X_h' \gamma + \epsilon_{ht}
\end{aligned} \tag{C.1}
$$

This model produces reduced form estimates that are not directly comparable to the IV coefficients obtained by estimating equation 3. A more useful comparison is between the average effect for all hospitals (θ_1) and the triple difference coefficient for the hospital attribute of interest (θ_3). Low-penalty-risk hospitals that lack the attribute of interest form the reference group. For example, if we are interested in government ownership, the reference group is low-penalty-risk private hospitals. Z_h is the baseline instrument used in the main analysis and provides identifying variation across hospitals. K_h is an indicator for the hospital feature of interest. In this example, it takes a value one for all government-owned hospitals.[16] θ_1 is effectively the D-D estimator for privately-owned hospitals. θ_2 captures the secular change post-HRRP for low-penalty-risk government-owned hospitals. θ_3 is the triple-difference estimator and quantifies if government-owned hospitals differentially changed their readmissions, relative to private hospitals (θ_1) and holding penalty risk constant. Standard errors are clustered by hospitals. Table A.6 presents the results with 30-day readmissions as the outcome.

C.3.1 Government ownership

Previous studies have shown that government owned hospitals operate under a soft-budget constraint (Duggan, 2000; Baicker, Staiger et al., 2005) and hence may not respond to high powered financial incentives such as HRRP. Table A.6 panel A presents corresponding results obtained for each condition separately. A similar proportion of government and privately owned hospitals were penalized under HRRP. The results suggest that, 1) low-penalty-risk government hospitals do not differentially respond post-HRRP ($\theta_2 \approx 0$), and 2) since θ_3 is positive (though not statistically significant), at-risk government hospitals improve by a smaller amount in comparison to private hospitals at the same level of penalty risk. Hence, the estimates indicate that government hospitals are less responsive. This is consistent with conclusions drawn by previous studies. None of the point estimates are statistically significant.

[16] The sample excludes federally owned hospitals and includes state, city, and county-owned hospitals. About 17% of hospitals in the data are government owned.

C.3.2 Hospital systems

Hospital consolidation has received considerable attention in the economics literature, especially regarding the question of whether mergers lead to improved operational efficiencies (Gaynor and Town, 2012; Schmitt, 2017). Hospital systems can leverage larger scale operations to invest in better IT architecture, better navigate complex regulations, and retain higher quality managers. These factors may lead to greater productivity (Tsai et al., 2015) and enable hospitals to respond faster to performance pay incentives. Further, hospital systems may push reforms at all member facilities regardless of whether a specific facility is expected to be penalized or not.

I formally test this conjecture. Specifically, I test if hospitals owned by systems expecting a large share of their hospitals to be penalized responded more than hospitals with the same penalty-risk but not owned by such a system. I obtained information on system ownership from the AHA survey files. I assign hospitals as standalone or system affiliated based on their status in 2009 or the first year they appear in the survey. Approximately 1,900 hospitals (55%) are classified as system owned or operated. I compute the leave-one-out proportion of hospitals in each system (weighted by number of index cases) that would be penalized based on their performance in 2007. I use a leave-one-out approach to avoid spurious correlation between a hospital's own status and the system's penalty share. Systems with a 'score' greater than 0.5 are deemed high penalty systems, i.e., the HRRP penalty should be highly salient for these systems. Standalone hospitals receive a score of zero. Section B.3.4 discusses construction of the measure in detail. About 61% of system hospitals (33% of all hospitals) are assigned to high-penalty systems under this definition.

Table A.6 panel B presents corresponding coefficients. Estimates of θ_2 imply that simply being owned by an at-risk system helps hospitals differentially improve post-HRRP, regardless of their own penalty status. This is in contrast to the corresponding result for government ownership or large hospitals. Hence, hospital systems may enforce revised protocols uniformly across their facilities. The effects are small but noteworthy. The triple-difference estimator indicates that—at least for heart failure—hospitals owned by at-risk systems differentially improved more than remaining hospitals, holding penalty risk constant.

System ownership seems to help produce a greater decline in readmissions, but it is not clear if this is due to greater focus on quality, selection, or both. A full investigation is beyond the scope of this paper. I did look at whether system-owned hospital admission decisions changed for patients returning to the ED within 30 days of discharging from another hospital owned by the same system. Using a triple-difference specification on the sample of patients returning to a different hospital (the third dimension being whether the hospital was owned by the same system as that for the index stay), I find that system-owned hospitals are less likely to readmit a returning patient if she happens to be from a hospital owned by the same system. The triple-difference coefficients (not presented) are large, consistently negative, and in the case of pneumonia, statistically significant. This evidence supports the interpretation that systems share and use detailed information on patients across member facilities and implement uniform protocols.

C.3.3 Patient volume

There is a large empirical literature documenting the correlation between patient volume and health outcomes (Luft, Bunker and Enthoven, 1979; Luft, Hunt and Maerki, 1987). Recent work has partially addressed concerns over unobserved selection in patient-provider matches and established that greater volume leads to improved quality (Gowrisankaran, Ho and Town, 2004; Gaynor, Seider and Vogt, 2005; Hentschker and Mennicken, 2017). The volume-outcome hypothesis implies that larger hospitals achieve higher quality through more patient practice or learning-by-doing. Consistent with this interpretation, I ask whether larger hospitals are also able to differentially improve their readmissions rate, holding penalty-risk constant. I assign hospitals to different tertiles of patient volume based on the number of index cases over 2007–09. I consider hospitals in the top tertile (approximately 1,000) as large for the purpose of this analysis.

Table A.6 panel C presents corresponding results. The results are somewhat similar to those for at-risk hospital systems—large hospitals differentially decrease readmissions post-HRRP, holding penalty-risk constant. Again, the effect of being large is particularly important in the case of heart failure patients, where the total effect more than doubles in magnitude at top tertile hospitals.

C.3.4 Spare bed capacity

Hospitals struggling to fill their beds may value readmissions even if there is a chance that an additional readmission may invite a penalty in the future. Holding all else equal, I would expect hospitals with more spare bed capacity to produce a lower decline in readmission rates. To test this hypothesis, I identify hospitals in the bottom (patient weighted) tertile of bed capacity utilization as of 2009. Over this period, hospitals in the bottom tertile had a utilization rate of 55% versus 80% for the remaining hospitals. I only observe total hospital beds, and therefore I can only calculate hospital capacity utilization. This is a crude measure since it reflects utilization at the hospital level and may not accurately reflect utilization of the departments relevant to the targeted conditions. This introduces measurement error into the analysis if the true determinant is spare capacity in the relevant department (e.g., cardiovascular medicine specifically). With this caveat in mind, I use this indicator to characterize hospital spare capacity. Panel D presents the corresponding triple-difference results. Due to data limitations the analysis in this panel was performed using the 20% Medicare sample. To make the estimates comparable to those presented in panels A through C, I weight hospitals by the number of index cases over the benchmark period 2009–11, as recorded by CMS.

The triple-difference coefficients are statistically insignificant, and inconsistent across the conditions. The triple-difference coefficient is large and positive for heart attack implying that hospitals with greater spare capacity produce a smaller decline in readmission rates than the remaining hospitals, holding penalty-risk constant. In fact, the coefficient is large enough to offset the main effect (0.15 versus -0.18), implying that there was no change in readmissions at these hospitals. The triple-difference coefficients are negative for both heart failure and pneumonia but are only meaningful in the case of heart failure—in fact, the results imply that most of the decline in readmissions is driven by hospitals in the bottom tertile by capacity utilization.

C.4 Non-targeted conditions

The main analysis exploits differences in penalty incentives across hospitals and quantifies the relative change over time within patients of the same condition. An alternative approach could exploit the fact that HRRP's penalty was not applied to all Medicare patients. Patients with non-targeted conditions account for more than 85% of all hospital stays on average. An approach using such patients as a comparison group ignores hospital penalty status and estimates an 'intent-to-treat' effect of introducing the penalty. A nice feature of this research design is that I can include all hospitals, and mean reversion is not a concern since I do not condition on penalty risk.

Some previous studies (Desai et al., 2016; Ody et al., 2019) have used this approach, selecting different subsets from among the non-targeted conditions as controls. I prefer to use all non-targeted patients as the comparison group, except those very likely to receive spillover effects. Accordingly, I exclude patients with conditions belonging to respiratory or circulatory condition categories or hip/knee replacement surgeries. I refer to this group as 'Others.' Table A.1 column 5 presents descriptive statistics for this cohort. This group is composed of patients admitted with several different conditions, and adverse outcomes occur less frequently on average.[17] Combining these patients with those admitted with a targeted condition, I estimate traditional D-D models of the form below:

$$Y_{iht} = \alpha_h + \delta_t + \kappa d_i + \beta\ d_i \cdot 1(t \geq 2012) + X_i'\gamma + \epsilon_{iht}, \tag{C.2}$$

where d_i is an indicator for a patient belonging to a condition targeted by HRRP. Hospital fixed effects α_h eliminate time invariant unobserved hospital features that affect patients of both targeted and non-targeted conditions. The remaining terms have the same interpretation as in the main specification. Standard errors are clustered by hospital to account for possible correlation across patients discharged from the same facility. β is the coefficient of interest and estimates the intent-to-treat (ITT) effect of introducing HRRP under the usual identifying assumption—in absence of the program, the targeted condition and comparison group would evolve along parallel trends. I assess the validity of this assumption by estimating year specific coefficients β_t using a dynamic variant of equation C.2. Figure A.10 plots these coefficients for all the key outcomes.

Figure A.10a presents unadjusted trends in aggregate readmission rates for the targeted (diamonds) and non-targeted (squares) conditions. For ease of comparison, both lines are normalized by their respective values in 2007. The plot clearly shows a decline for the targeted conditions, while the trend remains stable for the comparison group. Table 6 row 8 presents the coefficients obtained using equation C.2. The model includes a full vector of co-morbidity history. These coefficients can be directly interpreted as the average intent-to-treat effect, while the estimates in rows 1–7 must be scaled by 0.5, the mean probability of penalty.

The patterns are very similar to those seen in the main estimates. For example, there is a statistically significant decrease in all three outcomes, and the magnitude tends to be larger for heart attack patients, particularly in the case of mortality. The aggregate effect on 30-day readmissions across all three conditions is about 0.9 pp (not

[17] Section B.4 describes how the cohort is identified. This group accounts for more than 40% of all inpatient stays in the sample. The largest disease groups in this comparison group are Digestive (22%), Urinary (15%), Cancer (10%), and Endocrine (7%).

presented due to space constraints), which is only slightly smaller than the 1 pp implied by the main results. Similarly, the implied aggregate effect on one-year mortality is 0.32 pp, comparable to the 0.45 pp implied by the main estimates. I interpret the mortality coefficient for heart attack patients with caution since there is a declining pre-trend. These are intent-to-treat estimates and may be biased down due to positive spillovers to the non-targeted patients. A noteworthy point of departure from the main results is that this approach estimates a statistically significant increase in mortality among heart failure patients. However, this estimate also falls within the confidence interval implied by the coefficient from the main analysis.

C.5 Concerns on mortality effects

This section attempts to reconcile contradictions between mortality results in the main analysis and those reported by Gupta et al. (2018). The authors estimate an increase of 1.4 (5) pp in 30-day (one-year) mortality for heart failure patients. In contrast, the primary results in Table 3 panel B can rule out corresponding increases of more than 0.4 (0.8) pp even for hospitals that are certain to be penalized.

The most important difference between the two approaches is the research design. I use a dose-response type research design that differences out trends for hospitals at low-penalty risk and attributes differential changes to HRRP. In contrast, Gupta et al. (2018) consider the absolute change in mortality for heart failure patients post-2010 relative to the period 2006–10. A second difference is that their analysis was performed on a non-random sample of about 420 hospitals, while I use a national sample. A third, minor difference is that they consider post-discharge mortality, whereas I include in-hospital deaths.

In order to quantify the importance of each of these differences, I replicate their analysis using the 20% Medicare sample and then introduce changes step by step to move closer to my approach. Specifically, I estimate a model predicting one-year post-discharge mortality using patient observables (demographics and medical history) and hospital and quarter-of-sample fixed effects. Figure A.11a plots the effects relative to 2006Q1, aggregated to each calendar year (red diamonds). The figure is oriented on calendar years to match their approach, even though my main analysis is organized around July–June periods. There is clearly an increase in mortality over this period. Much of the increase occurred by 2011, and mortality rates stabilized after 2012.

Table A.10 panel B presents the estimated annual coefficients aggregated into the three periods used by Gupta et al. (2018)—1) pre-ACA, i.e., through March 2010; 2) announcement, April 2010–September 2012; and 3) implementation, October 2012–December 2014. Column 1 presents the corresponding results for all hospitals and shows that mortality for heart failure patients increased during implementation relative to the pre-ACA period by a precisely estimated 3.1 pp. Throughout this section I will refer to this coefficient as the aggregate mortality effect.

Gupta et al. (2018) use a sample of 420 hospitals that enrolled voluntarily in an information sharing program and hence may be selected on unobserved features. Khera, Dharmarajan and Krumholz (2018) discuss several limitations of using this hospital sample. To simulate the variability introduced due to using a subset of hospitals, I re-estimated the model 500 times, each time drawing a 15% sample with replacement or about 450 hospitals. Figure A.11a plots the 95% range of estimated values for each year (grey diamonds), showing that for some subsets of hospitals, mortality increased

by about 5 pp over this period. Table A.10 panel B columns 2 and 3 present the corresponding 5% and 95% bounds on the main effect, which are about 2.5–4 pp. Column 4 presents the corresponding estimate when I change the measure to include in-hospital deaths. The aggregate mortality effect drops from 3.1 pp to 2.8 pp.

My research design considers the differential change for high-penalty-risk relative to low-penalty-risk hospitals. Consistent with my main analysis, I classify hospitals in the top tertile by predicted readmission in 2006-07 as high-penalty-risk, and remaining hospitals as low-risk. Figure A.11b presents corresponding annual estimates separately for both groups. The hospitals at high-risk consistently have a lower mortality rate, and their increase over time also appears slightly lower in magnitude. Table A.10B columns 5 and 6 formally report the main effects for high and low penalty-risk hospitals, respectively. The difference is reported in column 7; it is negative and statistically insignificant.[18] Hence, hospitals that had greater incentives to respond to the penalty experienced a slightly lower increase in mortality.

Finally, I illustrate the concern about increased unobserved patient severity over this period. Figure A.11c presents the corresponding annual estimates for post-discharge mortality among patients with non-targeted 'other' conditions. This is the same group used in the alternate research design reported in Section VI.B and accounts for about 40% of all admissions. There is a similar, albeit lower magnitude, increase in mortality among patients who were ostensibly not affected by the program. This suggests that there could be other common factors operating over this period that led to increased mortality across a wide range of conditions. An important factor could be the increase in the share of beneficiaries that enrolled in Medicare Advantage. It is well known that MA plans tend to attract (or seek) healthier beneficiaries, and hence the residual fee-for-service patient pool may have increased in unobserved severity over time. Other contemporaneous policies could have also played a role. MedPAC (2018) discusses the Recovery Audit Contractor (RAC) program introduced by CMS in 2010 as a factor that decreased short-stay hospital admissions (typically of healthier patients) during this period. They show evidence that this program had substantial bite for heart failure.

References

Baicker, Katherine, Douglas Staiger, et al. 2005. "Fiscal Shenanigans, Targeted Federal Health Care Funds, and Patient Mortality." *The Quarterly Journal of Economics*, 120(1): 345–386.

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

David, Guy, Atul Gupta, and Lucy Kim. 2019. "The incidence of mandated performance pay penalties: Evidence from health care."

[18] Note that this estimate is still not perfectly comparable to the main estimates reported in Table 3B for two reasons. First, this is a reduced form and not the IV estimate. Second, the main analysis uses July 2011 onward as the post-HRRP period, which falls in between periods II and III in this analysis. Loosely speaking, a more comparable estimate would be an average value lying between the -0.0071 and -0.0019 reported in column 6.

Desai, NR, JS Ross, J Kwon, and et al. 2016. "Association between hospital penalty status under the hospital readmission reduction program and readmission rates for target and nontarget conditions." *JAMA*, 316(24): 2647–2656.

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

Gaynor, Martin, and Robert Town. 2012. "The impact of hospital consolidation – Update."

Gaynor, Martin, Harald Seider, and William B Vogt. 2005. "The volume-outcome effect, scale economies, and learning-by-doing." *The American Economic Review*, 243–247.

Gowrisankaran, Gautam, Vivian Ho, and Robert Town. 2004. "Causality and the volume-outcome relationship in surgery." *The Sixth Conference on the Industrial Organisation of Healthcare. Hyannis, Massachussets*, 22–25.

Gupta, Ankur, Larry A Allen, Deepak L Bhatt, Margueritte Cox, Adam D DeVore, Paul A Heidenreich, Adrian F Hernandez, Eric D Peterson, Roland A Matsouaka, Clyde W Yancy, et al. 2018. "Association of the hospital readmissions reduction program implementation with readmission and mortality outcomes in heart failure." *JAMA Cardiology*, 3(1): 44–53.

Hentschker, Corinna, and Roman Mennicken. 2017. "The Volume-Outcome Relationship Revisited: Practice Indeed Makes Perfect." *Health Services Research*, n/a–n/a.

Ibrahim, Andrew M, Justin B Dimick, Shashank S Sinha, John M Hollingsworth, Ushapoorna Nuliyalu, and Andrew M Ryan. 2018. "Association of coded severity with readmission reduction after the Hospital Readmissions Reduction Program." *JAMA Internal Medicine*, 178(2): 290–292.

Joynt, Karen E, JE Figueroa, and Ashish K Jha. 2016. "Opinions on the Hospital Readmission Reduction Program: results of a national survey of hospital leaders." *The American Journal of Managed Care*, 22(8): e287–e294.

Khera, Rohan, Kumar Dharmarajan, and Harlan M Krumholz. 2018. "Rising mortality in patients with heart failure in the United States: facts versus fiction." *JACC: Heart Failure*.

Luft, Harold S, John P Bunker, and Alain C Enthoven. 1979. "Should operations be regionalized? The empirical relation between surgical volume and mortality." *New England Journal of Medicine*, 301(25): 1364–1369.

Luft, Harold S, Sandra S Hunt, and Susan C Maerki. 1987. "The volume-outcome relationship: practice-makes-perfect or selective-referral patterns?" *Health Services Research*, 22(2): 157.

MedPAC. 2013. "A data book: Health care spending and the Medicare program."

MedPAC. 2018. "Mandated report: Effects of the Hospital Readmissions Reduction Program."

Mellor, Jennifer, Michael Daly, and Molly Smith. 2017. "Does it pay to penalize hospitals for excess readmissions? Intended and unintended consequences of Medicare's hospital readmissions reductions program." *Health Economics*, 26(8): 1037–1051.

Norton, Edward C, Jun Li, Anup Das, and Lena M Chen. 2018. "Moneyball in Medicare." *Journal of Health Economics*, 61: 259–273.

Ody, Christopher, Lucy Msall, Leemore S Dafny, David C Grabowski, and David M Cutler. 2019. "Decreases in readmissions credited to Medicare's program to reduce hospital readmissions have been overstated." *Health Affairs*, 38(1): 36–43.

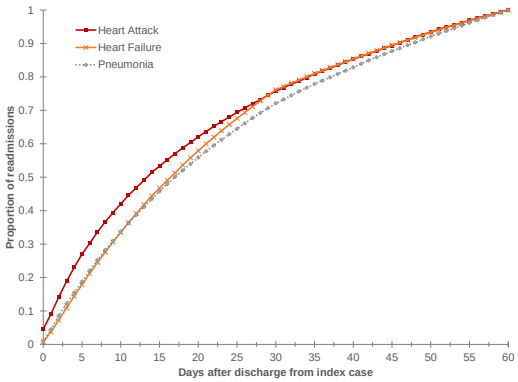
Schmitt, Matthew. 2017. "Do hospital mergers reduce costs?" *Journal of Health Economics*, 52: 74–94.

Tsai, Thomas C, Ashish K Jha, Atul A Gawande, Robert S Huckman, Nicholas Bloom, and Raffaella Sadun. 2015. "Hospital board and management practices are strongly related to hospital performance on clinical quality metrics." *Health Affairs*, 34(8): 1304–1311.

Ziedan, Engy. 2018. "The Intended and Unintended Consequences of the Hospital Readmission Reduction Program." *Available at SSRN 3350492*.

Appendix Figures and Tables

(a) Readmissions



(b) Mortality

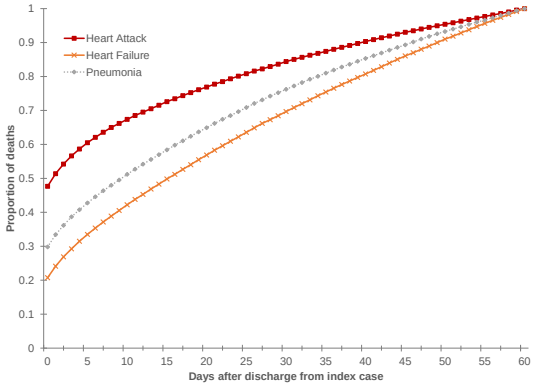


Figure A.1: Distribution of outcomes post-discharge

Note: This figure presents the cumulative proportion of readmissions (panel a) and deaths (panel b) over the 60 days post-discharge from the initial stay for each of the three targeted conditions, using data from July 2007 through June 2011. All in-hospital deaths are attributed to day zero.

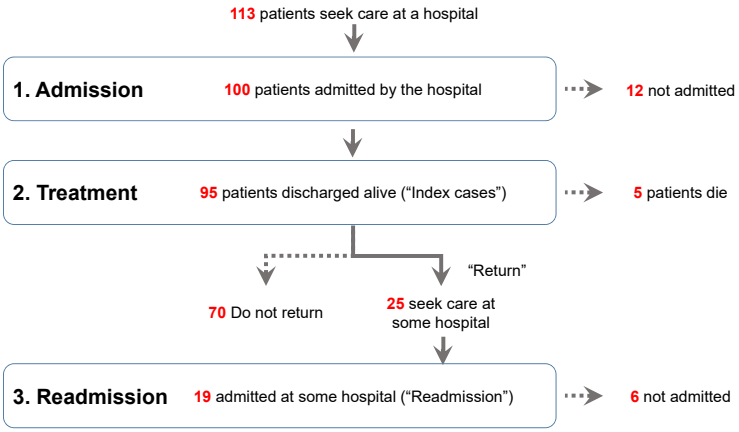


Figure A.2: Stylized readmission cycle

Note: This figure presents a stylized flowchart of the steps a patient could undergo during the course of a hospital admission and readmission. The numbers are normalized such that 100 patients are admitted to the hospital in a year. The proportions match actual proportions for Medicare hospital cases with heart attack, heart failure, and pneumonia over July 2006– June 2011. 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 these occur very rarely. The return and readmission can be to any hospital, not just the original discharging hospital.

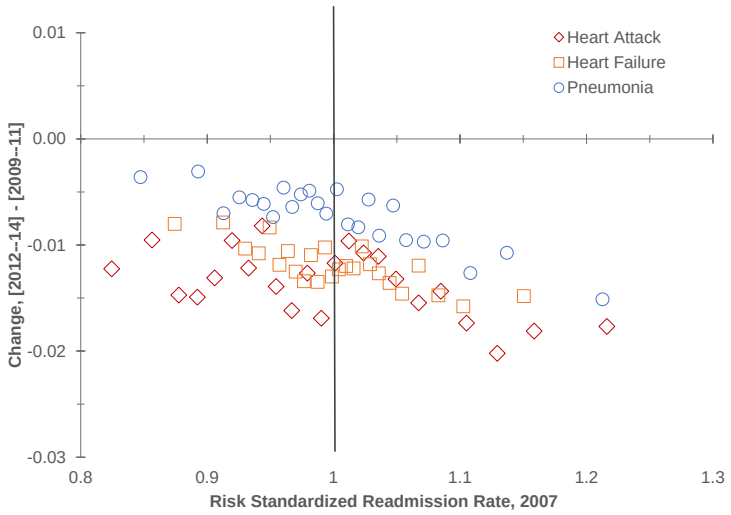


Figure A.3: Change in risk-standardized readmission rate (by condition)

Note: This figure plots the change in risk standardized readmission rate (RSRR) for each condition between 2012–14 (after) and 2009–11 (before), for 25 equal-sized hospital cells on the Y-axis against their 2007 RSRR value on the X-axis. Figure 2b summarizes this data by presenting the weighted mean across conditions.

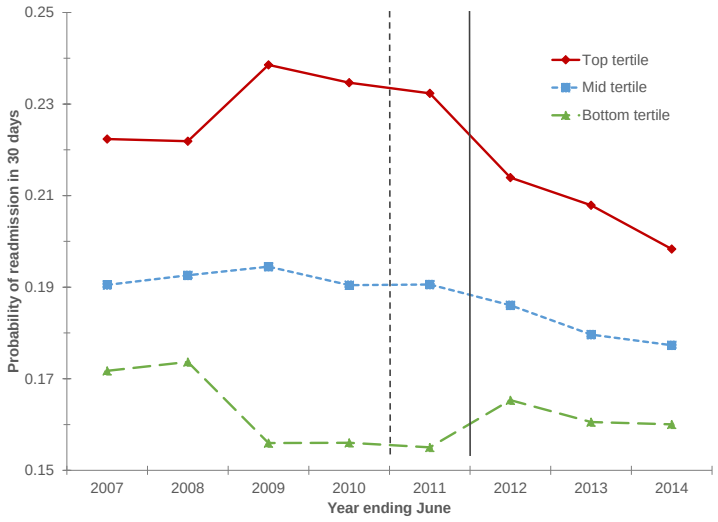


Figure A.4: Mean reversion

Note: This figure presents un-adjusted trends in the mean probability of readmission for the targeted conditions for hospital tertiles based on penalty likelihood P_{h1}. Values are computed as weighted averages across conditions. This illustrates the underlying variation when estimating equation 1 by OLS.

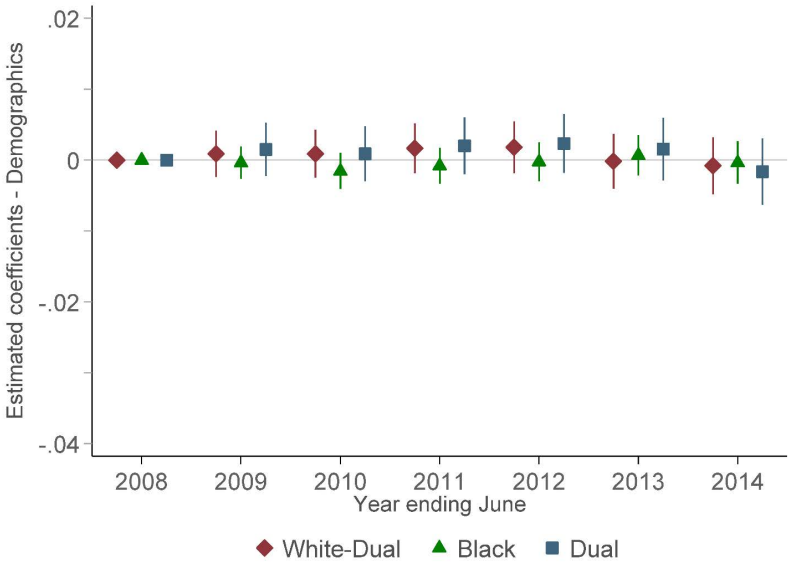
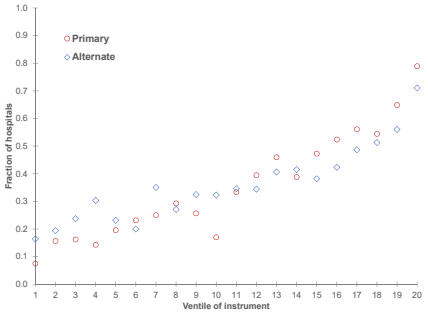


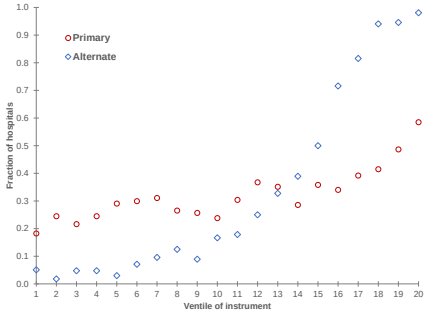
Figure A.5: Changes in demographics of arriving patients

Note: This figure plots coefficients obtained by estimating equation 5 with various demographics as dependent variables. These characteristics are used to construct the predicted readmission rate, $\tilde{R}_h$, used as an alternative instrument for expected penalty, as described in Section III.B. These regressions are estimated using the sample of all arriving patients in order to avoid conditioning on hospital admission decisions over the years 2008–2014. The regressions are estimated independently for heart attack, heart failure, and pneumonia, and the figure plots weighted average values. Models include hospital and year fixed effects. Standard errors are clustered by hospital and obtained using block bootstrap. Error bars indicate 95% confidence intervals.

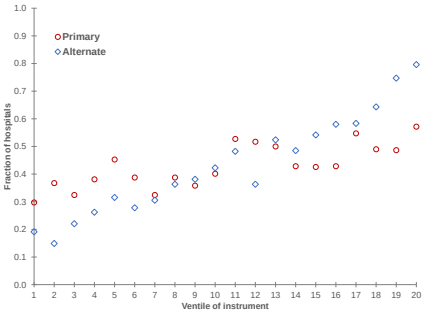
(a) High penalty risk



(b) High share of low income patients



(c) High share of minority patients



(d) Government owned hospitals

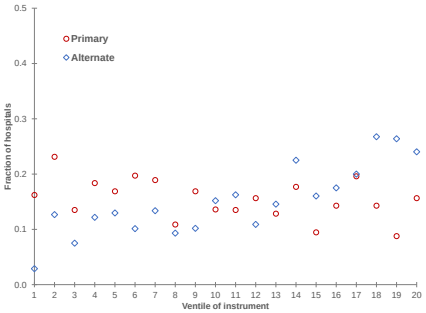


Figure A.6: Instruments and hospital features

Note: This figure illustrates how the two instruments move different types of hospitals toward high-penalty-risk, using data on pneumonia patients. Hospitals are divided into twenty equal-sized bins (ventiles) separately using the primary and alternate instruments. Each sub-figure plots the fraction of hospitals with a certain feature on the Y-axis, corresponding to each ventile on the X-axis. A hospital is defined as having high penalty risk or share of minorities, etc., if it is in the top tertile of hospitals for that attribute. The primary instrument is the predicted readmission rate in 2007 using patient co-morbidites, while the alternate instrument is the risk-adjusted readmission rate predicted by patient demographics excluded from CMS risk adjustment. Section III.B describes the instruments.

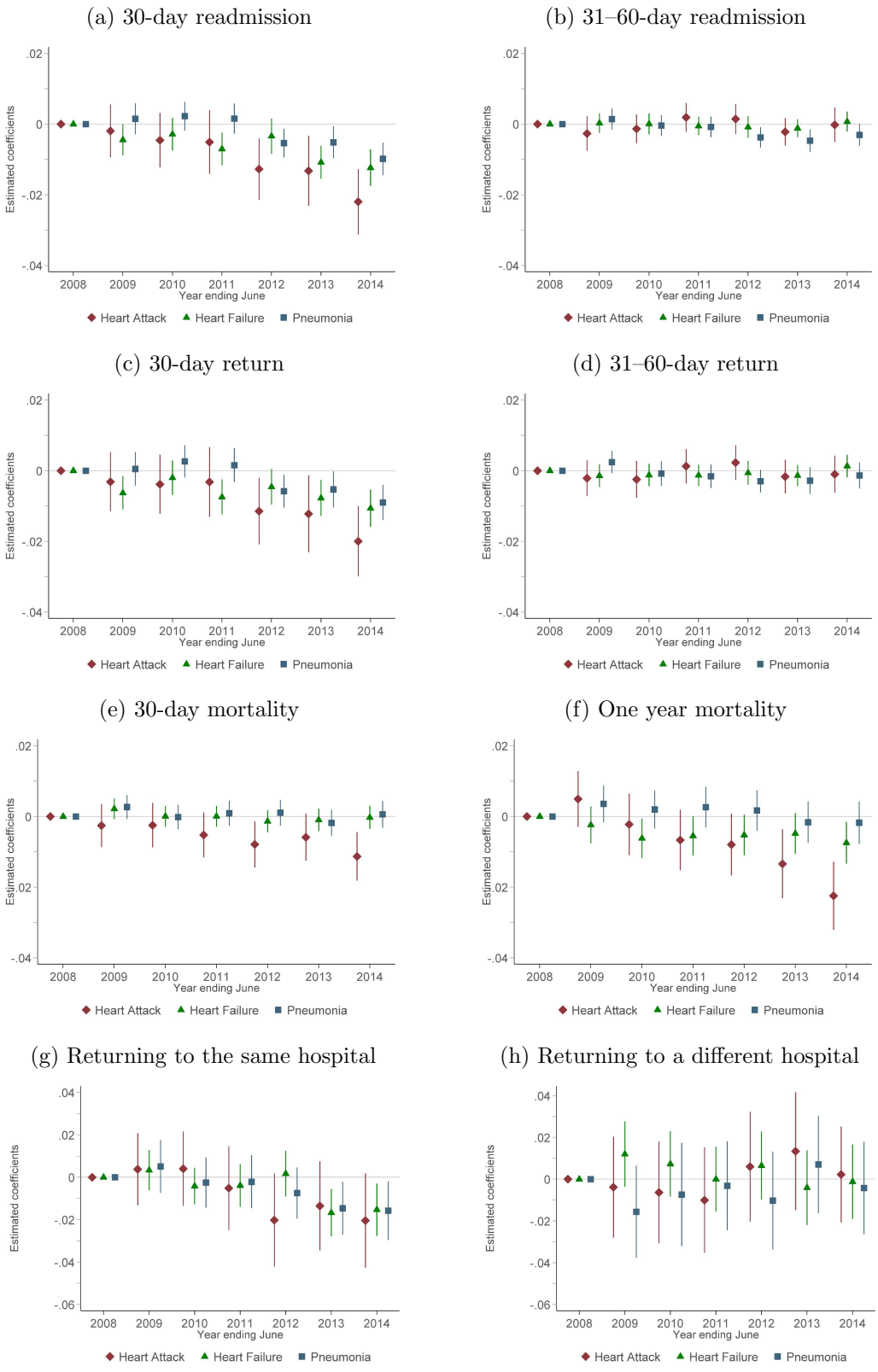


Figure A.7: Condition-specific event studies

Note: This figure presents estimated coefficients from equation 5 on all key outcomes. In all panels, 2008 is the reference year and the source of identifying variation across hospitals is the baseline instrument. These figures are exactly equivalent to figures 4,5a, 5b, and 7a except that they present coefficients for each condition separately. All models include hospital and year fixed effects. Standard errors are clustered by hospital. Error bars indicate 95% confidence intervals. To interpret magnitudes, note that the mean probability of readmission (return) pre-HRRP is 0.2 (0.27) and 0.06 (0.09) in 30 and 31–60 days, respectively. Thirty-day (one-year) mortality is 0.12 (0.35), and the probability of readmission for returning patients to the same (different) hospitals is 0.69 (0.74).

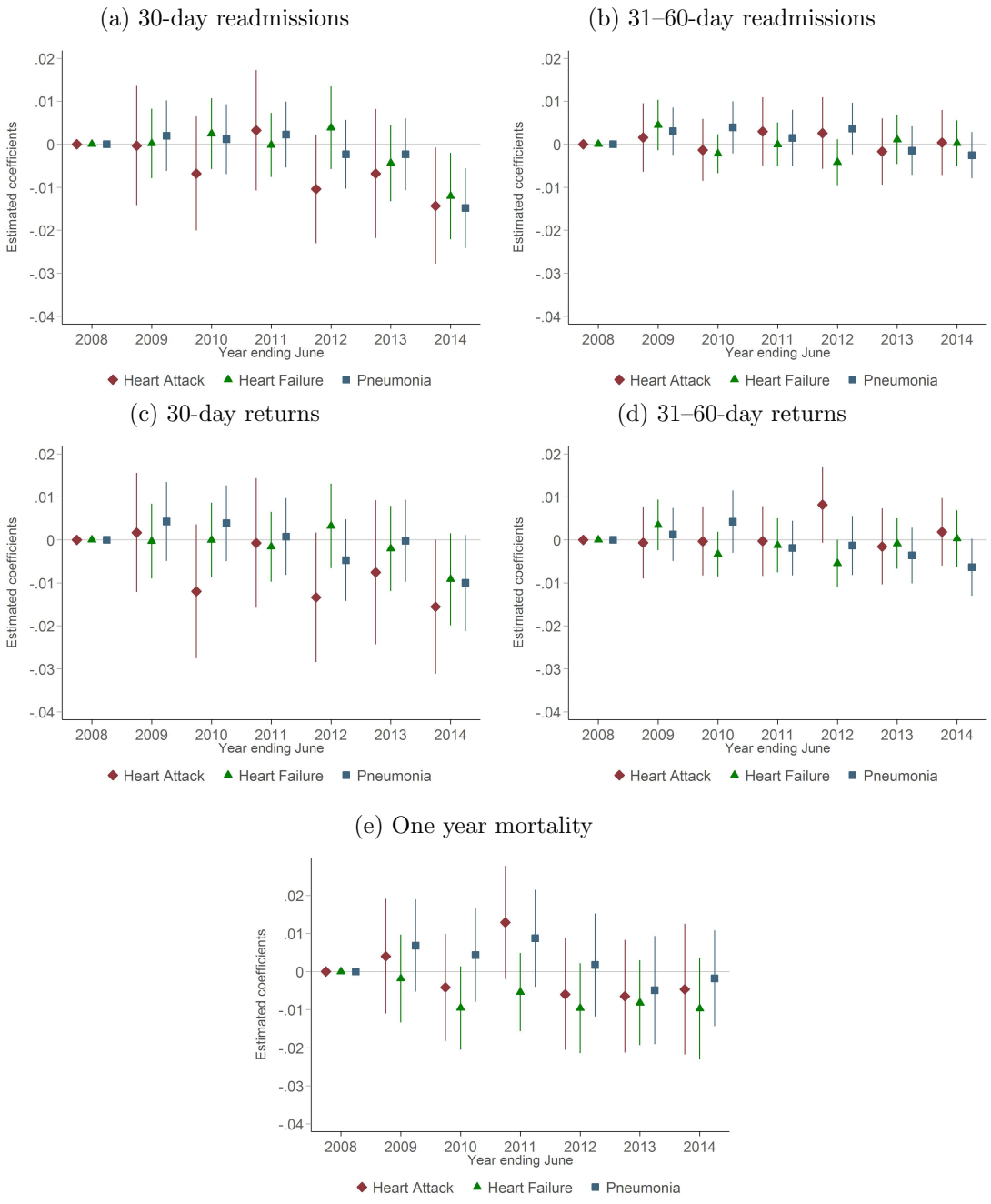


Figure A.8: Market-level results

Note: This figure presents dynamic effects on key outcomes using market-level (hospital referral region) variation in HRRP penalty probability (see Section VI.A). Coefficients are estimated for each year from 2008–14 using equation 5 for readmissions (panels a and b), returns (panels c and d), and mortality (panel e). In all panels, 2008 is the reference year and the source of variation is the HRR-level predicted readmission rate in 2007, obtained by aggregating the baseline instrument to the market level. These figures correspond to figures 4, 5a, and 5b. Models are estimated using the 20% Medicare sample described in Section B.2. Models include HRR and year fixed effects and are estimated for each condition separately. Standard errors are clustered by market. Error bars indicate 95% confidence intervals. To interpret magnitudes, note that the mean probability of readmission (return) pre-HRRP is 0.2 (0.26) and 0.06 (0.09) in 30 and 31–60 days following discharge, respectively. The mean one-year mortality rate is 0.35.

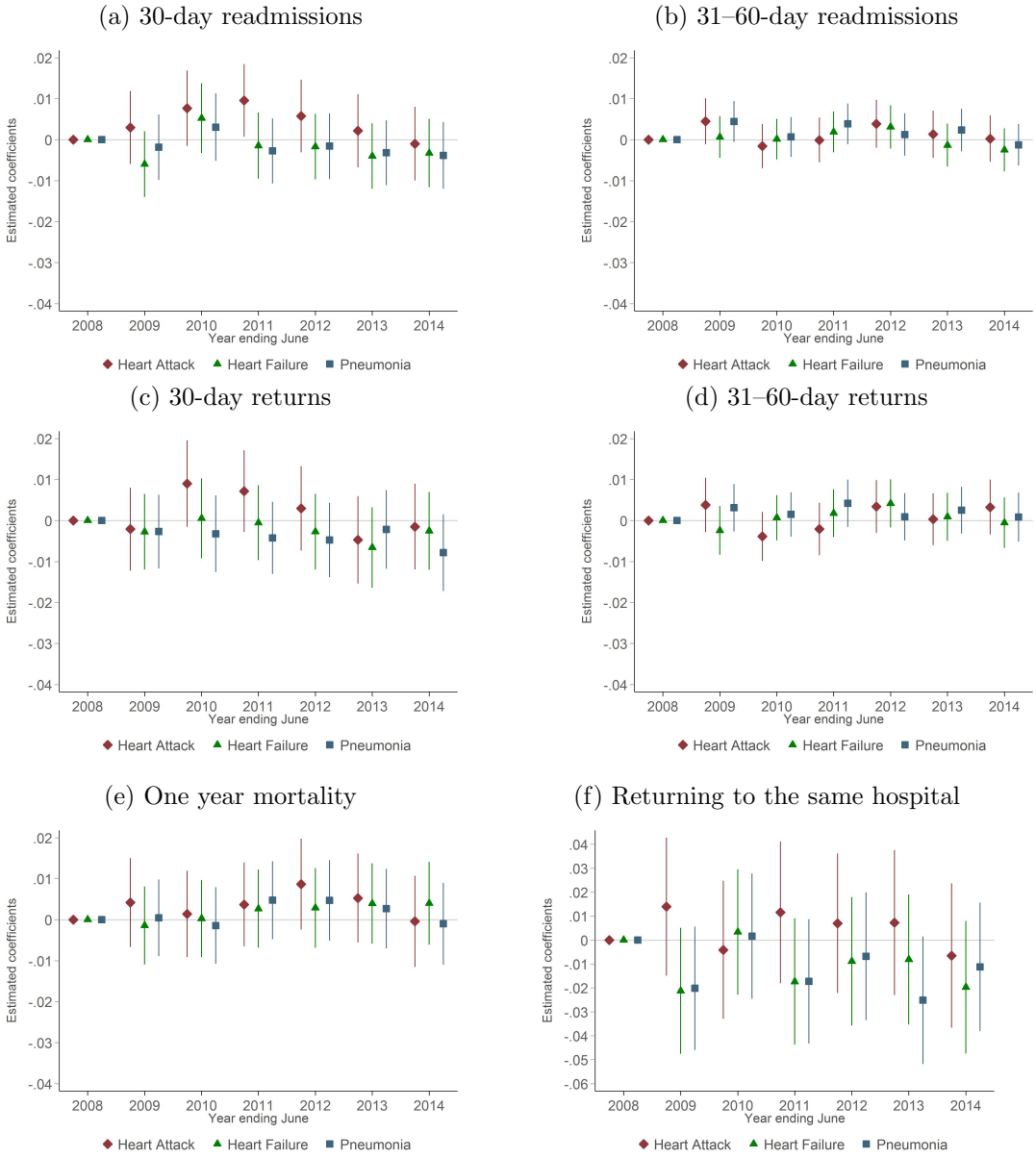


Figure A.9: Results for gastroenteritis patients

Note: This figure presents estimated effects on key outcomes for patients admitted with gastroentiritis (GI), a condition not targeted by HRRP, and relatively insulated from the targeted conditions. Models are estimated using the 20% Medicare sample described in Section B.2. Section B.4 describes how the GI cohort is constructed. Table A.1 column 4 presents descriptive statistics on this patient cohort. Estimates are obtained by estimating equation 5 with the baseline IV instrument as the source of identifying variation, except that it is applied to the GI cohort. Panel (f) presents effects on the probability of readmission on return to the same hospital within 30 days of discharge. Section VI.B discusses details. Models are estimated independently applying the penalty probability for each targeted condition. All models include hospital and year fixed effects. Standard errors are clustered by hospital. Error bars indicate 95% confidence intervals.

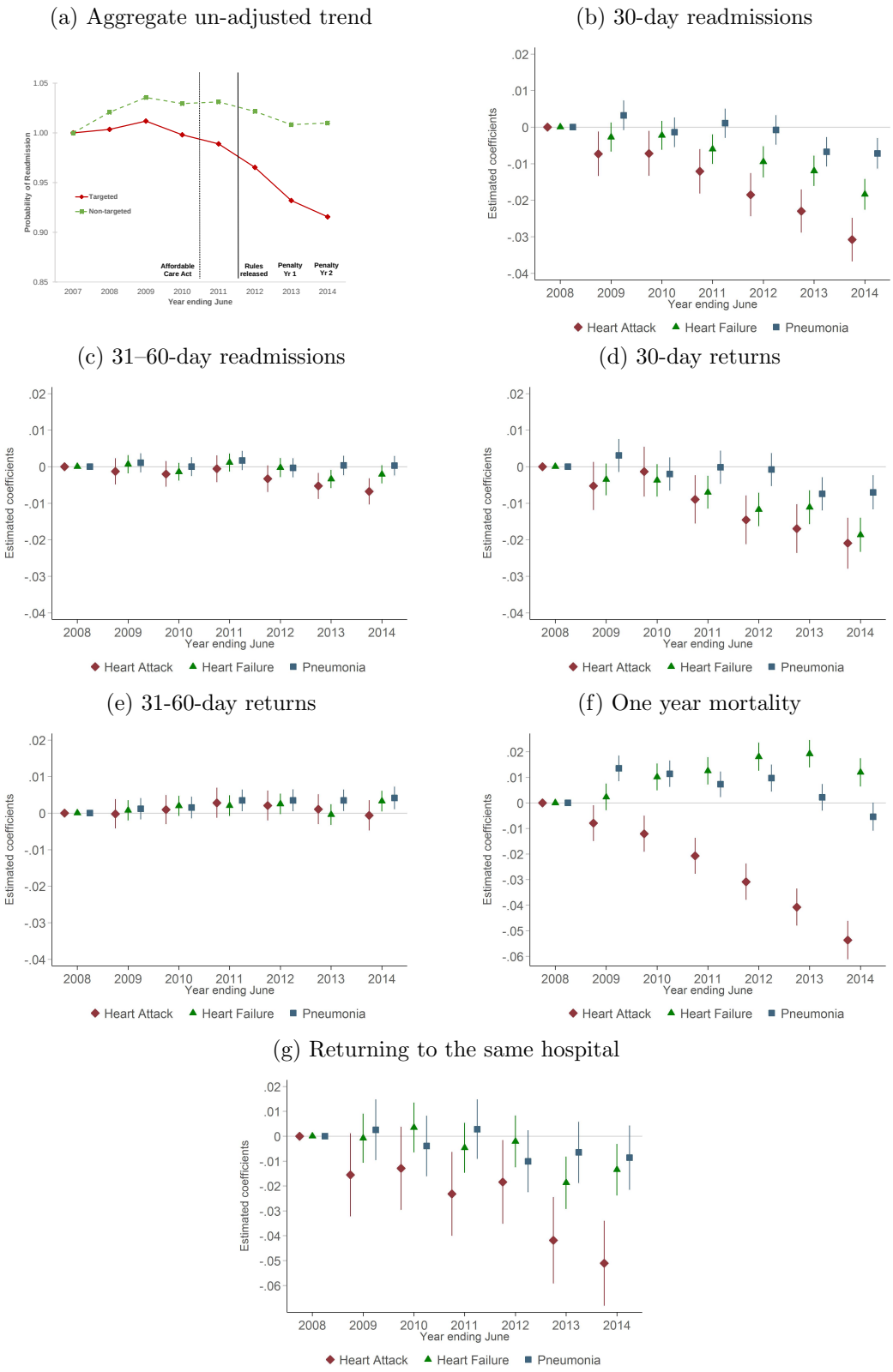
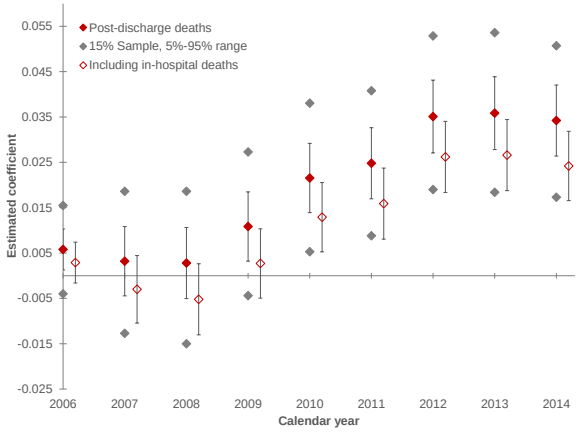


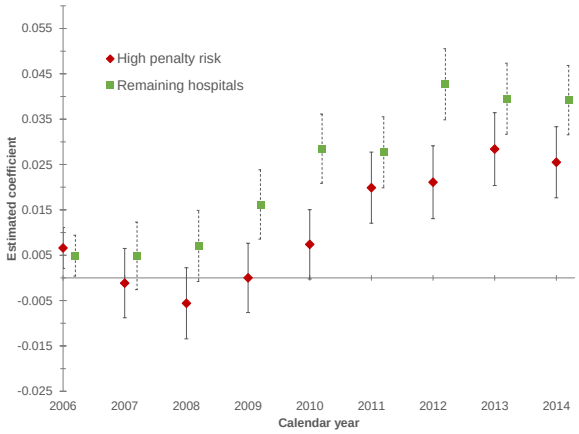
Figure A.10: Using non-targeted conditions as a comparison group

Note: This figure presents the aggregate trend in readmissions and estimated dynamic effects on key outcomes using a dynamic version of equation C.2. Models are estimated using the 20% Medicare sample described in Section B.2. The comparison group here is the 'others' cohort of patients admitted with conditions not targeted by HRRP, as described in Section VI.B. Table A.1 column 5 presents descriptive statistics on this patient group. All models include hospital and year fixed effects. Standard errors are clustered by hospital. Error bars indicate 95% confidence intervals.

(a) All hospitals



(b) High penalty risk vs. remaining hospitals



(c) Non-targeted ('Other') patients

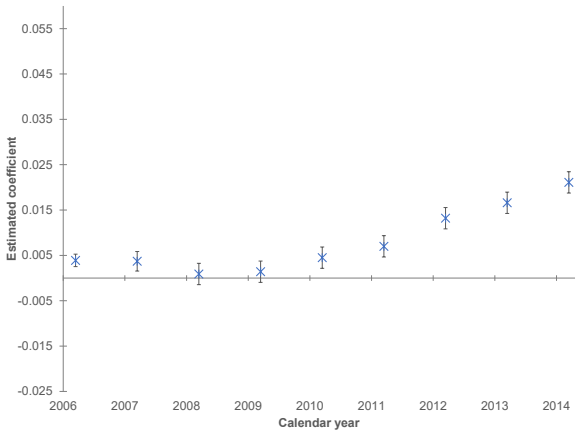


Figure A.11: Supplementary mortality analysis (heart failure)

Note: This figure presents supplementary analyses on one-year mortality for heart failure patients, discussed in Section VI.C. Models are estimated using the 20% Medicare sample described in Section B.2. Panel (a) examines changes in one-year mortality from calendar year 2006 through 2014, as well as the 95% range [grey diamonds] obtained when using a 15% random subset of hospitals. It plots estimated yearly effects for post-discharge mortality, as well as mortality including in-hospital deaths [hollow diamonds]. Panel (b) plots the estimated yearly effects separately for high-penalty-risk (top tertile) versus the remaining hospitals. Panel (c) presents corresponding estimates of post-discharge mortality for non-targeted patients—'other' patients used as a comparison group in Section VI.B. These coefficients are obtained by estimating models with patient risk factors and hospital and quarter-of-sample dummies. The quarterly coefficients are then aggregated to obtain annual estimates. Error bars depict 95% confidence intervals.

Table A.1: Additional descriptive statistics

	(1) Heart Attack	(2) Heart Failure	(3) Pneumonia	(4) GI	(5) Others
Panel A: Initial stay					
All initial stays	268,048	649,026	531,863	577,285	5,221,104
% admitted through ER	0.896	0.785	0.836	0.737	0.652
	(0.09)	(0.13)	(0.12)	(0.12)	(0.13)
Length of stay (days)	5.8	5.4	5.8	4.9	5.2
	(1.3)	(1.1)	(1.2)	(0.8)	(1.1)
Total payment (2016$)	17,618	9,883	9,063	8,102	9,389
	(5,223)	(3,127)	(2,474)	(2,024)	(2,785)
Base payment (2016$)	13,064	6,984	6,328	-	-
	(3,551)	(1,892)	(1,458)		
Index cases	242,828	623,098	502,876	562,066	5,075,936
P(Readmission) in 30 days	0.178	0.219	0.172	0.151	0.142
	(0.07)	(0.05)	(0.06)	(0.05)	(0.02)
Readm. Payment (2016$)	12,734	12,006	12,260	10,584	11,491
	(5,572)	(4,345)	(4,907)	(4,410)	(2,798)
Readm. LOS (days)	5.4	6.1	6.4	6.1	6.3
	(1.9)	(1.6)	(2.0)	(2.0)	(1.1)
Episode payment (2016$)	28,027	19,915	18,160	NA	NA
	(7,027)	(4,891)	(4,711)		
Panel B: Quality of care					
P(Return) in 30 days	0.250	0.286	0.237	0.216	0.208
	(0.07)	(0.06)	(0.06)	(0.05)	(0.03)
Mortality					
30-day	0.148	0.107	0.120	0.033	0.038
	(0.06)	(0.04)	(0.04)	(0.01)	(0.01)
One year	0.298	0.390	0.332	0.155	0.17
	(0.09)	(0.07)	(0.07)	(0.04)	(0.03)
Panel C: Admission decisions					
P(Admission)	0.990	0.876	0.838	0.377	0.368
	(0.03)	(0.09)	(0.10)	(0.10)	(0.11)
P(Readmission — Return)	0.646	0.685	0.656	0.628	0.618
	(0.14)	(0.11)	(0.13)	(0.13)	(0.08)
P(Observation status — Return)	0.110	0.082	0.074	0.084	0.078
	(0.12)	(0.08)	(0.09)	(0.09)	(0.05)

Note: This table uses the 20% random sample of Medicare beneficiaries to replicate descriptive statistics on key outcomes and describe additional outcomes and values for non-targeted conditions used in robustness checks. Section B.2 describes the 20% sample in detail. GI denotes gastroenteritis. Others is a group composed of conditions or procedures not targeted by HRRP. Section B.4 describes how these cohorts are constructed. The top half of panel A describes initial stays for the relevant condition, including when the patient died in-hospital. It lists the number of stays over the period July 2006–June 2014. Total payment includes patient cost sharing. Base payment is Medicare payment excluding medical education, capital cost, disproportionate share, and outlier payments. The bottom half of panel A describes index cases as defined by CMS. Payments include total Medicare and beneficiary payments made to providers. The values for readmissions payment and length of stay are in italics to reflect that they are computed only for patients who had readmissions. Panels B and C are identical to those in Table 1. All mean and standard deviation values pertain to the period July 2008–June 2011, which is the benchmark period used to determine penalty in the first year of the program, except in the case of observation stays where information is recorded starting in January 2011.

Table A.2: 20% versus 100% Medicare sample

	(1) Heart Attack	(2) Heart Failure	(3) Pneumonia
Panel A: Index cases			
Dep var: # Index cases in auxiliary			
# Index cases (orig.)	0.202	0.196	0.199
	(0.001)	(0.001)	(0.001)
Constant	0.422	1.705	1.139
	(0.292)	(0.389)	(0.334)
R-squared	0.97	0.97	0.96
Panel B: Readmission rate			
Dep var: Readm. rate in auxiliary			
Readm. Rate (orig.)	1.090	0.984	1.042
	(0.026)	(0.023)	(0.025)
Constant	0.001	0.013	0.003
	(0.004)	(0.005)	(0.004)
R-squared	0.54	0.39	0.39
Observations	1,824	2,899	2,916

Note: This table shows the correlation between hospital-level values in the 20% and 100% Medicare claims files. Both panels present coefficients from descriptive cross-section regressions using data aggregated to hospital level over 2009–11, the benchmark period for the first year of HRRP. Panel A shows coefficients from regressions with number of index cases in the original sample (as X) on number of index cases in the auxiliary sample (as Y). Panel B presents corresponding results with readmission rates as the variable of interest. The regression in panel B is weighted by number of index cases. In both cases, the sample included hospitals with 50 or more index cases over this period and is the sample used in the main analysis. Robust standard errors are displayed in parentheses.

Table A.3: 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. denotes fraction: for example "Frac. dual" indicates the fraction of patients amongst all arriving cases at a hospital in July 2006–June 2008 that were Medicaid-eligible. All regressions include HRR fixed effects. These coefficients are then used to predict the secondary instrument as described in Section III.B.

Table A.4: Impact on readmissions by interval

Interval	(1) Heart Attack	(2) Heart Failure	(3) Pneumonia	(4) Aggregate	(5) Mean
0–10 days	-0.0341	-0.0068	-0.0207	-0.0167	0.09
	(0.006)	(0.002)	(0.003)	(0.002)	
11–20 days	-0.0039	-0.0019	-0.0074	-0.0042	0.06
	(0.003)	(0.002)	(0.002)	(0.001)	
21–30 days	-0.0003	-0.0017	-0.0022	-0.0016	0.05
	(0.003)	(0.002)	(0.002)	(0.001)	
31–40 days	0.0005	0.0003	-0.0041	-0.0012	0.03
	(0.002)	(0.001)	(0.002)	(0.001)	
41–50 days	-0.0009	-0.0003	-0.0027	-0.0013	0.02
	(0.002)	(0.001)	(0.001)	(0.001)	
51–60 days	0.0024	0.0000	-0.0039	-0.0010	0.02
	(0.002)	(0.001)	(0.001)	(0.001)	
Observations	886,312	2,253,488	1,753,437	4,893,237	

Note: This table presents estimated effects on the probability of readmission over ten-day intervals post-discharge from the index stay. Each cell presents the main coefficient from a different IV regression. Columns 1–3 present results for each condition, and column 4 presents the weighted average values using number of admissions as weights. Each row presents estimates for a different time interval. Results are obtained via 2SLS using equation 3 with the baseline IV instrument, separately for each interval, i.e., using the probability of readmission in the interval (0–10 days, 11–20 days, and so on) as the outcome. All specifications include patient risk factors and hospital and year fixed effects. Standard errors are clustered by hospital and for the aggregate estimate are obtained using block bootstrap. The mean probability of readmission in each interval is presented in column 5. To interpret magnitudes, note that the mean probability of being penalized is approximately 0.5.

Table A.5: Intensity of care, and coordination

	(1) Heart Attack	(2) Mean	(3) Heart Failure	(4) Mean	(5) Pneumonia	(6) Mean
Panel A: 30 day episode						
Readmission charges ($)	-0.252	8,871	-0.0860	10,303	-0.229	8,182
	(0.131)		(0.060)		(0.072)	
Readmission payment ($)	-0.245	2,355	-0.0855	2,626	-0.195	2,125
	(0.116)		(0.0527)		(0.0631)	
All remaining charges ($)	0.0039	90,387	0.0203	52,765	-0.0295	47,551
	(0.025)		(0.016)		(0.017)	
All remaining payment ($)	0.0266	25,673	0.0228	17,289	-0.0423	16,036
	(0.0219)		(0.0157)		(0.0168)	
Panel B: Initial stay						
Facility (Part a) charge ($)	-0.0052	71,530	0.0049	38,081	-0.0334	34,766
	(0.028)		(0.017)		(0.018)	
Physician (Part b) charge ($)	0.250	6,002	-0.0044	2,677	-0.0707	2,173
	(0.082)		(0.046)		(0.048)	
Length of stay (days)	0.0122	5.8	-0.0163	5.4	-0.0069	5.8
	(0.023)		(0.014)		(0.014)	
Use of PCI	0.0443	0.38	0.0017	0.008	0.0002	0.001
	(0.015)		(0.001)		(0.001)	
Use of Coronary Care Unit	0.0497	0.34	0.0148	0.20	0.0035	0.10
	(0.025)		(0.021)		(0.009)	
Use of radiology imaging	0.0403	0.88	0.0146	0.98	0.0076	0.99
	(0.016)		(0.011)		(0.007)	
Use of supplies	-0.0217	0.85	0.0007	0.83	-0.0038	0.86
	(0.019)		(0.019)		(0.020)	
Supplies charges ($)	0.243	12,098	0.101	6,285	0.0414	2,239
	(0.075)		(0.051)		(0.053)	
Number of stays	193,943		464,388		371,991	
Panel C: Post-discharge						
Use of any post-acute care	0.0059	0.42	-0.0049	0.49	-0.0203	0.47
	(0.014)		(0.009)		(0.010)	
Post-acute care charges ($)	-0.102	14,927	0.0418	13,067	0.0421	16,361
	(0.061)		(0.028)		(0.037)	
P(post-acute care) within 7 days	0.0020	0.37	-0.0063	0.41	-0.0185	0.40
	(0.014)		(0.009)		(0.010)	
P(nursing home) within 7 days	-0.0208	0.16	0.0008	0.18	-0.0056	0.20
	(0.011)		(0.006)		(0.008)	
P(primary care visit) within 7 days	0.0029	0.36	0.0017	0.39	-0.0181	0.41
	(0.016)		(0.008)		(0.010)	
Use of prescription drugs	0.0098	0.45	0.0146	0.47	0.0189	0.47
	(0.016)		(0.010)		(0.010)	
Prescription drug payment ($)	0.103	566	-0.0151	478	0.0740	566
	(0.047)		(0.029)		(0.036)	
Number of stays	175,283		447,041		352,591	
Panel D: Process compliance						
Process of care score	0.0874	0.00	-0.113	0.00	0.131	0.00
	(0.063)		(0.046)		(0.066)	

Note: This table presents estimated effects on intensity of care and coordination post-discharge, discussed in Section IV.B. Each cell presents the IV coefficient from a different regression. Columns 1, 3, and 5 present estimated coefficients for heart attack, heart failure, and pneumonia patients, respectively. Columns 2, 4, and 6 present corresponding mean values. These models are estimated using equation 3 via 2SLS on the 20% sample of Medicare beneficiaries using the primary instrument. Section B.2 describes the sample. All spending variables are expressed in 2016 dollars, but the regressions use logged values. Panel A presents coefficients for total charges and payment (including out-of-pocket) for the 30-day episode for all index cases, split between readmission related (including zeros) and unrelated. Panel B presents coefficients pertaining to the initial hospital stay for all patients (including those dying in hospital) admitted with a targeted condition. Panel C presents estimated effects on the use of various types of care in the post-discharge period (through 30 days) and total episode charges only for index cases. Outcomes in italics (e.g., supplies charges) are modeled only for patients with positive values of that charge. Panel D presents IV coefficients on aggregated and normalized process of care scores at the hospital-year level. Section B.3.2 describes how these measures are constructed. Standard errors are clustered by hospital and presented in parentheses. To interpret magnitudes, note that the mean probability of being penalized is approximately 0.5.

Table A.6: Heterogeneity across hospitals

	(1) Heart Attack	(2) Heart Failure	(3) Pneumonia
A. Govt. ownership			
$Z_h \cdot 1(t \geq 2012)$	-0.1958	-0.1188	-0.2834
	(0.035)	(0.040)	(0.037)
$1(\text{Govt})_h \cdot 1(t \geq 2012)$	0.0012	0.0001	-0.0007
	(0.003)	(0.002)	(0.002)
Triple diff	0.0096	0.0631	0.0314
	(0.088)	(0.133)	(0.111)
B. Part of penalized hospital system			
$Z_h \cdot 1(t \geq 2012)$	-0.1909	-0.0744	-0.2943
	(0.036)	(0.046)	(0.043)
$1(\text{At-risk sys.})_h \cdot 1(t \geq 2012)$	-0.0012	-0.0031	-0.0029
	(0.002)	(0.002)	(0.002)
Triple diff	-0.0181	-0.1321	0.0757
	(0.080)	(0.095)	(0.081)
C. Hospital size			
$Z_h \cdot 1(t \geq 2012)$	-0.1764	-0.0696	-0.254
	(0.054)	(0.061)	(0.057)
$1(\text{Large})_h \cdot 1(t \geq 2012)$	0.0016	0.0000	-0.0015
	(0.002)	(0.001)	(0.001)
Triple diff	-0.0178	-0.0848	-0.0438
	(0.067)	(0.074)	(0.072)
D. Bed utilization			
$Z_h \cdot 1(t \geq 2012)$	-0.175	-0.0056	-0.216
	(0.097)	(0.094)	(0.106)
$1(\text{Low util})_h \cdot 1(t \geq 2012)$	0.0064	0.0038	0.0000
	(0.004)	(0.003)	(0.003)
Triple diff	0.147	-0.160	-0.0149
	(0.159)	(0.140)	(0.149)
Observations	10,872	17,222	17,348
P(Readm) 0-30 days	0.18	0.22	0.17

Note: This table presents reduced form effects on 30-day readmission rates using equation C.1 via OLS. These regressions are estimated at the hospital-year level. The panels test for differential responses by A. government owned hospitals, B. facilities owned/operated by a system that expects a large share of its hospitals to be penalized, C. hospitals in the top tertile by patient volume, and D. those in the bottom tertile by bed utilization, as discussed in Section IV.C. Section B.3.4 describes how these indicators are constructed. In each panel, the first row presents the coefficient on the interaction of post-HRRP and the baseline instrument. Hence, it is the D-D estimator of the main effect of the penalty. The second row captures the secular change post-HRRP among all hospitals with the feature of interest. The third row in each panel presents the triple-difference estimator, which captures the differential effect for hospitals with the feature of interest relative to the remaining hospitals, holding penalty-risk constant. To interpret magnitudes, note that the 90^{th}–10^{th} percentile difference in Z_h is about 0.05, and its standard deviation is 0.02. Standard errors are clustered by hospital and presented in parentheses.

Table A.7: Distortions during the initial stay

	(1) Heart Attack	(2) Heart Failure	(3) Pneumonia
Panel A: Probability of admission			
1. Average effect	-0.0042	-0.0183	-0.0153
	(0.004)	(0.005)	(0.008)
2. Selection on risk			
$\hat{R}_i$	-0.2610	0.144	0.920
	(0.283)	(1.062)	(0.837)
Inpatient history	-0.0007	0.0183	0.161
	(0.039)	(0.096)	(0.118)
Y Mean	0.99	0.88	0.84
Observations	194,834	535,097	448,476
Panel B: Vulnerable patients			
3. Black	0.0133	0.0017	-0.0023
	(0.007)	(0.006)	(0.005)
Y Mean	0.07	0.12	0.07
4. Dual eligible	-0.0008	0.0047	-0.0081
	(0.010)	(0.006)	(0.009)
Y Mean	0.18	0.24	0.27
Observations	192,459	464,388	371,991
Panel C: Up-coding			
5. Predicted score	-0.0034	-0.0021	-0.0053
	(0.002)	(0.001)	(0.001)
Y Mean	0.19	0.22	0.17
Observations	175,236	446,986	352,564
Panel D: Relabeling			
6. Share of target conds.	0.0006	-0.0067	0.0069
	(0.002)	(0.002)	(0.003)
Y mean	0.03	0.07	0.06
7. Target condition vol.	-0.103	-0.261	-0.034
	(0.0520)	(0.0379)	(0.0388)
8. Total volume	-0.130	-0.152	-0.104
	(0.032)	(0.027)	(0.031)
Observations	9,326	16,166	16,407

Note: This table presents results from testing various distortions during the initial admission. These models are estimated using the 20% Medicare sample described in Section B.2. Each cell of the table presents a coefficient from a different regression. Panel A row 1 presents average effects on the probability that a patient is admitted upon arrival. Estimates are obtained via 2SLS using equation 3 with the baseline instrument. Section III.B describes the instrument. Panel A row 2 presents triple-difference coefficients testing for evidence of selection on readmission risk at the ED using two measures of observed severity. The first relies on medical history and co-morbidity codes, while the second is an indicator for hospitalization in the previous six months. Estimates are obtained via OLS with the baseline instrument as the source of variation across hospitals. To interpret magnitudes for this model only, note that the 90^{th}–10^{th} percentile difference (s.d.) for Z_h is approximately 0.05 (0.02) across conditions. Models presented in panel A are estimated on the sample of all arrivals for the initial case. Panel B presents IV estimates for the share of Black and Dual eligible patients among all admissions. Panel C presents IV estimates for a measure of coded risk obtained using co-morbidity codes recorded on the index stay. Panel D presents IV estimates on the total volume of targeted conditions, the total Medicare volume, and the ratio of the two using hospital-year level data. The volume variables are simple counts of admissions and do not correspond to index cases. The share is computed as the ratio of target condition to total volume. All models include a vector of local market factors and hospital and year fixed effects. Standard errors are clustered by hospital. To interpret magnitudes (except for the triple-difference coefficients), note that the mean probability of being penalized across conditions is approximately 0.5.

Table A.8: Robustness: Additional variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	P(Return) in 30 days			P(Mortality) in 1 year			P(Readmission — Return) - same hosp.		
	Heart Attack	Heart Failure	Pneumonia	Heart Attack	Heart Failure	Pneumonia	Heart Attack	Heart Failure	Pneumonia
Main results (IV estimate)	-0.0286	-0.0061	-0.0268	-0.0266	-0.0029	-0.0086	-0.0384	-0.0225	-0.0488
	(0.006)	(0.003)	(0.004)	(0.006)	(0.004)	(0.005)	(0.014)	(0.007)	(0.009)
Panel A: Specification checks									
1. Demographics only	-0.0292	-0.0072	-0.0262	-0.0366	-0.0046	-0.0092	-0.0508	-0.0228	-0.046
	(0.006)	(0.003)	(0.004)	(0.007)	(0.004)	(0.005)	(0.016)	(0.008)	(0.010)
2. Time varying market controls	-0.0152	-0.0052	-0.0266	-0.0091	0.0015	-0.0112	-0.0223	-0.0245	-0.0246
	(0.013)	(0.007)	(0.008)	(0.015)	(0.008)	(0.009)	(0.035)	(0.016)	(0.022)
3. Perfect foresight	-0.0334	-0.0050	-0.0279	-0.0336	-0.0029	-0.0099	-0.0473	-0.0245	-0.0542
	(0.009)	(0.004)	(0.005)	(0.008)	(0.004)	(0.005)	(0.017)	(0.007)	(0.010)
4. Time-varying prob. penalty	-0.0118	-0.0053	-0.0198	-0.0122	-0.0031	-0.0076	-0.0305	-0.0224	-0.0273
	(0.005)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.008)	(0.004)	(0.005)
5. Alternate IV	-0.0236	0.0153	-0.0102	-0.0328	0.0046	0.0028	-0.0336	-0.0374	-0.0595
	(0.009)	(0.008)	(0.008)	(0.010)	(0.008)	(0.009)	(0.021)	(0.016)	(0.019)
6. Market-level variation	-0.0156	-0.00972	-0.0313	-0.0190	-0.0124	-0.0176	N/A	N/A	N/A
	(0.013)	(0.009)	(0.010)	(0.012)	(0.010)	(0.010)			
Panel B: Using other conditions									
7. Effects for GI patients	-0.0101	-0.001	-0.0054	0.002	0.005	0.009	-0.0253	-0.0101	0.0110
	(0.009)	(0.007)	(0.007)	(0.01)	(0.006)	(0.007)	(0.023)	(0.019)	(0.021)
8. DD against other conditions	-0.0110	-0.01	-0.0047	-0.0204	0.0066	-0.0056	-0.0198	-0.0107	-0.0086
	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.005)	(0.003)	(0.004)

Note: This table presents results from different robustness checks as discussed in Section VI. It covers three outcomes—the probabilities of return in 30 days (cols. 1–3), of death in 1 year (cols. 4–6), and of readmission for patients returning to the same hospital within 30 days (cols. 7–9). The top row presents the preferred coefficients presented in tables 3 and 4 for ease of comparison. There are no coefficients on the probability of readmission on return with market-level variation in penalty probability (row 6) since there is no variation in penalty across hospitals in the same market. To interpret coefficients in rows 1–7, note that the mean probability of being penalized is approximately 0.5. Row 8 presents results using a difference-in-difference research design with non-targeted 'other' conditions as the comparison group. All specifications except in row 6 include hospital and year fixed effects. Standard errors are clustered by hospital and presented in parentheses. The models in row 6 are organized around HRR as the panel variable instead of hospital. Coefficients in rows 2, 6, 7, and 8 are estimated using the 20% Medicare sample described in Section B.2.

Table A.9: Effects for patients with other conditions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Targeted conditions			**Non-targeted conditions**				
	Heart Attack	Heart Failure	Pneumonia	Cardiac Dys.	Stroke	COPD	GI	Renal
Panel A: Readmissions								
1. 0-30 days	-0.0349	-0.01	-0.0281	-0.0260	-0.0225	-0.0053	-0.0086	0.0009
	(0.006)	(0.003)	(0.003)	(0.009)	(0.010)	(0.013)	(0.020)	(0.014)
Y-mean	0.18	0.22	0.17	0.10	0.10	0.15	0.12	0.16
Panel B: Mortality								
1 Year	-0.0266	-0.0029	-0.0086	-0.0002	0.0029	-0.0015	-0.0016	0.0387
	(0.006)	(0.004)	(0.005)	(0.009)	(0.012)	(0.017)	(0.014)	(0.017)
Y-mean	0.30	0.39	0.33	0.14	0.23	0.26	0.16	0.40

Note: This table presents estimated effects on 30-day readmissions (panel A) and one-year mortality (panel B) for patients with conditions targeted (cols. 1–3) and not targeted (cols. 4–8) by HRRP. Columns 1–3 reproduce coefficients from Tables 2 and 3B for the reader's convenience. Columns 4–8 present coefficients obtained by estimating the baseline IV equation 3 using the same approach as for the targeted conditions, except that I assign hypothetical penalty status to hospitals for these non-targeted conditions by computing their risk-adjusted readmission rates during the first benchmark period 2009–11 and applying HRRP's rules. I also predict readmission rates in 2007 using patient covariates and use that as the instrument. The condition cohorts are identified using principal discharge diagnosis codes. Column 4 presents results for cardiac dysrythmia (dx 427), column 5 for stroke or cerebrovascular disease (dx 430–438), and column 8 for renal failure (dx 584–586). Patients with COPD are identified following the approach used by CMS to compute readmissions. Appendix B.4 provides more details on how these cohorts are constructed and lists all the diagnosis codes for gastroenteritis (GI, col. 7). To mitigate the potential for contamination when examining the non-targeted conditions, I exclude patients who also appear in the sample for any of the three targeted conditions.

Table A.10: Supplementary analysis on mortality

Panel A: Adjusting for HVBP	(1) Heart Attack	(2) Heart Failure	(3) Pneumonia
Main IV estimate	-0.0266	-0.0029	-0.0086
	(0.006)	(0.004)	(0.005)
Incl. HVBP penalty factors	-0.0254	-0.0037	-0.0069
	(0.006)	(0.004)	(0.004)

Panel B: Time series approach	(1)	(2)	(3)	(4) Heart Failure	(5)	(6)	(7)	(8) Other
	All hospitals	Sample 5%	Sample 95%	All cases	High penalty risk	Remaining	(5)-(6)	
I. Jul 2006 - Mar 2010	0.0057	0.0047	0.0065	-0.0012	-0.0014	0.0089		0.0021
	(0.003)	(0.010)	(0.007)	(0.003)	(0.006)	(0.004)		(0.001)
II. Apr 2010 - Sep 2012	0.0272	0.0269	0.0313	0.0183	0.0159	0.0333		0.0075
	(0.004)	(0.010)	(0.007)	(0.004)	(0.007)	(0.005)		(0.001)
III. Oct 2012 - Dec 2014	0.0364	0.0284	0.0448	0.0269	0.0286	0.0408		0.0194
	(0.004)	(0.010)	(0.007)	(0.004)	(0.007)	(0.005)		(0.001)
Estimated effects:								
a) II - I	0.0215	0.022	0.025	0.0195	0.0173	0.0244	-0.0071	0.0054
	(0.002)	(0.003)	(0.001)	(0.002)	(0.003)	(0.002)	(0.003)	(0.001)
b) III - I	0.0308	0.024	0.038	0.0281	0.03	0.0319	-0.0019	0.0173
	(0.002)	(0.000)	(0.000)	(0.002)	(0.003)	(0.002)	(0.004)	(0.001)

Note: This table provides additional results on mortality effects. Panel A tests robustness to including the hospital's penalty risk under the Value Based Purchasing (HVBP) program. I include the total performance score (TPS) for 2013-14. This is the first year when hospital mortality rates are considered by HVBP. The TPS varies from 0–100, but it is normalized to have a mean of zero and s.d. of one in this analysis. Section B.3.3 describes how these variables are constructed. All models control for time varying hospital case-mix characteristics and hospital and year fixed effects. Standard errors are clustered by hospital and presented in parentheses. To interpret magnitudes, note that the mean probability of being penalized is approximately 0.5. Panel B presents additional results on one-year mortality for heart failure patients. These models are estimated using the 20% Medicare sample. Section B.2 describes this sample. Except for column 4, all others present results using post-discharge mortality (pre-HRRP mean 0.36). Column 4 includes in-hospital deaths (mean 0.39). Column 8 presents results for Others, the non-targeted patients discussed in Section VI.B. The top 3 rows compute mean changes in mortality relative to 2006Q3, with standard errors in parentheses. The bottom rows present the estimated changes from period I (pre-HRRP) to period II (announcement), and period I to III (implementation). Section C.5 provides more details.

Table A.11: Acute care vs. Critical Access Hospitals

	(1) Heart Attack	(2) Heart Failure	(3) Pneumonia	(4) GI	(5) Others	(6) Total
1. Number of hospitals						
General Acute Care	3,053	3,164	3,177	3,183	3,322	3,339
Critical Access Hospitals	866	1,193	1,237	1,208	1,244	1,246
2. Hospital volume per year						
General Acute Care	14	30	22	24	193	584
Critical Access Hospitals	1	4	6	4	23	132
3. Proportion below 15 admissions						
General Acute Care	0.35	0.10	0.10	0.14	0.03	NA
Critical Access Hospitals	0.69	0.63	0.43	0.71	0.07	NA
4. CAH share of patients	2.2%	5.0%	10.0%	5.4%	4.2%	7.8%

Note: This table presents descriptive statistics to illustrate differences between two types of hospitals—General Acute Care and Critical Access. These values are obtained from the 20% Medicare sample and pertain to the period July 2008–June 2011, the benchmark period to determine the penalty for the first year of HRRP. Patients are assigned to conditions using the ICD 9 principal discharge diagnosis code, as described in Section B.4. GI denotes gastroenteritis. Others refers to the composite group of non-targeted patients, discussed in Section VI.B. Total presents statistics across all patients discharged by the hospital, regardless of condition. Row 3 presents the proportion of hospitals of each type that have fewer than 15 admissions in a condition over the benchmark period 2009–11. Section B.5 describes differences between GACs and CAHs in more detail.