(Re-)Examining the Impact of Hospitals' Participation in Voluntary Value-Based Programs Through Matching and Placebo Tests

Jason Buxbaum, MoonSoo Choi, John Giardina *

May 11, 2019

Abstract

We replicate and extend the results of Ryan et al. (2017), which examined the association between hospital performance in the mandatory Hospital Readmission Reduction Program (HRRP) and participation in voluntary value-based reforms administered by the Centers for Medicare & Medicaid Services (CMS). Ryan et al. (2017) reported an association between participation in the voluntary programs and greater decreases in readmission rates for acute myocardial infarction, heart failure, and pneumonia in the HRRP. We conducted additional analyses to test whether the Ryan et al. findings should be interpreted causally, or are instead driven by unobserved hospital characteristics and secular time trends. Specifically, we employed two approaches to identify possible confounding: (i) using a matching procedure for panel data to calculate a difference-in-differences effect of voluntary program implementation, and (ii) conducting a series of placebo tests using other publicly available data from CMS. The results of both analyses suggest that implementation of the voluntary value-based reforms do not have a causal effect on readmission reduction in association with the HRRP. Further work is needed to better understand the interactions between participation in voluntary CMS programs and mandatory CMS programs more generally.

*The authors can be reached at jasonbuxbaum@g.harvard.edu, mochoi@hbs.edu, and jgiardina@g.harvard.edu. We would like to thank Gary King, Max Goplerud, and Meredith Dost for leading the Government 2001: Advanced Quantitative Political Methodology course in the Spring 2019 semester and for the invaluable feedback they provided on this paper. We are also grateful to the three anonymous peer reviewers who shared valuable suggestions. Finally, we are grateful to Andrew Ryan, who generously provided us with the data analyzed in this paper.

1 Introduction

A central concern of health policy in the US is improving the value of care provided by the healthcare system. This requires policymakers to find ways to improve the quality and safety of care patients receive, while also limiting growth in the cost of care. To this end, the Centers for Medicare & Medicaid Services (CMS) has devoted significant resources to a range of mandatory and voluntary physician- and hospital-facing initiatives intended to improve value [1], [2]. In 2010, the Affordable Care Act [3] introduced the Hospital Readmissions Reduction Program ("HRRP") in an effort to improve the quality of patient care by preventing excess patient readmissions at acute care hospitals. Under the HRRP, hospitals with risk-adjusted 30-day readmission rates for Medicare patients admitted with certain conditions that exceed the national average are financially penalized.[2] Three of the conditions included in the HRRP are acute myocardial infarction (AMI), heart failure, and pneumonia; the program is designed to incentivize hospitals to improve care for patients with those conditions so that they are not readmitted in the future. While the vast majority of acute care hospitals must participate in the HRRP, participation in the voluntary programs is far from universal.

The interactions among these programs – especially the extent to which hospital participation in one or more programs delivers additional benefit beyond what would otherwise be achieved through participation in a mandatory program – is not well-understood. A recent paper by Ryan et al. (2017) investigated this issue in the context of the HRRP and four voluntary programs: incentives for meaningful use (MU) of electronic health records (EHRs), the Medicare Shared Savings Program (MSSP), the Pioneer Accountable Care Organization (ACO) program, and the Bundled Payment for Care Initiative episode-based payment program (BPCI).[1] The Ryan et al. research sought to determine if the combination of the HRRP with other value-based initiatives improved readmission rates to a greater degree than just the HRRP alone.

The structure of the paper is as follows. We first provide a brief overview of the Ryan et al. paper; we discuss their research questions, methodology, data sources, and their findings. We then take two approaches to studying the suitability of the Ryan et al. model.

First, we employ matching based on fixed hospital characteristics in order to see whether there is a substantial change in the treatment effect – i.e., the effect of hospitals' participation in voluntary value-based programs on readmission rates – compared to the initial estimates obtained through the parametric analyses conducted by Ryan et al. Because hospital participation in the matching programs varied with time, we used the time-varying matching approach proposed in Imai, Kim, and Wang (2019) [4]. Second, we conduct a

[1] MSSP is also an ACO program. Ryan et al. combine participation in Pioneer or MSSP for many of their analyses and label the category "ACO." We follow their example where appropriate.

series of placebo tests using the Ryan et al. models with other publicly available quality measures from CMS over the time period of interest. Given the nature of the voluntary programs, it is unlikely that these specific quality measures would change significantly due to participation in these programs.

The results of these two extensions could call into question the Ryan et al. conclusion that the HRRP and the voluntary programs are complementary with respect to readmission reduction. Specifically, finding that either:

1. voluntary program implementation does not impact readmission reduction during the HRRP era through the matching exercise, or

2. implementation of voluntary programs is associated with significant changes in the quality measures during the HRRP era through the placebo tests,

would tend to undermine confidence in the Ryan et al. estimation strategy. Finding either null effects through matching or significant effects through placebo tests would suggest that unobserved differences in hospital characteristics may explain both participation in the voluntary programs as well as greater success in readmission reduction.

2 Overview of Ryan et al. (2017)

To evaluate the effect of participation in four voluntary programs (MU, MSSP, Pioneer ACO, and BCPI) on readmission rates, Ryan et al. uses a set of hospital-level data on hospital performance under HRRP publicly available at the Centers for Medicare and Medicaid Services (CMS). These datasets were then merged with information on participation in four voluntary programs (MU, MSSP, Pioneer ACO, and BCPI) as well as a set of fixed hospital characteristics that is common in the literature (e.g., urban/rural location, share of Medicaid discharges, etc.) [6], [5]. In analyzing the data, Ryan et al. uses an interrupted time-series (ITS) analysis in examining the association between the readmissions rate and the HRRP by estimating three separate linear fixed-effects models for AMI (Acute Myocardial Infarction, or heart attack), heart failure, and pneumonia at the hospital level. In particular, they estimate the following linear regression model:

$$R_{it} = \beta_0 + \beta_1 t + \beta_2 \omega_t T_t + \beta_3 \omega_t T_t Z_i + \alpha_1 \omega_t T_t MU_{it} + \alpha_2 \omega_t T_t ACO_{it} + \alpha_3 \omega_t T_t BCPI_{it} + \alpha_4 \omega_t T_t MU_{it} ACO_{it} + \alpha_5 \omega_t T_t MU_{it} BCPI_{it} + \alpha_6 \omega_t T_t MU_{it} ACO_{it} BCPI_{it} + \lambda_i + \varepsilon_{it}$$

Here, R_{it} is the 30-day readmission rate for one of the conditions of interest for hospital i at time t; T_t is an indicator variable for the implementation for the HRRP program (since the program is mandatory, it has the same value for all the hospitals at time t); MU_{it}, ACO_{it}, and $BCPI_{it}$ are indicator variables for the implementation of the meaningful use, ACO, and BPCI programs, respectively, by hospital i at time t (the MSSP and Pioneer ACO programs are grouped together for this analysis); Z_i are time-invariant hospital characteristics including teaching hospital status, number of beds, disproportionate patient percentage, and urban location; λ_i are hospital fixed-effects; and ε_{it} is an error term (assumed to have a Normal distribution with mean zero). Additionally, Ryan et al. include the weighting term ω_t in order to account for the fact that some of the observations overlap with the implementation of the HRRP; ω_t equals the proportion of the observation at time t where the HRRP was active. The effect of this parameter is to weight changes in the readmission rate more highly when estimating the effect of the HRRP if the program was only partially implemented for that observation. The consideration of this weighting parameter becomes important when developing the matching extension presented below.

The primary quantities of interest in this specification are the α parameters. These represent the change in the effect of the HRRP after implementation of one of the voluntary programs.[2] For example, α_1 represents the change in the effect of the HRRP on readmission rates after only implementing a meaningful use program and $\alpha_1 + \alpha_2 + \alpha_4$ represents the change in effect after implementing both a meaningful use program and an ACO program. The results from Ryan et al. for these quantities of interest are presented in Table 1. They find that hospitals' participation in MU, ACOs, and BPCI is associated with a reduction in 30-day risk-standardized readmission after the implementation of HRRP, compared to those of the hospitals that participated in none of the programs.

Ryan et al. recognize that these quantities of interest only represent the association between the implementation of the voluntary programs, but to inform policy decisions it is often necessary to make a causal interpretation of these effects. The main concerns in making such a causal interpretation of these associations are that (a) there is a self-selection effect where hospitals more likely to participate in voluntary programs are also more likely to improve quality of care in general, and (b) there are secular trends in readmission rates across all hospitals correlated with the implementation of the voluntary programs. By including time-invariant hospital characteristics and a linear time trend in their model, Ryan et al. attempt to control for hospital self-selection into the voluntary programs and any secular trends in readmission

[2] All the voluntary programs launched after the HRRP, which is why terms for the effect of the voluntary programs alone are not included in the specification.

Table 1: Estimates of the Change in the Estimate of Hospital Readmission Reduction Program Effect Across Participation in Voluntary Reforms

	Percentage Point Change in Readmission Rate (95% CI for 2-Sided Test)		
	AMI	Heart Failure	Pneumonia
Number of Observations	15456	22280	22696
Program Participation:			
Only meaningful use (α_1)	-0.78 (-0.89, -0.66)	-0.97 (-1.08, -0.86)	-0.56 (-0.65, -0.47)
Only ACO programs (α_2)	-0.94 (-1.29, -0.59)	-0.83 (-1.26, -0.41)	-0.59 (-1.00, -0.18)
Meaningful use and BPCI ($\alpha_1 + \alpha_3 + \alpha_5$)	-1.02 (-1.22, -0.81)	-1.16 (-1.40, -0.92)	-0.71 (-0.90, -0.51)
Meaningful use and ACO ($\alpha_1 + \alpha_2 + \alpha_4$)	-1.00 (-1.18, -0.83)	-1.30 (-1.48, -1.11)	-0.83 (-0.99, -0.68)
All Programs ($\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 + \alpha_5 + \alpha_6$)	-1.27 (-1.58, -0.97)	-1.64 (-2.02, -1.26)	-1.05 (-1.32, -0.78)

rates. This allows them to argue the implementation of the voluntary programs actually caused the observed reduction in readmission rates.

To further support this claim, Ryan et al. compare the pre-treatment readmission rate in hospitals that implemented a voluntary program with the rate in hospitals that had not yet implemented the program. They find no difference in pre-treatement readmission rates between the two groups. Additionally, Ryan et al estimated models with randomly assigned participation dates for the voluntary programs, and found that the effect of program "participation" in these models was almost never statistically significant. Ryan et al. argued that these results support a causal interpretation of the effect of voluntary program participation on the impact of the HRRP on readmission rates and "lend support for CMS's multi-pronged strategy to improve hospital value." The remainder of this paper is focused on testing if this interpretation is correct, or if the effect of voluntary program participation on patient quality of care is more nuanced – or perhaps non-existent.

3 Extension 1: Estimating Effects with Matching

3.1 Deriving Quantities of Interest for Matching Procedure

In order to better account for selection bias and secular trends in readmission rates, we estimated the primary quantities of interest from Ryan et al. using a matching method. However, because the primary quantity of interest in Ryan et al. is essentially an interaction effect – and not a straightforward treatment effect – care

must be taken to define the quantity of interest for a matching procedure. To do this, it is helpful to return to a generalized version of the model of readmission rates used by Ryan et al.

We construct the model as follows: for hospital i at time t the expected readmission rate R_{it} is determined by an HRRP treatment indicator T_{it} (weighted for the proportion of the time t observation under the treatment, ω_t), hospital level fixed-effects λ_i, time trendγ_t, and the effect of the HRRP on readmission rates for hospital i at time t, η_{it}:

$$E[R_{it}|i,t] = \beta + \omega_t \eta_{it} T_{it} + \lambda_i + \gamma_t$$

Note that this model is actually more robust than the specification used by Ryan et al., which assumed a linear time trend. For the matching approach, we are able to take advantage of the parallel trends assumption for the difference-in-differences estimator, which is discussed below. Ryan et al. further specify the effect of the HRRP as follows, where P_{it} is an indicator for participation in a voluntary program (or group of voluntary programs) and Z_i indicate time-invariant hospital characteristics:

$$\eta_{it} = \alpha_0 + \alpha_1 P_{it} + \alpha_2 Z_i$$

The primary quantity of interest investigated by Ryan et al. is α_1: the change in the effect of HRRP on readmission rates from implementing voluntary programs (e.g., ACO programs or EHR meaningful use). In order to recover this quantity with a matching procedure in the context of this time series data, we used the difference-in-differences approach proposed by Imai, Kim, and Wang (2019) to take time trend into account.

First see that taking the difference of the readmission rates for a given hospital i before and after implementation of the HRRP (at times t' and t respectively), gives the following relationship:

$$E[R_{it}|T_{it} = 1] - E[R_{it'}|T_{it'} = 0] = \omega_t(\alpha_0 + \alpha_1 P_{it} + \alpha_2 Z_i) + \gamma_t - \gamma_{t'}$$

Now, suppose that hospital i implements the voluntary program at time t^*, so that we can recover the following difference in readmission rates before and after implementation of the program (since all the voluntary programs were implemented after HRRP):

$$\begin{aligned}
&[E[R_{it^*}|T_{it^*} = 1, P_{it^*} = 1] - E[R_{it'}|T_{it'} = 0]] - [E[R_{it^*-1}|T_{it} = 1, P_{it^*-1} = 0] - E[R_{it'}|T_{it'} = 0]] \\
&= E[R_{it^*}|T_{it^*} = 1, P_{it^*} = 1] - E[R_{it^*-1}|T_{it} = 1, P_{it^*-1} = 0] \\
&= \omega_{t^*}\alpha_1 + (\omega_{t^*} - \omega_{t^*-1})(\alpha_0 + \alpha_2 Z_i) + \gamma_{t^*} - \gamma_{t^*-1}
\end{aligned}$$

Then, the difference-in-differences approach can be used to account for the time trend by matching hospital i with hospital j at time t^*, where hospital j has not implemented the voluntary program at time t^*. By assuming that the time trends (i.e., $\gamma_{t^*} - \gamma_{t^*-1}$) and time-invariant hospital characteristics (i.e., Z_i and Z_j) for the matched hospitals are equal, we can recover α_1, which is the primary quantity of interest:

$$\begin{aligned}
&(1/\omega_{t^*})[[E[R_{it^*}|T_{it^*}=1, P_{it^*}=1] - E[R_{it^*-1}|T_{it}=1, P_{it^*-1}=0]] \\
&- [E[R_{jt^*}|T_{jt^*}=1, P_{jt^*}=0] - E[R_{it^*-1}|T_{it}=1, P_{it^*-1}=0]]] \\
&= \alpha_1
\end{aligned}$$

Therefore, following Imai, Kim, and Wang (2019), we can estimate the average treatment effect on the treated by averaging over difference-in-differences for some set of matched pairs M:

$$\hat{\alpha}_1 = \frac{\sum_{(i,j)\in M} \sum_{t=2}^{T} P_{it}(1-P_{it-1})(1/\omega_t)(R_{it} - R_{it-1} - (R_{jt} - R_{jt-1}))}{\sum_{(i,j)\in M} \sum_{t=1+L}^{T} P_{it}(1-P_{it-1})}$$

Here, the term $P_{it}(1-P_{it-1})$ is an indicator variable so that only periods when a hospital first implemented a voluntary program are included in the estimator.

This shows that it is possible to recover the quantity of interest from Ryan et al. with a matching procedure by using the implementation of the voluntary program as the treatment variable and the readmission rate as the outcome while also weighting for the proportion of each treated observation occurring after the implementation of HRRP. The key advantage of this matching approach is its ability to control for time trends compared to the the parametric approach employed by Ryan et al. Whereas the parametric approach described above assumed a linear time trend and fixed hospital effects, the matching approach only assumes a parallel time trend between matched hospitals. This approach can better control for secular trends in readmission rates and time-varying hospital characteristics that could drive self-selection (to the extent that the matched hospitals are similar in those time-varying characteristics). Additionally, it is important to note that this also means the matched difference-in-differences estimator is more robust to changes in the changes to the specification of the model from Ryan et al. For example, including lagged outcomes (i.e., lagged readmission rates) or transforming the time-invariant hospital characteristics (e.g., adding quadratic terms) in the model would lead to a different specification for parametric estimation, but doing so would still lead to the same difference-in-differences estimator described above as long as the hospitals were matched on these characteristics.

3.2 Calculating Quantities of Interest for Matching Procedure

In order to calculate the matched difference-in-difference estimator, the `PanelMatch` package for `R` developed by Imai, Kim, and Wang (2019) was used. This package essentially calculates a more flexible version of the matched estimator derived above, allowing the use of multiple matched controls (i.e., not just one-to-one matching) and the calculation of effects for the periods following implementation of the program. To create the matched sets, the package strictly matches observations based on lagged voluntary program implementation (i.e., it matches on treatment history), and then refines those matched sets using a distance metric or propensity score on a set of given covariates.

Using the `PanelMatch` package, the following difference-in-difference estimator was calculated, where hospital j is a weighted average of the controls matched to the treated hospital i:

$$\hat{\beta}_{PM} = \frac{\sum_{(i,j)\in M}\sum_{t=2}^{T} P_{it}(1-P_{it-1})((1/\omega_t)R_{it} - (1/\omega_{t-1})R_{it-1} - ((1/\omega_t)R_{jt} - (1/\omega_{t-1})R_{jt-1}))}{\sum_{(i,j)\in M}\sum_{t=1+L}^{T} P_{it}(1-P_{it-1})}$$

Note that the weighting factors used to adjust for partial implementation of the HRRP (i.e., ω_t used in this estimator are slightly different from the actual difference-in-difference estimator derived above; this is because we were limited in the extent to which we could modify the `PanelMatch` package. But, the difference between the two estimators should be relatively small:

$$\hat{\beta}_1 - \hat{\beta}_{PM} = \frac{\sum_{(i,j)\in M}\sum_{t=2}^{T} P_{it}(1-P_{it-1})(1/\omega_{t-1} - 1/\omega_t)(R_{it-1} - R_{jt-1})}{\sum_{(i,j)\in M}\sum_{t=1+L}^{T-F} P_{it}(1-P_{it-1})}$$

For any time t four or more years after the implementation of the HRRP, $\omega_t - 1 = 1$ so $(1/\omega_{t-1} - 1/\omega_t) = 0$ in those cases. Additionally, by matching on the pre-treatment trend in readmission rates, we can ensure that $R_{it-1} - R_{jt-1}$ is relatively close to zero. Both these factors should mitigate the difference between the estimator calculated using the `PanelMatch` function and the actual estimator associated with the model developed by Ryan et al.

Now, to actually run the matching procedure, we used the `PanelMatch` function to match on three years of lagged treatment history with the voluntary programs, and then used the Mahalanobis distance metric to prune the number of matched controls on time-invariant hospital characteristics (i.e., teaching hospital, number of beds, disproportionate patient percentage, and urban location) and lagged readmission rates for the three periods before implementation. We included the matched controls with the 10 lowest

Table 2: Estimates of Change in Readmission Rates Following Voluntary Reforms using Matching Procedure

Voluntary Reform	Percentage Point Change in Readmissions Rate (95% CI for 2-Sided Test)		
	AMI	Heart Failure	Pneumonia
Meaningful Use Only	-0.15 (-3.76, 3.45)	-0.10 (-3.88, 3.69)	-0.22 (-2.98, 2.53)
ACO Programs Only	0.00 (-1.04, 1.04)	-0.13 (-1.32, 1.06)	-0.10 (-0.98, 0.79)
Meaningful Use and BPCI	0.03 (-0.11, 0.16)	-0.01 (-0.15, 0.14)	0.03 (-0.1, 0.16)
Meaningful Use and ACO	0.01 (-0.88, 0.9)	-0.09 (-1.12, 0.94)	-0.06 (-0.78, 0.66)
All Programs	-0.09 (-0.29, 0.11)	0.02 (-0.19, 0.23)	-0.19 (-0.37, -0.01)

Mahalanobis distances in the control set for each treated observation, with equal weight placed on each member of the control set.

We calculated the effect of the implementation of the same set of voluntary programs analyzed by Ryan et al., as listed in Table 1: MU Only, ACO Only, MU & BPCI, MU & ACO, and All programs (MU, ACO, and BPCI). We also calculated these same measures without weighting for the proportion of the treated observation occurring after the implementation of HRRP and checked the the matched sets for balance, in order to assess the robustness of the results of the matching procedure. Additionally, since it could be argued that the effect of the voluntary programs on readmission rates takes multiple years to become apparent, we calculated the same difference-in-differences estimator for 1 and 2 years after the implementation of the voluntary program.

3.3 Results

After running the matching procedure, we found that implementation of the voluntary programs was generally not associated with a significant reduction in readmission rates. Table 2 presents the estimated change in readmission rates for the four conditions of interest following implementation of the voluntary programs; the only non-zero effect within a 95% confidence interval was a 0.19% reduction in pneumonia readmission rates following implementation of all the voluntary programs. Similar results were obtained when the matching procedure was run without the weighting scheme used to account for the proportion of observations occurring after the implementation of HRRP, as shown in Table 3.

Of course, the estimates derived by this matching procedure are only useful if the treated observations are similar to the control observations. In the context of matching on panel data, the observations must also be similar across lagged time periods. In general, it was found that the matched sets were balanced across the different covariates used for this procedure. For example, Figure 1 shows the difference in the level of the covariates between the matched treated and control observations (as measured on the standard deviation

Table 3: Estimates of Change in Readmission Rates Following Voluntary Reforms using Matching Procedure (Using Unweighted Observations)

Voluntary Reform	Percentage Point Change in Readmissions Rate (95% CI for 2-Sided Test)		
	AMI	Heart Failure	Pneumonia
Meaningful Use Only	-0.02 (-0.09, 0.06)	-0.04 (-0.11, 0.04)	0.01 (-0.05, 0.06)
ACO Programs Only	0.03 (-0.07, 0.13)	-0.05 (-0.15, 0.05)	-0.05 (-0.13, 0.02)
Meaningful Use and BPCI	0.02 (-0.06, 0.11)	-0.01 (-0.1, 0.08)	0.03 (-0.08, 0.13)
Meaningful Use and ACO	0.04 (-0.05, 0.13)	-0.05 (-0.15, 0.05)	-0.05 (-0.12, 0.03)
All Programs	-0.10 (-0.26, 0.06)	0.01 (-0.15, 0.16)	-0.20 (-0.35, -0.04)

scale) across the three years preceding the implementation of an ACO program; clearly, the observations were well matched, with all the differences between the covariates falling within 0.03 standard deviations. Similar results were seen for the matched sets used to evaluate the implementation of the other programs.

Additionally, as noted in the previous section, it could be argued that the voluntary programs have a significant effect on readmission rates, but that it takes some time for these effects to become apparent. This argument was evaluated by calculating the matched difference-in-differences estimator for 1 and 2 years following implementation of a voluntary program; it was generally found that the implementation of voluntary programs did not have an effect on readmission rates. For example, Figure 2 shows the effect of ACO program implementation on AMI readmission rates, with no observed change in effect after the first and second years following implementation.

4 Extension 2: Placebo Tests

4.1 Rationale and Approach

Placebo tests can provide useful evidence in situations where omitted variable bias and model dependence are likely. We therefore sought to identify publicly available measures reported at the hospital level which would not be expected to meaningfully change in conjunction with the adoption of the voluntary programs. We took as our pool of potential placebo test measures the numerous measures that have been available for public viewing on Hospital Compare. Other than readmissions, we found that the following quality measures were continuously reported through the CMS quality reporting programs between 2008 and 2015:

- Percent of heart attack patients given aspirin at arrival (abbreviated as AMI-1);
- Percent of heart attack patients given fibrinolytic medication within 30 minutes of arrival (AMI-7);
- Percent of heart attack patients given percutaneous coronary intervention (PCI) within 90 minutes of arrival (AMI-8);

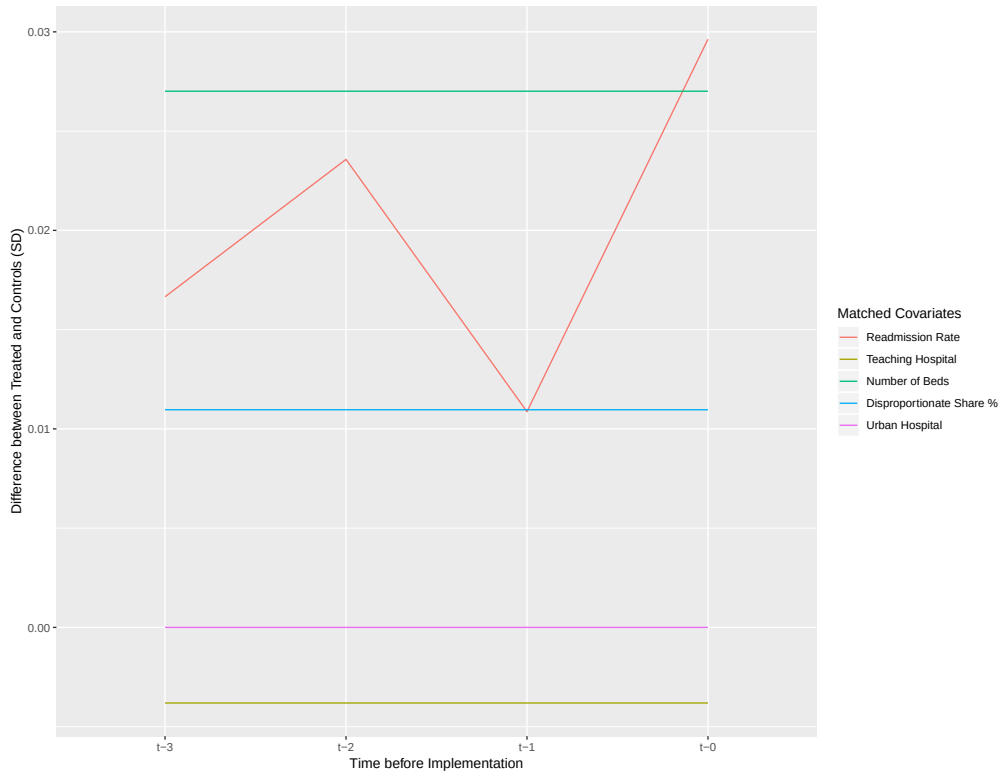

Figure 1: Covariate Balance for Matched Sets used to Calculate Reduction in AMI Admissions following ACO Implementation

- Percent of heart failure patients given an evaluation of left ventricular systolic (LVS) function (HF-2);
- Percent of surgery patients who received preventative antibiotic(s) one hour before incision (SCIP-1);
- Percent of surgery patients who received the appropriate preventative antibiotic(s) for their surgery (SCIP-2);
- Percent of surgery patients whose preventative antibiotic(s) are stopped within 24 hours after surgery (SCIP-3);
- Percent of surgery patients who received treatment to prevent blood clots within 24 hours before or after selected surgeries to prevent blood clots (SCIP-VTE-2); and
- Most survey items from the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) Survey.

The ACO programs, however, include a patient experience performance measure, suggesting that HCAHPS might make for a poor placebo [7]. Further analysis revealed that very few hospitals met the minimum volume threshold (more than 10 patients meeting the measure-specific denominator criteria) used after 2012 for AMI-7. We therefore decided to use only AMI-1, AMI-8, HF-2, SCIP-1, SCIP-2, SCIP-3, and SCIP-VTE-2 for placebo testing.

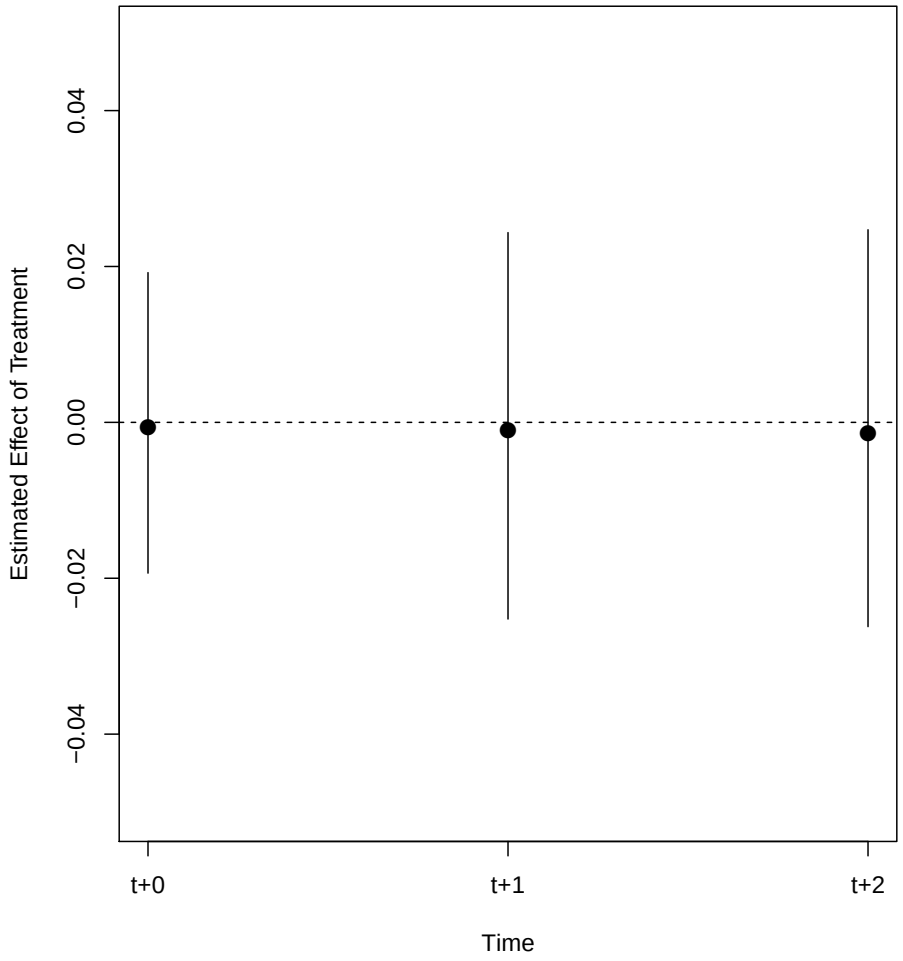

Figure 2: Estimates of Change in AMI Readmission Rates in 1 and 2 Years Following ACO Implementation

4.2 Procedure

We obtained year-specific Microsoft Access databases from the Hospital Compare data repository [8]. Each year-specific Access table pertaining to hospital-level quality and timeliness of care was extracted as a `.csv` file and brought into `R` for (extensive) data cleaning. Measure-specific longitudinal datasets of hospital performance were assembled. Within each measure-specific dataset, hospitals missing six or more observations (years) were excluded. For hospitals with seven observations – i.e., missing only one of eight years of data – we used the R package `Amelia`[9] to impute the one missing observation.[3]

We then executed the same analyses used by Ryan et al. (and reproduced in our replication) for each of the seven quality measures. In particular, we fit the same model described in Section 2, but replacing the readmission rate with each of the new quality measures as the dependent variable. As in Ryan et al, we included interactions between the post-HRRP period and participation in each of the the voluntary program(s). We prepared tables analogous to Table A1 and Table A2 replicated from Ryan et al.

4.3 Results

Between 2869 and 2808 hospitals met criteria for inclusion in AMI-1, HF-2, and the four surgical care quality measures (Table 4) – similar to the number of hospitals meeting criteria for inclusion in the heart failure and pneumonia readmission measures used by Ryan et al. (2785 and 2837, respectively, in Table A1). For AMI-8, only 1591 hospitals met criteria for inclusion in the study. Across the placebo quality measures, the hospital characteristics shown in Table 4 were qualitatively similar to the hospital characteristics in the Ryan et al. study (presented in Table A1) with the following exceptions:

- Among hospitals meeting criteria for inclusion in the AMI-1 measure, rates of participation in the voluntary programs were about one-third to one-half rates observed in the other programs.
- Hospitals meeting criteria for inclusion in the AMI-8 measure tended to be larger, more likely to be teaching hospitals, and more likely to be located in teaching areas.

Additionally, in Ryan et al. the average 30-day readmission rate (actually a percentage) varied from 17.9 to 23.6 percent, depending on condition. Among the placebo measures, average performance ranged from 86.5 to 96.3 percent.[4]

[3] Only linear time trends were used for imputation.

[4] This may have implications for the appropriateness of the linear model used by Ryan et al. for purposes of our placebo tests, which will be discussed later.

Table 4: Characteristics of Hospitals Included in Quality Measure Placebo Tests

Characteristic	AMI-1	AMI-8	HF-2	SCIP-1	SCIP-2	SCIP-3	SCIP-VTE-22
Hospitals, No	2860	1591	2869	2808	2809	2808	2811
Hospital-year observations, No	22880	12728	22952	22464	22472	22464	22488
Performance measure value, mean	96.35	86.52	95.69	94.07	95.78	92.71	90.94
Performance measure value, sd	8.00	16.91	9.68	9.95	7.59	10.20	12.81
Beds, mean	204.33	290.14	203.82	207.5	207.44	207.50	207.28
Beds, sd	178.82	194.00	178.78	178.93	178.93	178.93	178.94
Disproportionate Share Index, mean	0.28	0.27	0.28	0.28	0.28	0.28	0.28
Disproportionate Share Index, sd	0.17	0.16	0.17	0.17	0.17	0.17	0.17
Teaching hospital, %	32.38	48.21	32.28	32.98	32.97	32.98	32.94
Hospital located in urban area, %	70.94	88.37	70.93	71.97	71.98	71.97	71.90
Hospital participating during any time in study period, %							
Meaningful use of health information technology (MU)	68.53	92.90	96.83	94.16	94.13	94.09	94.88
Accountable care organization (ACO) programs	9.06	21.68	18.23	18.20	18.19	18.16	18.21
Bundled Payment for Care Initiative (BPCI)	5.00	16.15	11.75	12.00	12.00	12.00	11.99

Abbreviations: AMI-1, AMI-8, HF-2, SCIP-1, SCIP-2, SCIP-3, and SCIP-VTE-2 refer to hospital-level measures of quality of care. See text for descriptions.

Results from the interrupted time series models showed no significant improvement or deterioration in performance on the placebo measures in conjunction with entering the post-HRRP period for four of the seven measures (row 2, Table 5). The effects were qualitatively small, with point estimates ranging from -0.74 to 0.95 percentage points. Applying the same models used in Ryan et al. to estimate the effect of the voluntary programs, we found that hospitals not participating in any voluntary programs were expected to improve their performance on the placebo quality measures by 0.53 to 4.50 points depending on the measure (all increases significant) when calculated with all other covariates set at their means (row 3, Table 5). For the AMI-1 measure, participating in three of three voluntary programs was associated with an expected improvement of 3.46 points (95% CI: 1.51, 5.40) over the HRRP period. For the measures other than AMI-1, participation in MU and BPCI, MU and an ACO program, or all three programs was associated with statistically significant decreases in performance (Table 5, rows 6-8). Point estimates for the expected magnitude of performance decrease associated with participating in all three programs ranged from -2.32 percentage points (95% CI: 2.89, -1.74) for SCIP-2 to -9.20 percentage points (95% CI: -11.40,-7.00) for AMI-8.

5 Discussion and Conclusion

Our analyses suggest that participation in voluntary value-based programs following HRRP implementation does not have a direct causal effect on readmission rates for AMI, heart failure, or pneumonia. This suggests that the associations between program implementation and readmission rates observed by Ryan et al. can be better explained by unobserved hospital characteristics or trends.

The results from the matching procedure indicate that between hospitals that have similar lagged readmission rates and time-invariant characteristics, the implementation of one or more of the voluntary programs is not significantly associated with changes in readmission rates for AMI, heart failure, or pneumonia. This suggests that these voluntary programs do not have a causal effect on readmission rates, and either (1) there are some unobserved characteristics present in the matched hospitals that affect both voluntary program adoption and readmission rates or (2) there is some non-linear secular trend in readmission rates that is correlated with program implementation.

It is important to note, however, that there are some limitations to this matching analysis. Primarily, the weighting scheme used by Ryan et al. to account for partial implementation of the HRRP during the observed time periods could not be perfectly replicated with matching procedure used. As discussed in

Table 5: Estimates of Association Between Hospital Readmission Reduction Program (HRRP) Start and Improvement on Placebo Test Measures Across Participation in Voluntary Initiatives

	AMI-1	AMI-8	HF-2	SCIP-1	SCIP-2	SCIP-3	SCIP-VTE-2
Observations	17477	11363	22745	21749	21752	21719	21881
Estimate of HRRP	0.39 (-0.08, 0.85)	0.95 (0.10, 1.81)	-0.74 (-1.03, -0.46)	-0.06 (-0.45, 0.32)	-0.09 (-0.39, 0.21)	1.07 (0.70, 1.43)	0.45 (-0.02, 0.93)
Change Over HRRP if No Voluntary Program Participation	0.71 (0.30, 1.13)	4.50 (3.79, 5.21)	1.05 (0.85, 1.26)	2.11 (1.83, 2.39)	0.53 (0.30, 0.76)	2.92 (2.64, 3.20)	1.64 (1.29, 1.99)
Additional Change over HRRP Time Period if Hospital Participates in:							
MU Only	-0.32 (-0.80, 0.17)	-4.13 (-4.93, -3.33)	-2.18 (-2.48, -1.89)	-2.59 (-2.97, -2.20)	-0.68 (-0.93, -0.43)	-2.14 (-2.49, -1.78)	-1.25 (-1.73, -0.77)
ACO Only	-2.71 (-4.96, -0.47)	-2.27 (-4.57, 0.02)	-1.13 (-1.91, -0.34)	-2.34 (-2.98, -1.69)	-0.49 (-1.03, 0.06)	-2.15 (-3.01, -1.30)	-0.41 (-1.78, 0.95)
MU & BPCI	-2.26 (-3.30, -1.22)	-8.94 (-10.47, -7.40)	-4.59 (-5.04, -4.14)	-5.29 (-5.91, -4.67)	-1.90 (-2.32, -1.47)	-5.18 (-5.86, -4.50)	-4.45 (-5.26, -3.63)
MU & ACO	-1.39 (-2.13, -0.65)	-6.73 (-8.13, -5.34)	-3.73 (-4.13, -3.33)	-4.57 (-5.13, -4.00)	-1.64 (-1.99, -1.29)	-4.27 (-4.87, -3.67)	-3.56 (-4.33, -2.79)
All programs	3.46 (1.51, 5.40)	-9.20 (-11.40, -7.00)	-4.85 (-5.45, -4.26)	-5.82 (-6.64, -5.00)	-2.32 (-2.89, -1.74)	-5.43 (-6.38, -4.49)	-4.62 (-5.79, -3.45)

All values are percentage points, except row 1. 95 percent confidence intervals shown in parentheses. Abbreviations: ACO, Accountable Care Organization; BPCI, Bundled Payment for Care Initiative; MU, Meaningful Use. AMI-1, AMI-8, HF-2, SCIP-1, SCIP-2, SCIP-3, and SCIP-VTE-2 refer to hospital-level measures of quality of care. See text for descriptions.

Section 3.2, however, the fact that the treatment and control hospitals used for the analysis were well balanced on pre-treatment readmission rates suggests that this limitation will not significantly impact the conclusions from this analysis. Additionally, it is important to note that some of the confidence intervals of the effects calculated by the matching analysis are rather large. For example, in Table 2 the effect of the implementation of the meaningful use programs on AMI readmissions ranges from -3.76% to 3.45%. Thus, the lack of significant effects calculated by the matching procedure may indicate that there were not enough treated observations in the dataset to detect the effect of the voluntary programs on readmission rates, rather than indicating that there is no effect at all.

The conclusions from the matching analysis is further supported by the results from the placebo tests. We believe it is unlikely that changes in the placebo measures would be caused by participation in the voluntary programs since these measures are not related to the voluntary programs. The actual results of the placebo tests appear to suggest that increased participation in the CMS voluntary initiatives during the HRRP era has been associated with deterioration in hospital performance for six of seven process-oriented measures of care quality examined. While this may appear to indicate that the voluntary programs are leading to worsening quality, we think this is unlikely. Rather, we believe this shows the tendency of the model developed in Ryan et al. to capture underlying changes in hospital performance driven by unobserved covariates, rather than specific causal effects attributable to the voluntary programs. Some explanations for the decreases in quality performance that might be compelling in other contexts – for instance, crowd-out of effort for non-incentivized measures, as observed in the National Health Services' Quality Outcomes Framework pay-for-performance scheme [10] – appear unlikely given the relative simplicity of these processes, the backsliding observed, and the lack of any other recent US findings pointing to such a result.

As noted above, this finding of significant yet implausible results from the placebo tests suggests that the original Ryan et al. estimates may be capturing the effects of unobserved differences on tendency to both join voluntary programs and reduce readmissions. However, it is possible that the model used to evaluate these placebo measures is misspecified, which is an important limitation of this conclusion. In particular, the quality measures have a maximum value of 1.00, and the mean values were near 1.00 many of these measures (Table 4). This violates the assumption that the outcome is Normally distributed, which is inherent in the Ryan et al. model. In addition, the characteristics of hospitals meeting criteria for inclusion in the placebo tests varied somewhat from the characteristics of the hospitals meeting criteria for inclusion in the original Ryan et al. analysis (see Table 4 versus Table A1 in the Appendix and discussion in results section). This also suggests that findings from our placebo tests should be interpreted with caution.

Still, by combining the conclusions from the matching analysis and placebo tests, the results from this paper suggest that – at the very least – the relationship between the HRRP, voluntary value-based initiatives, and readmission rates is not as clear-cut as the Ryan et al. analysis indicates. This emphasizes the need for further work to better understand the interactions among the multitude of voluntary and mandatory value-based reforms that CMS has implemented and their effect on patient care and costs.

A Replication Results

We present the tables and figures that we replicated from the ones available in Ryan et al. (2017). In particular, we replicated all the tables and figures in the original paper, but not the additional ones in the supplemental analysis. The following two tables and three figures are replications of the corresponding tables and figures in Ryan et al. (i.e., Table A1 below replicates Table 1 in the original paper). Note that in Figure A1, the results for the other and ACO categories for 2015 do not match the results given in the original article. We believe this was a publication error, since the authors' original code seems to match the graph we replicated below. All the other results seem to match exactly. We replicated the results using `R`.

Table A1 below shows the summary statistics for the characteristics of hospital cohorts included in the datasets. Table A1 displays information such as the number of hospitals, the number of beds, proportion of hospitals in the urban area, etc. Figure A1 complements the information in Table A1 by showing percentage of hospitals that participated in voluntary value-based reforms from 2010 to 2015. As Figure A1 suggests, the participation pattern has varied substantively from year to year from 2010 to 2015.

Table A2 shows the results of the interrupted time-series analysis performed by Ryan et al. Because `R`, unlike `STATA`, does not have a built-in function that calculates marginal effect estimates from panel data, we replicated the results from Ryan et al. by developing a new function that both calculates the marginal effects of HRRP implementation from a fixed effects model (using the `plm` function to fit the initial model) and then calculates the first difference effects from implementing a set of voluntary programs. In order to calculate the appropriate standard errors, the delta-method was applied to these differences in marginal effect estimates. This procedure was derived from the Stata documentation [11]. Figure A2 shows the plot of 30-day readmission rates between 2009 and 2015. Lastly, Figure A3 shows a detailed analysis of impact of participation in the voluntary programs across various hospital characteristics. Again, we used the same panel fixed effect model margin estimation function developed for and used in Table A2.

Table A1: Characteristics of Hospital Cohorts

Characteristic	AMI	Heart Failure	Pneumonia
Hospitals, No.	1932	2785	2837
Hospital-year observations, No.	15456	22280	22696
30-d risk standardized observations rate (2008-2015), mean (SD)	18.7 (1.8)	23.6 (2.3)	17.9 (1.7)
Beds, mean (SD)	263 (187)	208 (179)	205 (179)
Disproportionate Share Index, mean (SD)	0.26 (0.15)	0.28 (0.17)	0.28 (0.17)
Teaching hospital, %	42.8	32.9	32.3
Hospital located in urban area, %	83.3	71.4	70.8
Hospital participating during any time in study period, %			
Meaningful use of health information technology	98.1	97.6	97.5
Accountable care organization programs	21.8	18.6	18.3
Bundled Payment for Care Initiative	15.9	12.1	11.9

Table A2: Estimates of the Association between the Hostpital Readmission Reduction Program and Readmissions Across Participation in Voluntary Reforms

	Percentage Point Change in Readmission Rate (95% CI for 2-Sided Test)		
Characteristic	AMI	Heart Failure	Pneumonia
Observations	15456	22280	22696
Overall Estimate of HRRP	-1.46 (-1.59, -1.32)	-2.13 (-2.28, -1.99)	-1.32 (-1.44, -1.19)
Estimate of HRRP if hospitals participate in no value-based reforms	-0.76 (-0.93, -0.60)	-1.30 (-1.47, -1.13)	-0.82 (-0.97, -0.67)
Change in estimate of HRRP if hospitals participate in:			
Only meaningful use	-0.78 (-0.89, -0.66)	-0.97 (-1.08, -0.86)	-0.56 (-0.65, -0.47)
Only ACO programs	-0.94 (-1.29, -0.59)	-0.83 (-1.26, -0.41)	-0.59 (-1.00, -0.18)
Meaningful use and BPCI	-1.02 (-1.22, -0.81)	-1.16 (-1.40, -0.92)	-0.71 (-0.90, -0.51)
Meaningful use and ACO	-1.00 (-1.18, -0.83)	-1.30 (-1.48, -1.11)	-0.83 (-0.99, -0.68)
All Programs	-1.27 (-1.58, -0.97)	-1.64 (-2.02, -1.26)	-1.05 (-1.32, -0.78)

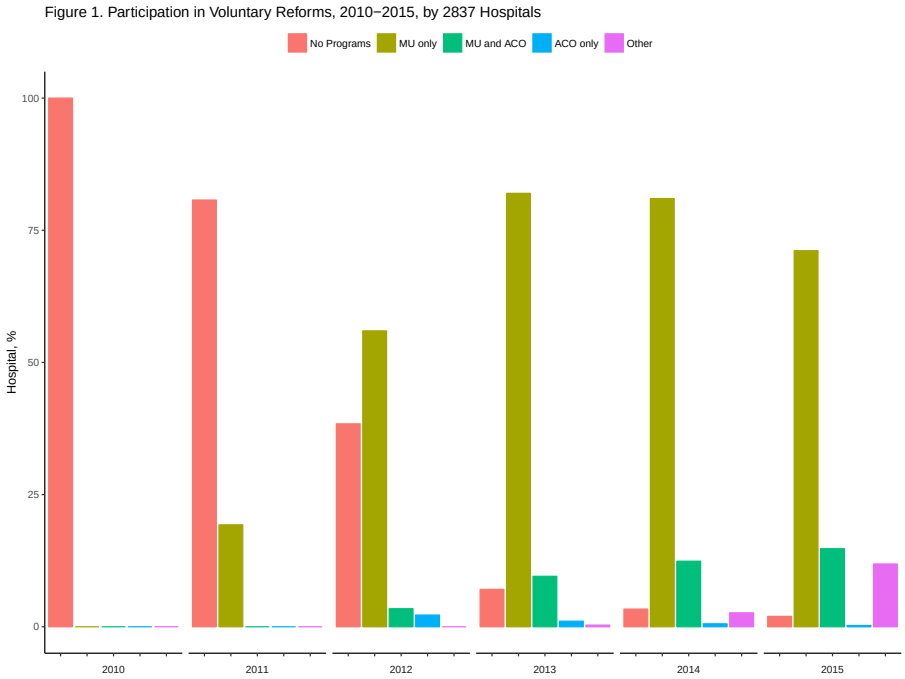

Figure A1: Participation in Voluntary Reforms, 2010-2015 by 2837 Hospitals.

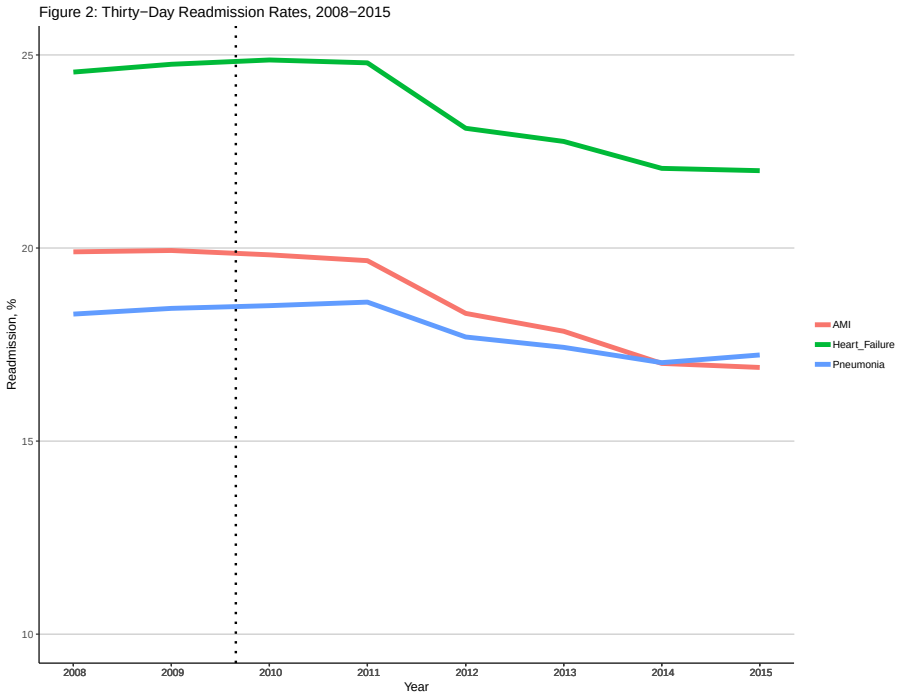

Figure A2: 30-Day Risk-Standardized Readmission Rates for Acute Myocardial Infarction (AMI), Heart Failure, and Pneumonia, 2008-2015. The dotted black line represents the time HRRP came into effect.

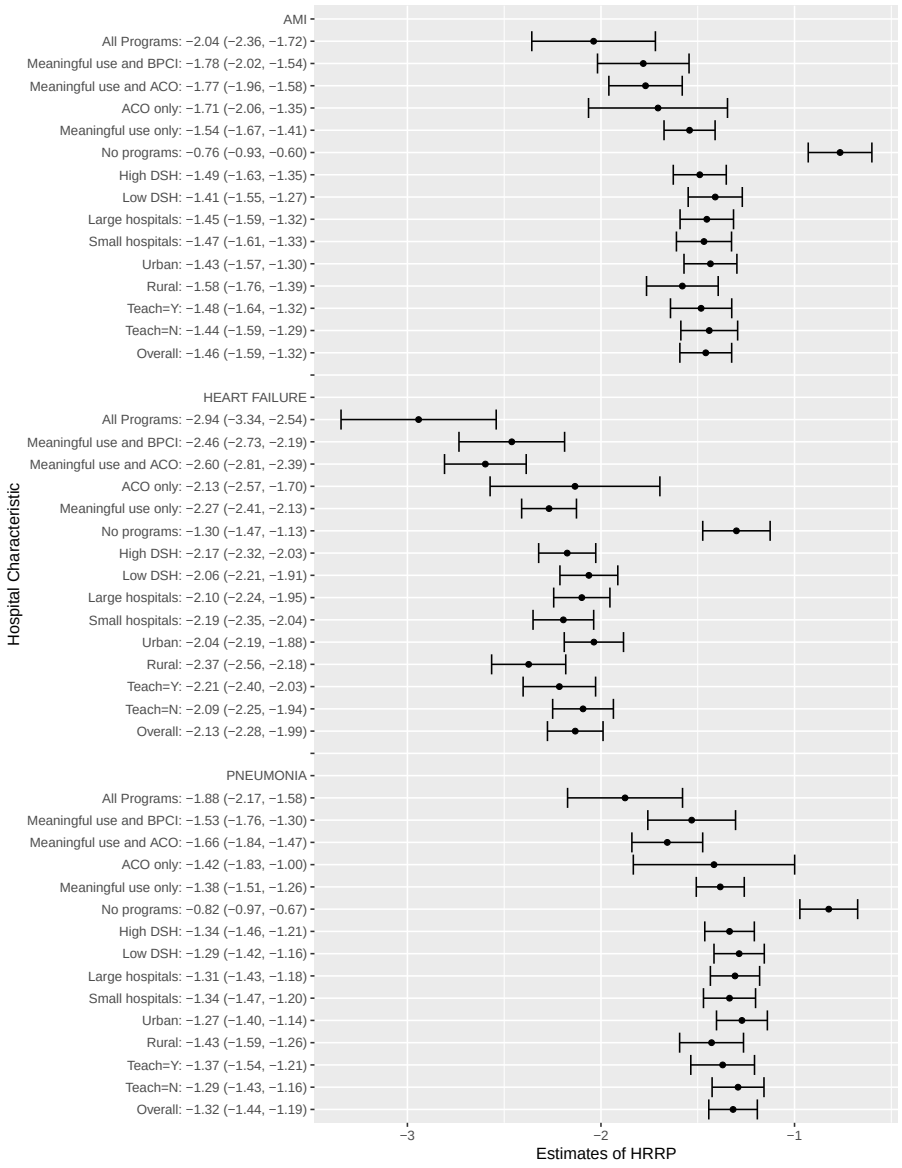

Figure A3: Variation in the Impact of the Hospital Readmissions Reductions Program Across Hospital Characteristics and Participation in Voluntary Reforms. The bars show the 95% Confidence interval of the estimated effect on readmissions reduction. High DSH (and Large Hospitals) and Low DSH (and Small Hospitals) indicate the 75% and the 25% percentile of the DSH index (and the number of beds), respectively.

References

[1] "Where Is HITECH's $35 Billion Dollar Investment Going?" In: *Health Affairs Blog* (2015). URL: DOI:%2010.1377/hblog20150304.045199.

[2] Cristina Boccuti and Giselle Casillas. *Aiming for Fewer Hospital U-turns: The Medicare Hospital Readmission Reduction Program*. Feb. 2018. URL: https://www.kff.org/medicare/issue-brief/aiming-for-fewer-hospital-u-turns-the-medicare-hospital-readmission-reduction-program/.

[3] "Patient Protection and Affordable Care Act (ACA). HR 3590." In: (2011). URL: https://www.gpo.gov/fdsys/pkg/BILLS-111hr3590enr/pdf/BILLS-111hr3590enr.pdf.

[4] Kosuke Imai, In Song Kim, and Erik Wang. "Matching Methods for Causal Inference with Time-Series Cross-Sectional Data". In: *Working Paper*. (2019). DOI: 10.1214/09-STS313.Matching.

[5] Karen E Joynt and Ashish K Jha. "Characteristics of hospitals receiving penalties under the Hospital Readmissions Reduction Program". In: *Jama* 309.4 (2013), pp. 342–343.

[6] Elizabeth H Bradley et al. "Contemporary data about hospital strategies to reduce unplanned readmissions: what has changed?" In: *JAMA internal medicine* 174.1 (2014), pp. 154–156.

[7] "Centers for Medicare and Medicaid Services. Guide to quality performance scoring methods for accountable care organizations." In: (2014). URL: http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Downloads/2012-11-ACO-quality-scoring-supplement.pdf.

[8] "Hospital Compare Data Archive". In: (2019). URL: https://data.medicare.gov/data/archives/hospital-compare.

[9] James Honaker, Gary King, Matthew Blackwell, et al. "Amelia II: A program for missing data". In: *Journal of statistical software* 45.7 (2011), pp. 1–47.

[10] Tim Doran et al. "Effect of financial incentives on incentivised and non-incentivised clinical activities: longitudinal analysis of data from the UK Quality and Outcomes Framework". In: *Bmj* 342 (2011), p. d3590.

[11] Stata Corp. *Stata 15 Base Reference Manual*. Stata Press, 2017.

[12] Andrew M. Ryan et al. "Association between hospitals' engagement in value-based reforms and readmission reduction in the hospital readmission reduction program". In: *JAMA Internal Medicine* 177.6 (2017), pp. 862–868. ISSN: 21686106. DOI: 10.1001/jamainternmed.2017.0518.