NBER WORKING PAPER SERIES

DID THE HOSPITAL READMISSIONS REDUCTION PROGRAM REDUCE READMISSIONS? AN ASSESSMENT OF PRIOR EVIDENCE AND NEW ESTIMATES

Engy Ziedan
Robert Kaestner

Working Paper 29545
http://www.nber.org/papers/w29545

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2021

Partial funding for this research was provided by the Agency for Health Research and Quality (AHRQ) under award R01 HS025586. We thank Atul Gupta for sharing information about the analysis in Gupta (2021). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Did the Hospital Readmissions Reduction Program Reduce Readmissions? An Assessment of Prior Evidence and New Estimates
Engy Ziedan and Robert Kaestner
NBER Working Paper No. 29545
December 2021
JEL No. I11,I18

ABSTRACT

In this article, we provide a comprehensive, empirical assessment of the hypothesis that the Hospital Readmissions Reduction Program (HRRP) affected hospital readmissions. In doing so, we provide evidence as to the validity of prior empirical approaches used to evaluate the HRRP and we present results from a previously unused approach to study this research question—a regression-kink design. Results of our analysis document that the empirical approaches used in most prior research assessing the efficacy of the HRRP often lack internal validity. Therefore, results from these studies may not be informative about the causal consequences of the HRRP. Results from our regression-kink analysis, which we validate, suggest that the HRRP had little effect on hospital readmissions. This finding contrasts with the results of most prior studies, which report that the HRRP significantly reduced readmissions. Our finding is consistent with conceptual considerations related to the assumptions underlying HRRP penalty; in particular, the difficulty of identifying preventable readmissions, the highly imperfect risk adjustment that affects the penalty determination and the absence of proven tools to reduce readmissions.

Engy Ziedan
Department of Economics
School of Liberal Arts
Tulane University
206 Tilton Hall
6823 St. Charles Avenue
New Orleans, LA 70118
eziedan@tulane.edu

Robert Kaestner
Harris School of Public Policy
University of Chicago
1307 East 60th Street (Room 3057)
Chicago, IL 60637
and NBER
kaestner@uchicago.edu

Introduction

The fee-for-service payment approach of traditional Medicare has long been criticized for its inherent incentive to provide too much, low-valued care and there is a widespread belief that this problem can be solved by linking provider payments to performance (outcomes) instead of the quantity of care (Institute of Medicine 2007; Cromwell et al. 2011; Mendelson et al. 2017). It was this logic that led to the inclusion of several pay-for-performance initiatives in the 2010 Patient Protection and Affordable Care Act (ACA). One of the most prominent of the ACA pay-for-performance initiatives was the Hospital Readmissions Reduction Program (HRRP). The goal of the HRRP was to reduce avoidable hospital readmissions that were thought to be substantial, for example 27% of all admissions (Van Walraven et al. 2011; Auerbach et al. 2016)—and costly—$17 billion in 2016 (Bailey et al. 2019).

Given the visibility of the ACA and the wide scope of the HRRP, which applies to over 3000 hospitals, there has been substantial interest in evaluating the consequences of the HRRP. Prior studies that have assessed whether the HRRP has reduced readmissions and affected other outcomes (e.g., mortality). However, many of these studies are single, or two, state analyses (e.g., Carey and Lin 2015; Chen and Grabowski 2017; Demiralp et al. 2017; Zingmond et al. 2018; Mellor et al. 2017);[1] many do not identify the impact of being penalized, but instead analyze differences between hospitals that differ along a dimension likely to be correlated with being penalized (e.g., McGarry et al. 2016; Chen and Grabowski 2017 and Ibrahim et al. 2018); and a few studies including some of the most well-cited national analyses use an interrupted time series approach that has acknowledged limitations with respect to identifying causal relationships (e.g., Zuckerman et al. 2016 and Chahabra et al. 2019). In addition, the prior literature has not produced a consensus finding. While it is clear that hospital readmissions have declined over the last decade, it is uncertain whether the HRRP had led to this decline (Lisk and Stensland 2018).[2] Therefore, it remains largely unknown whether the HRRP achieved its intended goal and, more generally,

[1] National studies include: Zuckerman et al. 2016; Desia et al. 2016; Wasfy et al. 2017; Ibrham et al. 2018; Chhabra et al. 2019; Gupta 2020)
[2] http://medpac.gov/docs/default-source/default-document-library/readmissions-january-2018-public9bd311adfa9c665e80adff00009edf9c.pdf?sfvrsn=0

whether the type of pay-for-performance mechanism embedded in the HRRP may be a productive way to improve the efficiency of health care.

In this article, we provide a comprehensive, empirical assessment of the hypothesis that the HRRP affected hospital readmissions. In doing so, we provide evidence as to the validity of prior empirical approaches used to evaluate the HRRP and we present results from a previously unused approach to study this research question—a regression-kink design. We also provide a discussion of conceptual issues related to how hospitals may respond to the HRRP, which has largely been missing from previous studies of the HRRP. The conceptual framework suggests that the effect of the HRRP, if any, is likely to be small and to differ by the cause of hospitalization (and for readmission after that hospitalization) and by hospital characteristics, for example, the share of patients in penalized conditions of total Medicare patients.

Results of our analysis document that the empirical approaches used in most prior research assessing the efficacy of the HRRP lack internal validity. Therefore, results from these studies may not be informative about the causal consequences of the HRRP. Results from our novel, regression-kink approach, which we validate, suggest that the HRRP had little effect on hospital readmissions. This finding contrasts with the conclusions of most prior studies, which report that the HRRP reduced readmissions. However, our finding is consistent with conceptual considerations related to the assumptions underlying the HRRP penalty; in particular, the difficulty of identifying preventable readmissions, the highly imperfect risk adjustment that affects the penalty determination and the absence of proven tools to reduce readmissions.

Our article contributes to the literature on pay-for-performance incentives in health care settings (Doran et al. 2016; Mendelson et al. 2017). The HRRP is a particularly important pay-for-performance program because of its prominence and wide coverage, and because the financial incentive is a penalty, which differs from the more common approach of using bonuses to reward performance (Cromwell et al. 2011). There is also a dynamic aspect to the incentive structure of the HRRP that is unique. Hospitals compete against each other to be in the lower half of the distribution of (adjusted) readmissions and hospitals have potentially different ways (actions) of competing. For example, an obvious and hoped for response is for a hospital to improve the quality of care it provides. The fact that we find little effect of the HRRP on performance (i.e., readmissions) suggest that this type of pay-for-performance incentive may not be efficacious, or that the implementation details of such a performance incentive need to be carefully considered.

Our article also contributes modestly to applied research in causal methods. We develop a novel test of the validity of a difference-in-differences-in-differences research design that is easily applied and straightforward to assess. We show that the test is consistent with an event-study specification that is typically used to assess difference-in-differences models. However, the test we propose is more parsimonious, which is its main strength.

Conceptual Considerations

As noted, most prior studies of the HRRP are empirical assessments, which is reasonable given that ultimately it is the consequences of the HRRP that are important in terms of assessing the efficacy of the program. However, it is helpful to consider some of the conceptual issues underlying the HRRP because doing so helps distinguish between consequences of the HRRP that are more or less scientifically plausible and because a conceptual model can highlight empirical predictions that can be assessed.

Structure of HRRP Incentive

Before discussing some of the conceptual issues, it is helpful to review the structure of the HRRP. The HRRP imposes financial penalties on hospitals with excess readmissions of patients admitted for one of several conditions (e.g., heart failure).[3] Excess readmissions are defined as having more than the average (expected) number of readmissions for a condition after adjusting for patient characteristics

[3] Initially, there were three conditions: AMI, heart failure and pneumonia. Currently, there are six conditions: the original three, Chronic Obstructive Pulmonary Disease (COPD), Coronary Artery Bypass Graft (CABG) and total hip or knee arthroplasty.

(comorbidities), and more recently, patient socioeconomic status (dual Medicaid-Medicare eligibility).[4] A hospital is penalized if it has excess readmissions for any of the specified HRRP conditions and the penalty increases with the number of excess readmissions and/or the number of conditions for which the hospital has excess readmissions. An important feature of the HRRP is that the financial penalty imposed applies to all Medicare payments and not just payments for targeted conditions. The penalty is substantial—currently up to 3% of Medicare payments.[5] In 2017 and 2018, the average penalty was over $500 million for penalized hospitals.[6]

Will a Hospital Respond?

Given the structure of the HRRP incentive, it is reasonable to assume that hospitals will respond by assessing whether the cost of responding, for example, taking actions to reduce readmissions, is less than the benefit of responding, which is the savings from not being penalized. This simple cost-benefit calculus has several implications about hospital behavior that we now discuss.

Perhaps the most important implication is that hospitals will only respond to the HRRP if in fact they can identify and prevent readmissions. Identifying a potentially preventable readmission is inherently a subjective exercise and while there are undoubtedly some readmissions that are preventable, the share of readmissions classified as such (determined ex post) is uncertain according to two systematic reviews (Vest et al. 2010; van Walraven et al. 2011) and small (11%) according to MEDPAC (2018). Even if the share or preventable hospitalizations is sizable and larger than calculated by MEDPAC (2018), it is still necessary to be able to identify these cases ex ante to respond in an efficient manner. However, common approaches to identifying potentially preventable readmissions ex ante are often not very accurate (Kansagara et al. 2011; Elham et al. 2020). The uncertainty over the share of readmissions that are preventable and the difficulty of identifying them ex ante suggests that hospitals have limited ability to

[4] For archived (round 1) and later adjustments to the HRRP penalty formula please see https://qualitynet.cms.gov/inpatient/hrrp

[5] 3% reduction in Medicare payments represents a substantial financial penalty because Medicare represents approximately 35 percent of hospital revenue. Moreover, hospital profit margins are approximately four percent and twenty percent of hospitals have negative margins (American Hospital Association, 2015)

[6] https://catalyst.nejm.org/doi/full/10.1056/CAT.18.0194

target admissions amenable to interventions to reduce potentially preventable readmissions. So, the cost of reducing a readmission includes the resources spent on admissions that are in fact not preventable, which are the large majority of admissions. The higher the cost of reducing preventable readmissions, the less likely it is that a hospital will expend the resources to avoid a HRRP penalty.

In addition to being difficult to identify a preventable readmission, there is, at best, mixed evidence of demonstrably effective interventions to reduce readmissions, particularly interventions that can be effectively implemented by the hospital to influence readmissions occurring after seven days post-discharge Graham et al. 2018). These delayed readmissions are the least responsive to the quality of care of the inpatient stay directly controlled by the hospital and depend on patient behavior and outpatient care.

Some quotes from recent reviews make the point about the lack of effective interventions.[7]

"Of interest (and concern), the authors found no consistent evidence from RCTs that any one intervention by itself significantly reduced hospital readmission. Of the 16 RCTs, only 5 yielded significant reductions in hospital readmission." (Kripalani et al. 2014, p 3).

"Burke and colleagues recently performed an updated systematic review… They included 61 interventions, 42 of which have been studied in RCTs. ... Just under half (47.5%) of interventions demonstrated a statistically significant reduction in readmissions. Consistent with prior reviews, no singular intervention component significantly reduced readmissions, though a trend was present for patient education and engaging social and community supports (p=0.06 for each)." (Kripalani et al. 2014, p. 5)

"Controlled trials [6-13] and cross-sectional studies [14-16] have identified different practices that may be effective, such as patient education, discharge planning, and telephone follow-up; however, no specific clinical practice has been shown to consistently reduce readmissions.[5] Furthermore, longitudinal studies

[7] Also, see Finkelstein et al. (2020) who conducted a RCT to evaluate the Camden Core Model care-transition program. The Camden program had no effect on readmissions.

have failed to link specific hospital practices with subsequent reduction in risk-standardized readmission rates (RSRRs), although employment of a greater number of these practices seems to be helpful." (Brewster et al. 2016, p.600)

To recap, prior evidence has demonstrated that it is difficult to identify and target preventable readmissions. This implies that to reduce readmission, a hospital would have to take actions, and presumably spend resources, on many admissions that are not preventable raising the cost of responding. Moreover, there is limited evidence that there are effective actions that a hospital can take to reduce readmissions—even if it could identify them. These considerations suggest a limited response by hospitals to the HRRP and, therefore, a small if any change in readmissions as a result of the HRRP.

Another aspect of the HRRP that may limit the ability of a hospital to respond is the 30-day period that is used to define a readmission. A study by Graham et al. (2018) reported that readmissions within 7 days of discharge are more likely to be amenable to changes in hospital practices than admissions after this period that are more likely due to the care received from other providers. Chin et al. (2016) came to a similar conclusion. However, most readmissions (two-thirds) during the 30-day widow occur after 7 days (Fingar et al 2017).[8] This evidence, again, suggests that it may be very costly for the hospital to respond to the HRRP because any action the hospital may take will, at best, affect only a portion of the readmissions that will be attributed to it.

A third aspect of the HRRP that may limit the ability of hospitals to respond is that the definition of excess readmissions is based on an extensive, but still limited set of observable comorbidities of patients. Hospitals with patients that are more (less) likely to be readmitted for unmeasured reasons will be more (less) likely to be penalized even if they are providing optimal care. This feature of the HRRP has been one of the most heavily criticized (American Hospital Association 2015; Joynt et al. 2016; Soumerai and Sullivan 2019). Even the recent addition of the share of patients that are dual eligible for

[8] https://www.hcup-us.ahrq.gov/reports/statbriefs/sb230-7-Day-Versus-30-Day-Readmissions.jsp#:~:text=Overall%2C%2036.1%20percent%20of%20all,highest%207%2Dday%20readmission%20rates.

Medicaid and Medicare (i.e., crude socioeconomic indicator) to the calculation of adjusted readmissions is only a partial solution to the problem of unmeasured confounders. Indeed, the models used by the Centers for Medicare and Medicaid Services (CMS) to adjust readmissions for comorbidities and that is intended to put hospitals on an equal footing for comparison (and penalty) purposes explain relatively little of the total variance in readmissions for some conditions (Krumholz et al. 2008). The inability of the CMS models to explain variation in readmissions across hospitals using patient characteristics is consistent with the studies noted earlier about the inability of hospitals to identify potentially preventable readmissions.

The implication of this measurement issue is that much of the variation in readmissions, and therefore, the determination of the HRRP penalty is likely due to omitted differences in patient-level factors. If so, then hospitals have little incentive to change practices because it will have little effect on the probability of being penalized.[9] Evidence that the HRRP consistently penalizes the same hospitals is provided by Thompson et al. (2017) who report that 53% (65%) of hospitals received penalties in every year (four) of the first five years of the HRRP program and 11% (15%) of hospitals were never penalized (or penalized just once). Overall, it is likely that only about 20% of hospitals face much uncertainty over whether or not they will be penalized by the HRRP. Most hospitals know that they will or will not be penalized and it is largely because of unmeasured patient risk factors and not the standard of care.

Finally, there is the dynamic aspect of the HRRP that may influence hospital behavior. Assuming, perhaps heroically, hospitals have some ability to influence readmissions in a cost-effective way and that the calculation of excess readmissions effectively puts hospitals on an equal footing in terms of patient severity, then the fact that every year 50% of hospitals are penalized for each condition makes the response of one hospital depend on the responses of other hospitals. Consider a simple case where there are just two hospitals. One hospital is providing optimal (less substandard) care and the other is providing

[9] The hospital would, in this case, have an incentive to alter patient characteristics if it knew which characteristics were associated with readmission. Ziedan (2020) examined this issue and found no evidence that patient characteristics change with hospital penalty status.

substandard care. The substandard hospital will be penalized because it will have more (adjusted) readmissions. If it responds by providing the optimal (less substandard) level of care, then in the next HRRP period, it will be random which of the two hospitals will be penalized.

This simple logic implies that, ideally, if the HRRP was effective, over time the distribution of (adjusted) readmissions would narrow (coalescing around an optimal level of care), the mean number of readmissions would be decreasing, and the probability of being penalized would become increasingly random. Evidence, however, does not suggest this has occurred. As noted, the probability of being penalized is highly persistent even after five years of the program (Thompson et al. 2017) and the distribution of adjusted readmissions has not narrowed over time. It is true that the mean of readmissions has been declining, but this is true for all types of readmissions and not just those targeted by the HRRP (Lisk and Stensland 2018).

Condition-specific Responses

While there are reasons to expect that the HRRP may have a limited effect on hospital behavior and readmissions, if hospitals do respond to the HRRP, it is unlikely that the response will be the same across HRRP conditions or across hospitals. First, the ability to identify, ex ante, potentially preventable readmissions differs by condition (Weinreich et al. 2016; Siracuse and Chamberlain 2016; Chamberlain et al. 2018; Smith et al. 2018). Thus, conditions for which potentially preventable readmissions can be most easily identified are the ones that are expected to have the largest change. For example, there is some evidence that readmissions are more easily predicted ex ante for heart failure patients than pneumonia and AMI patients. Therefore, all else equal, hospitals would be more likely to respond to HRRP by altering practices related to heart failure, but the all else equal caveat is not to be taken lightly. There still needs to be an effective intervention even if patients likely to be readmitted can be reasonably well identified.

Second, there may be significant differences in the cost of effective interventions across conditions. Hospitals may have to use different amounts of resources to reduce readmissions for one condition (e.g., AMI) than for another condition (e.g., pneumonia). Third, even if the marginal costs of intervention are the same across conditions, for example, because a common post-discharge protocol is used, the cost of the interventions to reduce readmissions may differ because of the relative number of patients with a condition. Consider two hospitals that were penalized only because they had excess readmissions for AMI patients. In one hospital, AMI patients represent 10% of all Medicare admissions and in the other hospital AMI patients represent 5% of all Medicare admissions. The latter hospital can avoid the penalty by spending fewer additional resources to prevent readmissions of AMI patients than the former hospital. Therefore, the hospital with relatively fewer AMI admissions will be more likely to respond; every dollar invested by the hospital with relatively fewer AMI patients to reduce AMI readmissions will bring a greater benefit (reduced penalty on all Medicare admissions) compared to a hospital with relatively more AMI admissions.

The upshot of this discussion is that it is unlikely that the HRRP will have similar effects on each of the targeted conditions. Differences in the ability to detect and the cost to prevent avoidable readmissions are very likely to differ by condition. Of course, there may be a set of circumstances that would lead to this result (similar effects), but it seems improbable.

An Empirical Assessment of the Evidence of Research Designs used in Prior Empirical Studies

There is a substantial body of research assessing the effects of the HRRP. Prior studies use four research designs: interrupted time series; difference-in-differences (DD); difference-in-differences-in-differences (DDD) and instrumental variables (IV). We assess the validity of each approach. Data for all of the analyses comes primarily from the 100% sample of Medicare administrative inpatient records reported in the MEDPAR files from 2006 to 2014. This period covers the periods, or much of the periods, analyzed by most prior studies. From here onward we refer to the data year 2006 to 2007 as 2007, and so on. In several of our analyses we use the baseline 2007 (2006 to 2007) readmissions and utilize 2008(2007 to 2008) to 2014(2013 to 2014) in our regression models. The MEDPAR files contain detailed information on all inpatient episodes of care for fee-for-service Medicare enrollees. The outcome we examine is 30-day readmissions. In addition, we use the Inpatient Prospective Payment System (IPPS) files that provide information about what hospitals were penalized and the excess readmission ratios for each HRRP targeted condition that are used to determine the penalty. We also extracted various hospital level covariates from the IPPS files including the number of beds and the hospital's teaching status.

Our sample is limited to hospitals that were assessed and not exempt from the penalty by CMS. A total of 3,128 hospitals were assessed for the HRRP penalties. This represents a major share of hospitals[10]. We focus on the first-round penalty, as has much of the prior research. In the first round, 2241 hospitals received penalties, for at least one condition. Of those 2241 hospitals 1,910 received a penalty below the maximum 1 percent cap. Another 887 hospitals received no penalties for any condition.

The Validity of the Difference-in-differences (DD) Design and Prior Evidence Using It

Several studies have used a difference-in-differences design, or the closely-related interrupted time series, to study the effect of the HRRP (e.g., Zuckerman et al. 2016; Desai et al. 2016; Wasfy et al. 2017). The basic DD design is well known. In the context of the HRRP, the DD approach compares changes in readmissions (other outcome) before and after the HRRP of hospitals that were penalized, or more likely to be penalized, for example because of higher baseline admissions (Demiralp et al. 2017), to changes in readmissions before and after the HRRP of hospitals that were not penalized, or less likely to be penalized. The key assumption of the DD design used in these prior studies is parallel trends—that in the absence of the HRRP, the time-series pattern of readmissions would be the same in hospitals that were and were not penalized, or hospitals that were more or less likely to be penalized. We assess the validity of this critical assumption.

The regression model analog of the DD approach is:

$$\begin{aligned} &READMIT_{jt} = a_j + \delta_t + X_{jt}\Gamma + \beta PENALTY_{jt} + e_{jt} \\ (1)\quad &j = 1,\ldots,N \\ &t = 2008,\ldots,2014 \end{aligned}$$

[10] Not all hospitals were assessed for the HRRP. Hospitals that were not considered for the HRRP penalties, included hospitals with too few cases to evaluate (less than 25 cases during the entire 3-year assessment period), psychiatric, rehabilitation, long term care, children's, cancer, critical access hospitals, and all hospitals in Maryland. in addition, we exclude hospitals with less than 50 cases during the entire 3-year assessment period, because CMS used a Bayesian shrinkage method that assigns these small hospitals a score close to the threshold but below 1.0.

In equation (1), j in an index of hospitals and t is an index of years. The model indicates that the share of admissions readmitted within 30 days (*READMIT*) in hospital j and year t depends on hospital fixed effects (α_j), year fixed effects (δ_t), demographic characteristics (age, race/ethnicity, gender) of patients (*X*) in hospital j in year t, and an indictor for whether the hospital was penalized by the HRRP (*PENALTY*). The demographic characteristics, which vary over time for each hospital, are indicators for age under 65, age 65-70, age 70-75, age 80-85, age over 85, share Hispanic, share Black, share White, share other race, and share female.

For studies that used proxies for being penalized (e.g., Demiralp et al. 2017; McGarry et al. 2016) the variable *PENALTY* is replaced with an indicator for having that proxy (e.g., high baseline readmissions). These studies estimate differences in outcomes for hospitals that have different amount of exposure and probability of being penalized and not differences between hospitals that were and were not treated, or what may be referred to as the treatment-on-treatment effect (i.e., effect of penalty on readmissions). It is necessary to note that the difference-in-difference model assumes that unpenalized hospitals do not respond. However, this may not be the case, as we noted in the discussion of the conceptual model. If unpenalized hospitals are responding, then the difference-in-difference model yields an estimate of the effect of the HRRP for penalized hospitals versus unpenalized hospital and not the total effect of the HRRP. If so, then the estimate will be smaller than the true effect.

The first year that the HRRP penalty was imposed was 2012 (officially October 1, 2011 was the start of penalty period, but we date the start to August 1, 2011 when it was announced which hospitals would be penalized).[11] So, the variable *PENALTY* is zero prior to 2012 for all hospitals and equal to one in 2012 and subsequent years for hospitals that were penalized.

[11] Several studies, particularly those that use an interrupted time-series design, use the start of the ACA (March 2010) as the post-HRRP period. No justification is given for this choice, which assumes that hospitals immediately started to respond to the HRRP despite not knowing if they would actually be penalized. CMS Hospital Compare data on readmissions by HRRP condition were available starting in July 2009 and provided some information about likely penalty status. This issue is addressed below when we conduct event-study analyses that shows the effect of the HRRP in all years from 2008 to 2014.

We estimate equation (1) for all targeted conditions, all non-targeted conditions and for each of the three conditions targeted at the beginning of the HRRP: AMI, heart failure (HF) and pneumonia (PN). Estimates are obtained using weighted least squares where the weight is the number of patients admitted to the hospital in a year. Standard errors are constructed using a robust-cluster approach that accounts for potential non-independence of observations within hospital.

We also estimate an event-study specification of equation (1) that allows there to be separate year effects for penalized and unpenalized hospitals. The regression model for this analysis is:

$$(2)\quad READMIT_{jt} = a_j + \delta_t + X_{jt}\Gamma + \sum_{t=2008}^{2014} \beta_t PENALTY_{jt} + e_{jt}$$

Equation (2) is identical to equation (1) except that we allow the difference between penalized an unpenalized hospitals to differ by year. Estimation proceeds as described above (weighted least squares and robust-cluster standard errors).

Figure 1 describes trends in 30-day readmissions by the HRRP penalty status in round 1 for the years 2006-2007 to 2013-2014 for each of the targeted conditions and for all targeted conditions aggregated. Across AMI, Heart Failure and Pneumonia, the figure shows that readmissions were decreasing among hospitals penalized and unpenalized prior to the HRRP. Additionally, there were diverging trends between the two groups – the unpenalized group in some years decreased readmissions by more than the penalized group. In Figure 2, we show trends in readmission rates pre and post the HRRP, by tertiles of baseline readmissions in 2006-2007, which has been used to classify hospitals more or less likely to be penalized. Here too, there is evidence of divergence in trends prior to the actual HRRP period. These simple time-series graphs highlight the parallel trend assumption and suggest that it may be violated. We turn next to regression estimates.

Table 1 report estimates of equation (1). We show results for each of the three HRRP conditions that were targeted in the first round of the HRRP program—AMI, heart failure and pneumonia—and for all targeted and all non-targeted conditions. We show estimates from models that use an indicator of whether a hospitalized was penalized for a specific condition (i.e., excess readmissions for that condition above average), which we refer to as condition-specific penalty, and for an indicator that the hospital was penalized for any (or all) conditions, which we refer to as hospital-wide penalty. We show simple DD estimates in the top row and estimates from an event-study specification in the bottom part of the table.

There are a few notable points revealed by estimates in Table 1. First, as indicated by the simple DD estimates in the top row, hospitals that were penalized by the HRRP, either measured by condition-specific or hospital-wide penalty, had significantly lower readmissions post-penalty than hospitals that were not penalized. The declines were on the order of 5% (hospital-wide penalty) to 10% (condition-specific penalty). Second, the declines in readmissions were very similar across targeted conditions, and there was a significant, although smaller (2%) decline in readmissions for non-targeted conditions. Third, the magnitude of the decline in readmissions was approximately 100% larger when the condition-specific penalty was used instead of the hospital-wide penalty. The larger effects associated with the condition-specific penalty suggests that hospitals respond largely on a condition-specific basis in response to a condition-specific penalty.

How do these results fit with the conceptual issues raised earlier? Consider the magnitudes of the DD estimates. Is it plausible that a hospital could decrease readmissions by 10%? If we assume that 27% of readmissions are preventable (e.g., van Walraven et al. 2011), a 10% decline in readmissions suggests that a penalized hospital was able to intervene and prevent 37% of the preventable readmissions. This is a relatively large impact, perhaps too large given the weak evidence related to effective interventions to reduce readmissions and the other considerations described earlier that suggest that HRRP is likely to have small effects. It is also somewhat surprising that readmissions declined by approximately the same magnitude across all three targeted conditions. Given that the ability to identify and effectively intervene to prevent a readmission likely differs across conditions, as does the likely cost of intervention, we do not expect this to be the case.

The event-study estimates suggest the answer to our skepticism regarding these DD estimates. As indicated by estimates of the effect of the condition-specific penalty, it is clear that trends in readmissions between penalized and unpenalized hospitals across all three conditions were diverging from 2008 onward, which is prior to passage of the ACA in 2010 and prior to any hospital actually receiving a penalty. The fact that the divergence occurred before the ACA and since 2008 casts doubt on the assumption that the decline in hospital readmissions that occurred in what is often referred to as the "anticipation" period between passage of ACA (i.e., 2010) and the date when penalties were actually imposed and highlighted by previous studies (e.g., Zuckerman et al. 2016) is due to the HRRP. In addition, Ody et al. (2019) show that much of the post-2011 decline in risk-adjusted readmissions was due to a change in electronic coding of comorbidities and not due to the HRRP. When the hospital-wide penalty is used, there is less, although still substantial, evidence of divergent trends in readmissions between penalized and unpenalized hospitals. Note that in these models, approximately 80% of hospitals were penalized and the comparison is between hospitals that were penalized for any or all conditions and hospitals that were not penalized at all.

In sum, the DD approach for studying the effect of the HRRP on readmissions and other outcomes does not appear to be valid because the underlying assumption of parallel trends between penalized and non-penalized hospitals does not seem to be applicable, at least in studies using national data (as opposed to single state studies) such as Desai et al. (2016) and Wasfy et al. (2017). Indeed, evidence of divergent trends was reported in Desai et al. 92016) but not highlighted. Trends in readmissions between penalized and non-penalized hospitals were diverging prior to 2010 when the HRRP was announced as part of the ACA and prior to when the first HRRP penalty was imposed in 2012. The magnitude of the divergence in trends prior to the HRRP was often as large as the post-HRRP divergence, and while the direction of the diverging trends was not always the same, it is the fact that there is divergence that is critical and not the direction of the divergence. It is not appropriate to use the *ex-post* results from a likely flawed research design to infer the true estimate because the counterfactual outcome remains unknown. While it is clear that readmissions were falling over the period of analysis, they were falling for targeted and non-targeted conditions and falling at a differential rate for hospitals that were and were not penalized up to 3 years before the policy became effective. The article by Ody et al. (2019) proposed one explanation for this uniform decline post-2011 in adjusted readmissions—a change in electronic coding of transactions. readmissions. Based on these findings, estimates of the effect of the HRRP from DD analyses are not likely to be informative of the causal effect of the HRRP.

Finally, while estimates in Table 1 are from a difference-in-differences approach that compares penalized and unpenalized hospitals (i.e., effect of treatment-on-treated), we also present evidence below that difference-in-differences models that rely on proxies for being penalized, specifically, baseline readmissions (e.g., Demiralp et al. 2017) are also likely invalid. We show this as part of the analysis of the Gupta (2021) study because that study incorporates a difference-in-differences approach that relies on the parallel trend assumption for hospitals with low- and high-baseline readmissions. Similarly, we show next that studies that compared changes over time in targeted and non-targeted conditions (e.g., Zuckerman et al. 2016) are also unlikely to be valid because of differential pre-trends between targeted and non-targeted conditions.

The Validity of the Difference-in-differences-in-differences Design and Prior Evidence Using It

A few studies have used a triple difference research design (Chen and Grabowski 2017; Mellor et al. 2017; Ody et al. 2019; Gupta 2021). The DDD approach can be viewed as the difference in two separate DD designs where one DD model is estimated for a treated group and another DD model is estimated for an untreated group. Studies that use the DDD approach to study the HRRP consider admissions for HRRP targeted conditions as treated group and non-targeted conditions (e.g., GI admissions) as the untreated group. Viewed this way, we can illustrate the DDD model approach using two DD models. The first DD model is for the treated group (TREAT=1) and is similar to equation (1):

$$(3) \quad READMIT_{jt}(TREAT=1) = a_j + \delta_t + X_{jt}\Gamma + \gamma Z_{jt} + \beta PENALTY_{jt} + e_{jt}$$

To denote that it is the treated group of conditions we use the TREAT=1 notation shown in equation (2). The only difference between equations (1) and (3) is that we have included another variable (Z) in equation (3) to represent unmeasured factors that affect readmissions. It is the presence of such factors that biases DD estimates, for example, because it is correlated with the penalty and, therefore, will cause trends in readmissions to diverge between treated and untreated hospitals, as demonstrated above.

An analogous DD model for the untreated group (i.e., TREAT=0) is:

(4) $READMIT_{jt}(TREAT=0) = a'_j + \delta'_t + X_{jt}\Gamma' + \gamma Z_{jt} + e_{jt}$

Equation (4) omits the penalty indicator because the penalty is assumed to have no effect on admissions for conditions that are not targeted by the HRRP. We also denote new parameters using the superscript ´. We explore shortly the consequences if this assumption is violated, but for now we proceed as if the assumption is valid. The key aspect of equation (4) is that the coefficient on the unmeasured factor (Z) is the same as that in equation (3) for the treated group—i.e., parallel trends between targeted and non-targeted conditions.

Taking the difference between equations (3) and (4), or the difference-in-differences-in-differences, yields:

(5) $READMIT_{jt}(TREAT=1) - READMIT_{jt}(TREAT=0) = \ddot{\alpha}_j + \ddot{\delta}_t + X_{jt}\ddot{\Gamma} + \beta PENALTY_{jt} + v_{jt}$

In equation (5), the dependent variable is the difference in readmissions for treated and untreated conditions in hospital j and year t. The DDD model includes the two-way interactions between hospital and treated conditions and year and treated conditions, as indicated by the inclusion of hospital and year effects in the difference model, which are differentiated by the use of the (..) superscript. The effect of unmeasured factors, which was assumed to be the same for both treated and untreated conditions, is eliminated by taking the difference between equations (3) and (4). If the assumption of common unmeasured effects holds, the DDD estimate yields the true effect of the HRRP penalty on readmissions.

A simple test of the validity of the DDD approach is to estimate a slightly different version of equation (5):

(6) $READMIT_{jt}(TREAT=1) = \rho READMIT_{jt}(TREAT=0) + \ddot{\alpha}_j + \ddot{\delta}_t + X_{jt}\ddot{\Gamma} + \beta PENALTY_{jt} + v_{jt}$

The difference between equations (5) and (6) is that we do not restrict the coefficient on readmissions for non-treated conditions to be equal to one, as does the standard DDD model. As we show in the Appendix (see Appendix A), the validity of the standard DDD approach depends on whether ρ=1, which is a test of whether the effect of unmeasured factors on readmissions is the same for treated and untreated conditions.

Before presenting estimates of DDD models, we review the bias of a DDD approach if the policy actually affects untreated conditions, for example, because the hospital responds to the HRRP by reallocating resources toward treated conditions that are part of the performance indicator and away from untreated conditions. In this case, the model for the untreated units (i.e., TREAT=0) is:

$$(7)\quad READMIT_{jt}(TREAT=0) = a'_j + \delta'_t + X_{jt}\Gamma' + \gamma Z_{jt} + \pi\beta PENALTY_{jt} + e_{jt}$$

In equation (7), we have allowed the HRRP penalty to affect untreated conditions and show it as a multiplicative factor (π) of the effect of the HRRP penalty on treated conditions (β) to simplify the algebra that comes next. Now, taking the difference between equations (3) and (7) yields:

$$(8)\quad \Delta READMIT_{jt} = \ddot{\alpha}_j + \ddot{\delta}_t + X_{jt}\ddot{\Gamma}_k + \beta(1-\pi)PENALTY_{jt} + v_{jt}$$

In this case, even if the unmeasured factor has the same impact on the treated and untreated conditions, the DDD estimate will be biased because it measures the difference in the treatment effects for treated and untreated (differentially treated) conditions. The test of the validity of the DDD approach remains the same, although this is just testing whether unmeasured factors affect readmissions for treated and untreated conditions the same and not whether there is an effect of the HRRP on untreated conditions.[12]

Table 2 reports DDD estimates that use two different sets of conditions as untreated: all non-targeted readmissions, as was used, for example, in Desai et al. (2016, Zuckerman et al. (2016) and Ody et al. (2019), and readmissions for GI conditions, which were used by Carey and Lin (2015) and Gupta (2021). We report estimates from equations (5) and (6) with the latter providing an explicit test of the validity of the DDD approach. For equation (5), which is the standard DDD specification, we show estimates only for the specification that used all non-targeted conditions as untreated group.

DDD estimates in Table 2 are similar to the DD estimates reported in Table 1 and suggest that an HRRP penalty reduced readmissions significantly. While the DDD estimates obtained using all non-targeted conditions as untreated conditions are smaller than DD estimates, they remain relatively large

[12] The same problem applies to difference-in-differences analyses that define treatment using baseline readmissions (e.g., Demiralp et al. 2017; Chen and Grabowski 2017), baseline Medicare admissions (e.g., McGarry et al. 2016), or any other variable used to classify hospitals.

and indicate that hospitals that were penalized were able to reduce readmissions by between 5% and 8%. As noted, for several reasons we expect the HRRP to have modest effects, which this effect size may be if the share or preventable readmissions is larger than most believe, or if there are more efficacious interventions than widely believed. And it is also the case that these DDD estimates are near identical across conditions, which we have argued is unlikely unless there is a common intervention to reduce readmissions across conditions. Estimates using the condition-specific penalty indicator (top panel) are again approximately twice the size of estimates from models that use the hospital-wide penalty indicator (bottom panel). DDD estimates are also relatively similar whether we use all non-targeted readmissions or GI readmissions, as untreated conditions, but the DDD estimates obtained from models using GI readmissions are larger and almost identical to DD estimates in Table 2.

Coefficients on readmissions for untreated conditions (equation 6 specification) indicate that the DDD model that uses GI readmissions as untreated admissions and a control is not valid, as estimates of the effect of GI readmissions on HRRP targeted conditions is significantly different from one in all cases. It is notable that DDD estimates from these models are larger than from models using all non-targeted conditions and near identical to DD estimates. In analyses that use all non-targeted conditions as the control, we cannot reject the hypothesis that estimates of the effect of these other conditions on targeted readmissions is equal to one except for the case of pneumonia. However, as shown in the Appendix, in some circumstances, the underlying assumption of the DDD specification may still be violated even if we cannot reject $\rho=1$. While rejecting $\rho=1$ always indicates an invalid research design, not rejecting this equality does not always indicate that the design is valid.

To further asses the common trends assumption of the DDD approach, we also conducted event-study analyses in the DDD context to assess the validity of the DDD specification. Table 3 reports the event-study results in the context of the DDD analysis. For this analysis, we use all non-targeted conditions as the non-treated control because it was clear from Table 2 that analyses that use GI as the control condition are not valid. First, we note that for five of the seven analyses, there is clear evidence that the DDD design is not valid, as supported by significant estimates associated with the pre-2012 penalty-by-year interactions.[13] A similar finding of diverging trends between targeted and non-targeted conditions was reported by Zucherman et al. (2016) in their interrupted time series analyses, but the divergence was not highlighted. Ody et al. (2019) also reported differential trends in pre-HRRP period between targeted and non-targeted conditions.

Overall, estimates in Tables 2 and 3 demonstrate that the DDD design to study the HRRP is unlikely to be valid. Although, our analysis applies to only studies using national data and not single-state studies (e.g., Chen and Grabowski 2017 and Mellor et al. 2017). The evidence in Tables 2 and 3 also implies that DD analyses that use targeted and non-targeted conditions in an intention-to-treat framework, which measures the difference in readmissions between conditions more or less likely to be penalized (and not whether a hospital was penalized) are also unlikely to be valid (e.g., Demiralp et al. 2017; Ody et al. 2019). Moreover, the DDD relies on the untested assumption that the control conditions are not affected by the HRRP. For both reasons, DDD estimates of the effect of the HRRP are not likely to be informative of the causal effect of the HRRP.

The Validity of the Instrumental Variables Design and Evidence Using It

One previous paper used an instrumental variables (IV) approach (Gupta 2021). The way that the IV approach was implemented by Gupta (2021) is closely related to the DD approach. Specifically, Gupta (2021) instruments for the HRRP penalty (actually a forecasted penalty, which we discuss further below) using baseline hospital-specific readmissions in 2007 (2006-07) interacted with a post-HRRP penalty indicator. This approach is similar in spirit to McGarry et al. (2016) and others that use baseline characteristics to divide hospitals by whether they are more or less likely to be penalized and conduct a DD analysis (this would be the reduced form estimate in Gupta 2021).

Little theoretical justification was provided for the validity of this instrument (e.g., exclusion restriction), which is a Bartik (1991)-like instrument. The idea underlying this approach is that baseline

[13] In Table 2, the test of the validity of the DD only identified a problem in the case of pneumonia. The failure to detect the invalid design using the specification in Table 2 highlights that this test may not always identify an invalid design (see Appendix).

characteristics make it more or less likely that a hospital will be penalized when the HRRP comes into effect and, therefore, examining changes in readmissions pre-to-post HRRP for hospitals stratified by baseline characteristics can be used to obtain an estimate of the effect of the HRRP. Of course, this approach still relies on the assumption that absent the HRRP, the pre-to-post changes in readmissions of hospitals stratified by baseline characteristics would be the same.

Algebraically, the IV model used by Gupta (2021) can be illustrated using the following regression model specifications:

$$(9)\quad READMIT_{jt} = a_j + \delta_t + X_{jt}\Gamma + \beta(\widehat{PENALTY_j * POST_t}) + e_{jt}$$

Equation (9) is the model of interest and identical to equation (1), although in this case treatment is indicated using an interaction between a hospital-specific indicator of whether the hospital was penalized in the first round of the HRRP and an indicator that the period is 2012 or after (*POST*). Also note that instead of using the actual HRRP penalty indicator a predicted penalty (i.e., predicted penalty-by-post) is used. The predicted penalty is obtained by the following first-stage regression:

$$(10)\quad PENALTY_j * POST_t = \breve{a}_j + \breve{\delta}_t + X_{jt}\breve{\Gamma} + \pi(READMIT0607_j * POST_t) + u_{jt}$$

In equation (10), we have used the superscript (˘) to denote new parameters. The key aspect of this model is the use of baseline (2006-07) readmissions interacted with a dummy variable indicating the post-penalty period (2012) as an instrument. As Gupta (2021) shows, this instrument is a strong predictor of the HRRP penalty (or forecasted penalty). Hospitals with relatively more admissions in 2006-07 are more likely to be penalized by the HRRP in 2012, as presumed by McGarry et al. (2016) and others.

The critical issue for the IV approach to be valid is whether the exclusion restriction is valid. Validity requires that baseline readmissions affect readmissions in subsequent years (e.g., post-HRRP) only because it makes it more likely that a hospital will be penalized. In the absence of the penalty, hospital readmissions would be the same for hospitals with different levels of baseline admissions, which is just the parallel trend assumption of a DD analysis. Given that the model described by equations (9) and (10) is just identified, there is no way to definitively test this assumption. However, evidence of the likely validity of the exclusion restriction can be obtained in a variety of ways. One way is to conduct the IV analysis using an event-study specification (e.g., Jaeger et al. 2020). In this case, the outcome equation is:

$$(11)\ READMIT_{jt} = a_j + \delta_t + X_{jt}\Gamma + \sum_{t=2008}^{2014} \beta_t (PE\widehat{NALTY_t * YEAR_t}) + e_{jt}$$

Equation (11) is an event-study specification often used in a DD analysis. It allows the difference in readmissions between penalized and unpenalized hospitals to differ in each year from 2008 to 2014 (leaving one year omitted as reference). The only difference between this specification and the specification of equation (9) is that the *POST* dummy variable is replaced with a set of year dummy variables. The analogous first stage equation to predict the penalty variables is:

$$(12)\ PENALTY_j * YEAR_{t=k} = \breve{a}_j + \breve{\delta}_t + X_{jt}\breve{\Gamma} + \sum_{t=2008}^{2014} \pi_t (READMIT0607_j * YEAR_t) + u_{jt}$$

Equation (12) is used to predict the interaction between the HRRP penalty and year indicators. In this equation too, the POST dummy variable (of equation 10) has been replaced with a set of year dummy variables. There will be six (k=6) first-stage equations for each of the six (seven years with one omitted as reference) interactions shown in equation (11). For each first stage equation, there are six instruments (one year omitted) constructed by interacting year dummy variables with the baseline readmissions, but the only instrument that, in practice, matters in terms of significance is the interaction between baseline readmissions (*READMIT0607*) and year when the year of this interaction equals the year of the interaction with the penalty (*PENALTY*YEAR*).[14]

Estimates from equation (11) can be used to test the validity of the IV approach. Estimates of the coefficients on the interactions between the penalty indicator and year (*PENATY*YEAR*) in years prior to the HRRP (2012) should be zero. If not, then it suggests that trends in hospital readmissions between hospitals with different levels of baseline admissions differ prior to the penalty, which implies that such differences would persist in years post-penalty even without the penalty.

[14] We do not restrict the estimates of the "off" year interactions to be zero.

A second way to assess the validity of the IV approach is to exploit the multiplicity of potential instruments stemming from the nature of the instrumental variables approach (Goldsmith-Pinkham et al. 2019). In his application, Gupta (2021) used baseline readmissions to stratify hospitals more or less likely to be penalized, but any number of baseline variables could be used as instruments and using these additional instruments makes the model an over identified one that can be used to test the likely validity of the exclusion restriction.

An expanded IV model that uses multiple potential instruments is the following:

$$(13)\ READMIT_{jt} = a_j + \delta_t + X_{jt}\Gamma + \sum_{k=1}^{K} \lambda_k'(W_{jk} * POST_t) + \beta P\hat{ENALTY}_{jt} + e_{jt}$$

and

$$(14)\ PENALTY_{jt} = \breve{a}_j + \breve{\delta}_t + X_{jt}\breve{\Gamma} + \sum_{k=1}^{K} \pi_k'(W_{jk} * POST_t) + u_{jt}$$

Equations (13) and (14) represent the overidentified IV model. The variable *W* is a vector representing several hospital-specific baseline characteristics including readmissions, which is the instrument used by Gupta (2021). An overidentification test can be conducted by including all of the baseline characteristics except readmissions in the second stage (equation 13) and seeing whether the coefficient estimates on the interactions in equation (13) are statistically significant. Significant estimates suggest that changes in readmissions before and after the HRRP penalty differ by baseline hospital characteristics conditional on being penalized, which is evidence against the exclusion restriction in the case of baseline readmissions.

Table 4 presents OLS (difference-in-difference) estimates in the top row and IV estimates in panels two and three.[15] Estimates from the just identified model are presented in the middle panel and estimates from the model given by equations (12) and (13) are presented in the bottom panel. We also show estimates when the actual penalty is used as an indicator of treatment (left side of table) and when the actual penalty is replaced by a forecasted, or expected penalty. As noted above, Gupta (2021) did not use the actual penalty as an indicator of treatment, but instead used the forecasted penalty, which is a

[15] Appendix Table 1 presents analogous IV estimates that use the hospital-wide penalty instead of the condition-specific penalty. Results from these models are similar to those in Table 3.

function of readmissions in the period before the penalty period and used by CMS to determine the penalty.[16] The assumption for using a forecasted penalty is that hospitals used the previous period's readmissions to forecast their probability of being penalized and responded according to the probability of being penalized and not the actual penalty.[17] The instrument remains the same in this case.

OLS estimates in Table 4 have already been discussed earlier in the section on DD analyses and, therefore, we will not discuss them again. The just identified IV estimates in the middle panel are relatively large. When the actual penalty is used, the condition-specific HRRP penalty is associated with: a 31% reduction in AMI readmissions; a 13% reduction in heart failure readmissions; and a 15% reduction in pneumonia readmissions. We again note that effect sizes this large may be implausible if the actual number of preventable readmissions is 27% or even less. Just identified IV estimates obtained using the forecasted penalty are similar in magnitude to those that use the actual penalty. When the predicted penalty is used, the condition specific HRRP penalty is associated with a 22% reduction in AMI readmissions; a 9% reduction in heart failure readmissions and a 18% reduction in Pneumonia readmissions. These IV estimates are very close to what Gupta (2021) reported, although the IV estimate here pertaining to heart failure is larger than in Gupta (2021. Specifically, Table 2 in Gupta (2021) reports the following coefficients (standard errors) for AMI, HF and PN respectively: -0.0349 (0.006), -0.01 (0.003), -0.028 (0.003). Our estimates for this analysis (reported in Table 4) for AMI, HF and PN respectively are: -0.041(0.006), -0.023(0.003), -0.030 (0.004).

If the forecasted penalty was the correct measure to use instead of the actual penalty, one would expect the estimate to be larger when the expected penalty is used because it better classifies hospitals to treated and untreated groups, but this is not the case two out of three times.

---

[16] Gupta (2021) used a kernel regression of the actual penalty on readmissions in the prior three-year period to construct an expected penalty. We used a logit regression model in this analysis to construct the expected penalty.
[17] While this is certainly possible, no evidence to substantiate this claim was provided. Moreover, whether a hospital was penalized was known at the start of the penalty period (August 2011). So, in the penalty period onward, the hospital knew whether it was penalized. Only if the hospital responded earlier than August 2011 would using the expected penalty be appropriate. But CMS used data through most of 2011 to determine the excess readmission ratio. Therefore, the hospital would have had to be calculating their excess readmission ratio contemporaneously to CMS and also know the distribution of the same ratio for all hospitals to predict whether it would be penalized.

The last set of estimates in Table 4 are from the event-study specification of the IV model and these are shown in the bottom panel. The key estimates here are those on the interactions between the penalty indicator and year in years prior to 2012. A valid IV design would require that these estimates be zero. For both AMI and heart failure, estimated coefficients on these interaction terms are statistically significant and large suggesting that the exclusion restriction of the IV approach is unlikely to be valid. Only in the case of pneumonia does the IV design appear to hold up. Again, very similar results are obtained when the forecasted penalty is used as an indicator of treatment instead of the actual penalty. In Appendix Table 1, we replicate the analysis of Table 4, but use as the measure of treatment an indicator of whether the hospital was penalized for any condition (i.e., hospital-wide penalty). Estimates in that table are qualitatively the same as in Table 4 and indicate the same problems with using the IV approach. It is important to note that the lack of validity of the IV approach that uses baseline readmissions as an instrument also provides evidence that DD intention-to-treat analyses based on classifying hospitals using baseline readmissions are also invalid (e.g., Demiralp et al. 2017; Ody et al. 2019).

Table 5 presents IV estimates from the model of equations (13) and (14).[18] This is a just identified model with the excluded instrument of baseline readmissions interacted with a dummy variable indicating the post-penalty period. Several other interactions between baseline characteristics (shares of race/ethnicity of patients; bed size of hospital; whether hospital is a teaching institution; and share of patients with diabetes and pneumonia) and a post-penalty dummy variable are included in both the first and second stages. For continuous baseline measures (e.g., diabetes, pneumonia, bed size), we used dummy variables indicating tertiles of the distribution of the baseline characteristic.[19] Because any or all of these baseline characteristics could plausibly be used as instruments, including them in the second stage allows for tests of over identification.

[18] Appendix Table 2 presents analogous IV estimates that use the hospital-wide penalty instead of the condition-specific penalty. Results from these models are similar to those in Table 5.
[19] Using different thresholds and classification did not change results.

The IV estimates from this model are similar to those in Table 4 (shown in top row of Table 5) and indicate that the HRRP penalty (actual) is associated with a 33% reduction in AMI readmissions; a 15% reduction in heart failure readmissions and a 15% reduction in pneumonia readmissions. Notably the IV estimate for this model when the expected penalty is used are much larger than the corresponding estimates in Table 4 suggesting declines in AMI readmissions of 61% and pneumonia admissions of 50%. Effects of this magnitude seem implausible given our earlier discussion. More importantly, for each outcome, there is evidence that the exclusion restriction may not be valid as some of the interactions between baseline characteristics and a dummy variable indicating the post-penalty period are statistically significant. In the case of AMI, estimates associated with interactions between share of patients who are black and bed size of hospital are statistically significant; and for heart failure and pneumonia the significant estimates pertain to diabetes and bed size. Estimates in Appendix Table 2, which use an indicator for a hospital wide penalty, or the forecasted hospital wide penalty, provide similar evidence that the IV approach using baseline readmissions (in this case all readmissions) interacted with a post-HRRP dummy variable is likely invalid.

Overall, estimates in Tables 4 and 5 suggest that the IV approach is unlikely to be valid. There is substantial evidence that the exclusion restriction associated with the use of interactions between one or more baseline characteristics and time is unlikely to be valid. Baseline (2006-2007) readmissions that predict the penalty, predict reductions in readmissions up to 3 years before the HRRP was implemented. Also, characteristics that make the hospital more or less likely to be penalized by the HRRP also influence readmissions directly (perhaps because of being correlated with confounders) and not just because of the HRRP penalty. This finding is consistent with the DD analysis that showed that penalized and unpenalized hospitals had different trends in readmissions prior to the HRRP (and prior to ACA even). So, as with the DD and DDD analyses, the IV estimates are unlikely to be informative about the causal effect of the HRRP on readmissions.

A Novel Approach to Study the Effect of the HRRP: The Regression-kink (RK) Design

The HRRP penalizes hospitals with excess readmissions, which are defined using the ratio of actual readmissions to expected readmissions, which are the predicted number of readmissions given a hospital's patient's comorbidities.[20] The excess readmissions ratio is measured in a prior three-year period. A hospital with an excess readmission ratio in the prior three-year period of less than or equal to one is not penalized, but a hospital with an excess readmission rate greater than one is penalized, and penalties grow with the excess readmission ratio. Therefore, there is a "kink" in the relationship between the size of the HRRP penalty and the excess readmission ratio at the value of one. The penalty is zero when the excess readmissions is below one and the penalty increases with the excess readmissions ratio for values greater than one. Accordingly, if the HRRP penalty causes hospitals to reduce readmissions, then there should be a similar "kink" in the relationship between readmissions and the excess readmission ratio. The "kink" structure of the HRRP penalty determination lends itself to the use of a regression-kink (RK) research design (Card et al. 2015).

Identification of the effect of the HRRP in the RK design comes from a change in the effect (slope) of the excess readmission ratio on readmissions at the threshold between when the penalty is zero or becomes positive (Card et al. 2015). The main advantage of the RK approach is that, in contrast to previous studies, the RK uses variation in the penalty amount due to a kink in the penalty formula and does not rely on any type of parallel trend assumption, as do the DD, DDD and IV approaches that have been used in the past and shown above to be problematic.

The RK approach also allows us, in contrast to all prior studies, to obtain an estimate of the effect of the size of the penalty on readmissions. Estimating the effect of the size of the penalty is arguably a more interesting parameter given that a large majority (e.g., 83%) of hospitals are penalized. The large share of hospitals penalized suggest that the effect of size of the penalty is the important parameter to

[20] A detailed outline of the readmissions measure and methodology is described on the AHRQ website: https://www.qualitymeasures.ahrq.gov/summaries/summary/49197/unplanned-readmission-hospitalwideallcause-unplanned-readmission-rate-hwr?q=readmissions.

know and not simply whether a hospital was penalized or not. This fact has largely been ignored in prior studies that almost uniformly analyze the effect of being penalized, or the effect of the likelihood of being penalized.

The regression kink design is implemented using regression methods and model specifications as follows (Card et al. 2015). The equation of interest is:

$$(15)\quad READMIT_{jt} = b_0 + b_1 RATIO_{j(t-1)} + b_2 PENALTY_{jt} + n_{jt}$$

Equation (15) indicates that readmissions of hospital "*j*" in year "*t*" depends on the hospital's excess readmission ratio (*RATIO*) and the amount of penalty (*PENALTY*) that hospital "*j*" in year "*t*" incurred. Estimates of equation (15) are likely biased because the amount of the penalty is likely correlated with unmeasured factors that influence both the penalty and readmissions, such as hospital size and patient characteristics. To address the possible bias, we instrument for the penalty using a RK specification:

$$(16)\qquad PENALTY_{jt} = b_0' + b_1' RATIO_{j(t-1)} + b_2' (RATIO_{j(t-1)} * ABOVE_j) + v_{jt}$$

In equation (16), the size of the readmission penalty (*PENALTY*) of hospital "*j*" in year "*t*" depends on the excess readmission ratio (*RATIO*) in year "*t-1*" and the interaction between a dummy variable indicating that the excess readmission ratio is greater than one (*ABOVE*). This regression model mimics the CMS formula that mechanically determines the readmission penalty. The readmission penalty is zero when the excess readmissions ratio is less than or equal to one and then the penalty is a linear function of the excess readmission ratio after the threshold of one.[21] The instrumental variables approach in this context is valid as long as the interaction term in equation (16) is excludable—i.e., does not belong in equation (15). We believe this is a reasonable assumption because there is no reason (theoretical or empirical) to expect a change in the

[21] We also estimated a model that uses a quadratic specification of the excess readmission ratio and analogous interaction terms. The linear model cannot be rejected.

relationship between the excess readmissions ratio at the threshold (*RATIO=1*) other than because of the change in the penalty.

Table 6 reports the estimates from equation (16) for each of the three HRRP conditions and verifies that the regression mimics the penalty formula; estimates show that the coefficient on the excess readmission ratio (b_1') is virtually zero, which is expected because the penalty is zero prior to the excess readmission threshold. After the threshold, the HRRP penalty grows with the excess readmissions. Estimates indicate that every 0.10 unit increase in the excess readmissions ratio, which is about a one standard deviation change, is associated with approximately a .3% to .4% increase in the penalty (e.g., from 0 to .4%). The relationship between the change in excess readmission ratio and penalty is relative similar across conditions. Estimates indicate that the maximum penalty, which was 1% (1.0) in 2012, is reached when the excess readmissions ratio is approximately 0.3 units above the penalty threshold.

The reduced form effect of the HRRP on readmissions is then obtained using the following regression:

$$(17)\quad READMIT_{jt} = \ddot{b}_0 + \ddot{b}_1 RATIO_{j(t-1)} + \ddot{b}_2 (RATIO_{j(t-1)} * ABOVE_j) + v_{jt}$$

In equation (17), the share of admissions that are readmitted (*READMIT*) in hospital "*j*" in year "*t*" depends on the excess readmission ratio, and the interaction between the indicator of the threshold and the excess readmission ratio. If there is a causal effect of the HRRP penalty on outcomes, the coefficient ($\ddot{b}_2$) should be negative.

Note that the dependent variable is measured in year "*t*" (or after) which refers to the first year after the penalty was announced by CMS and known with certainty by the hospital (August 2011). In round 1, CMS announced the penalty in August 2011 based on an analysis of data from 2008 to 2011 (denoted as t-1 in equations 14-16), but penalties did not start until October 2012. Given this announcement and known penalty status, we use years spanning Aug. 2011 to Aug. 2014 as the post penalty period. However, as we noted earlier, there may be some uncertainty about the likelihood of being penalized, which would affect the interpretation of the RK estimates. We discuss this below.

One complication in applying the RK design is that the readmissions penalty is a function of the excess readmission ratio for three conditions: Heart Attack (AMI), Heart Failure (HF) and Pneumonia (PN). Thus, a hospital can be penalized if it has an excess readmission ratio greater than one on any, or all, of these conditions. This circumstance makes it difficult to identify the appropriate counterfactual hospital. For example, consider a hospital with excess readmission ratios of 0.9, 1.01 and 1.3 for AMI, HF and PN, respectively. For this hospital, the ideal counterfactual hospital might be one with excess readmission ratios of 0.9, 0.99 and 1.3 for AMI, HF and PN, respectively. This example reveals the dimensionality problem in defining appropriate comparison hospitals if we used all three excess readmission ratios.

To address this issue, we stratify the sample and focus on the effect of one cause of a penalty at a time. For example, to estimate the effect of a hospital incurring a penalty due to excess AMI readmissions, we limit the sample to hospitals with excess readmission ratios less than one for HF and PN (i.e., not penalized for HF and PN). Thus, we have a sample of hospitals that we can order with respect to the AMI excess readmission ratio that all have excess readmission ratios for HF and PN that are less than one.[22] One advantage of this approach is that it is straightforward. It allows for the use of one excess readmission ratio as the running variable, and, therefore, relies on a standard regression kink design. Second, we are making a particularly interesting comparison between a penalized hospital and what may be considered good performing hospitals—those that were not penalized for any condition. Out of the 2,569 penalized hospitals by CMS in round 1 (FY2013), 234 hospitals were penalized for only having excess AMI readmissions, 362 hospitals were penalized for only having excess HF readmissions and 315 hospitals were penalized for only having excess pneumonia readmissions. Therefore, the RK analysis includes 35% of hospitals that

---

[22] Including controls for the excess readmission ratio for the two other conditions (HF and PN in the AMI sample) does not alter estimates.

were penalized. The fact that we only use a portion of penalized hospitals, and that they are only penalized for one condition, may limit the generalizability of our findings.

Table 7 presents the results of the reduced form RK regressions for readmissions. We show results for several years: two prior years (2010 and 2011) and three post-HRRP penalty years (2012-2014). In each panel, the 2009-11 excess readmissions ratio is used as the running variable. This is the excess readmissions ratio used to determine the round 1 HRRP penalty. We include analyses of readmissions in years prior to the actual implementation to assess whether we observe any significant effects in those years, which would suggest an invalid research design (Landais 2015). In fact, we do not find any effects in these years.

The most noteworthy finding revealed in Table 7 is that all estimates associated with the interaction between the excess readmissions ratio and the indicator for being above the penalty threshold are small and statistically insignificant (at 0.05 level). There is only one marginally significant estimate and that pertains to heart failure in 2014. The magnitude of this estimate indicates that a 0.1 change in the excess readmissions ratio, which is about a one-standard deviation change, is associated with a two percentage point (10%) decline in heart failure readmissions. This is the largest estimate in Table 7. It is also noteworthy that not all estimates are negative. Estimates pertaining to AMI are positive, but small and insignificant.

Table 8 presents the IV estimates of the effect of the HRRP penalty on readmissions. One advantage of the IV is it provides estimates of the effect of the size of the penalty on readmissions and not just whether a hospital was penalized. Prior studies have not provided such an estimate. As in the reduced form, we show estimates from periods before and after the implementation of the HRRP. We expect estimates from earlier periods to be statistically insignificant, which they are.

As our model is just identified, the IV estimates are similar to the reduced form estimates, but scaled by first stage estimate.[23] IV estimates measure the change in readmissions for a hospital

[23] Due to rounding in tables, the IV estimate is not exactly equal to the reduced form estimate in Table 7 divided by first stage estimate in Table 6.

that received the maximum penalty of 1% versus an unpenalized hospital. All IV estimates in Table 8 are small to moderate in magnitude and not statistically significant at 0.05 level. Consistent with reduced form estimates, IV estimates pertaining to AMI are positive. Estimates related to pneumonia are small and insignificant. Estimates of the effect of the round 1 penalty on heart failure readmissions are negative and the 2014 estimate is marginally significant. Consider the magnitude of this estimate. It implies that receiving the maximum HRRP penalty is associated with a 4.3 percentage point (21%) reduction in heart failure readmissions.

How does this estimate compare to the magnitudes of estimates from other approaches (DD, DDD, IV) reported earlier? Estimates for heart failure from these other approaches are in the range of -0.02 and -0.03, which is 25% to 50% smaller than the IV estimate for heart failure in Table 8. However, these other estimates refer to the effect of the average penalty and not the maximum penalty of 1%, as does the IV estimate. The average hospital penalty for heart failure was approximately 0.2% and using this figure to scale the IV estimate implies an effect size of -0.009 (or 0.9 percentage points). Thus, putting our IV estimate on a comparable basis to the previously reported estimates indicates that the IV estimate in Table 8 is less than half the size of the previous estimates and quite small suggesting a 4% reduction in heart failure readmissions for penalized hospitals. We acknowledge that the linear interpolation used to get comparable estimates may introduce substantial bias, but such a comparison highlights that the IV estimates in Table 8 are quite small (across all conditions) relative to estimates obtained by other approaches. Overall, we conclude that the IV estimates do not show much evidence that the HRRP penalty reduced readmissions for any of the conditions. At most, there is weak evidence that the HRRP penalty reduced readmissions for heart failure.

We have also validated the regression-kink design in a couple of ways. As noted, we find no evidence of effects in prior periods when the HRRP was not in effect (Landais 2015). We also find (see Appendix Table 3) that the addition of several baseline (i.e., 2007) hospital-specific covariates to the model has virtually no effect on reduced form or IV estimates.

Interpreting the Regression-kink (RK) Results

What may account for the null results we find? One answer is that hospitals do not respond to the penalty or respond ineffectively. Earlier, we provided several reasons why this may be the case: it is difficult and costly to identify preventable readmissions; there is, at best, only weak evidence that common interventions are effective, particularly with respect to readmissions after 7 days (the majority); and the penalty formula is flawed providing little incentive for hospitals to respond. Another explanation is that the RK obscures a true effect. We consider this issue next.

One way the RK may obscure a true effect is if some unpenalized hospitals close to being penalized respond to avert future, expected penalties. If so, then we expect that the true "kink" point is to the left of the threshold that is set using the excess readmissions threshold of one. The opposite behavior—some penalized hospitals close to the threshold, do not respond—would produce a "kink" to the right of the threshold that is set using the excess readmissions threshold of one. It is unlikely that both types of behavior are occurring simultaneously. If unpenalized hospitals just to the left of the threshold are responding to avert future penalties, then it suggests that penalized hospitals to the right of the threshold are responding similarly. And if penalized hospitals just to the right of the threshold are not responding, then it suggests that unpenalized hospitals to the left of the threshold are not responding. If whether a hospital was penalized or not, or the hospital lies in the excess readmissions distribution, does not provide a signal to the hospital about future penalties, as just described, then the penalty cannot bring forth a response (and is consistent with our null finding).

We assess the extent of this potential problem by testing for additional kink points to the left and right of the threshold. We use a randomization procedure that produces a distribution of placebo estimates in the region around the actual HRRP penalty threshold (Ganong and Jager 2018; Landais, 2015). The procedure is carried out by changing the excess readmission threshold a large number of times from above to below the actual HRRP penalty threshold (0.9 to 1.06), For each pseudo threshold, we re-estimate the regression kink model for each pseudo threshold to construct a distribution of estimates. We chose 0.9 (appx 10th percentile) to 1.1 (90th percentile) because we maintain a bandwidth of +/- 0.1 either side of each randomized threshold that we test and 100% of the hospitals have an Excess Readmissions Ratio between 0.8 and 1.2. Thus using 0.9 to 1.1, allows us to test whether the kink in hospital response to the HRRP, is located as early as the 10th percentile of the excess readmission ratio and up to the 90th percentile of the excess readmission ratio.

We plot the distribution of estimates and the distribution of the R-square statistic from each regression. If the true threshold was to the left (or right) of the actual threshold, we would expect to see this reflected in the R-square (a better fit).

Figures 3 through 5 show the randomized estimates for three years 2012, 2013 and 2014: Figure 3 shows estimates for AMI; Figure 4 shows estimates for heart failure; and Figure 5 shows estimates for pneumonia. The visual evidence in Figures 3 through 5 is compelling. For AMI and pneumonia, none of the estimates shown (Figures 3 and 5) are statistically significant or significantly different from the estimate obtained using the threshold of 1.0. All estimates are also very small and are within a very tight range. For heart failure, 2012 estimates are all statistically insignificant, virtually zero and within a narrow range. In 2013 and 2014, about one-third of the randomized estimates are statistically significant, but all remain small and are not statistically different from the estimate obtained using the threshold of 1.0. Overall, the randomized estimates in Figures 3 through 5 provide no evidence that there was a different threshold and that the insignificant results we report in Table 7 are due to using the wrong threshold. This point is further evidenced in Figure 6, which graphs the R-square statistics from the randomized regressions for each condition. There is no evidence that the R-square statistic increases at thresholds away from 1.0. Therefore, we conclude that the null findings in Tables 7 and 8 are not due to using the wrong threshold, but more likely because hospitals were either not responding to the HRRP penalty or not responding effectively.

Conclusion

The HRRP is one of the most significant and salient pay-for-performance programs used by the federal government to improve the quality of care provided to people covered by Medicare. It has received much attention for some of the imperfections associated with its implementation and for its consequences. In this article, we revisited the evidence on the consequences of the HRRP for the primary outcome targeted—hospital readmissions. Past studies have generally shown that the HRRP reduced hospital readmissions. However, our assessment of the validity of the empirical approaches used in past studies suggest that the research designs used are unlikely to be valid, which limits the usefulness of the evidence produced by those studies.

We also provide novel evidence of the effect of the HRRP on hospital readmissions that is based on a regression-kink design. We established the validity of this approach. In contrast to past research, our results suggest that the HRRP had no clinically or statistically significant effects on hospital readmissions. While our findings differ from prior research, they are consistent with implementation imperfections that characterize the HRRP (e.g., imperfect risk adjustment method), the known difficulty of identifying ex ante hospital readmissions that are preventable, and the absence of consistent evidence on interventions that prevent readmissions. Therefore, our findings are not surprising while the relatively large findings in past studies are.

References

(1) American Hospital Association (2015). Rethinking the Hospital Readmissions Reduction Program.

(2) Auerbach, A. D., Kripalani, S., Vasilevskis, E. E., Sehgal, N., Lindenauer, P. K., Metlay, J. P., ... & Schnipper, J. L. (2016). Preventability and causes of readmissions in a national cohort of general medicine patients. *JAMA internal medicine*, 176(4), 484-493.

(3) Bailey MK (IBM Watson Health), Weiss AJ (IBM Watson Health), Barrett ML (M.L. Barrett, Inc.), Jiang HJ (AHRQ). Characteristics of 30-Day Readmissions, 2010-2016. HCUP Statistical Brief #248. February 2019. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/reports/statbriefs/sb248-Hospital-Readmissions-2010-2016.pdf.

(4) Bartik, T. J. (1991). Who benefits from state and local economic development policies?. W.E. Upjohn Institute for Employment Research.

(5) Bowles, K. H., Hanlon, A. L., Glick, H. A., Naylor, M. D., O'Connor, M., Riegel, B., Shih, N., & Weiner, M. G. (2011). Clinical effectiveness, access to, and satisfaction with care using a telehomecare substitution intervention: a randomized controlled trial. *International Journal of Telemedicine and Applications*, 2011.

(6) Brewster, Amanda L., Cherlin, Emily J., Ndumele, Chima D.; Collins, Diane; Burgess, James F. Jr.,; Charns, Martin P.; Bradley, Elizabeth H.; Curry, Leslie A. What Works in Readmissions Reduction, Medical Care: June 2016 - Volume 54 - Issue 6 - p 600-607

(7) Card, D., Lee, D.S., Pei, Z. and Weber, A. (2015), Inference on Causal Effects in a Generalized Regression Kink Design. *Econometrica*, 83: 2453-2483.

(8) Carey, K., & Lin, M. Y. (2015). Readmissions to New York hospitals fell for three target conditions from 2008 to 2012, consistent with Medicare goals. *Health Affairs*, 34(6), 978-985.

(9) Chamberlain, R. S., Sond, J., Mahendraraj, K., Lau, C. S., & Siracuse, B. L. (2018). Determining 30-day readmission risk for heart failure patients: the readmission after heart failure scale. *International Journal of General Medicine*, 11, 127.

(10) Chen, M., & Grabowski, D. C. (2019). Hospital readmissions reduction program: intended and unintended effects. *Medical Care Research and Review*, 76(5), 643- 660.

(11) Chhabra, K. R., Ibrahim, A. M., Thumma, J. R., Ryan, A. M., & Dimick, J. B. (2019). Impact of Medicare readmissions penalties on targeted surgical conditions. *Health Affairs*, 38(7), 1207-1215.

(12) Chin, D. L., Bang, H., Manickam, R. N., & Romano, P. S. (2016). Rethinking thirty-day hospital readmissions: shorter intervals might be better indicators of quality of care. *Health Affairs (Project Hope)*, 35(10), 1867–1875.

(13) Cromwell, J., Trisolini, M. G., Pope, G. C., Mitchell, J. B., & Greenwald, L. M. (Eds.). (2011). *Pay for Performance in Health Care: Methods and Approaches*. Research Triangle Park, NC: RTI Press.

(14) Davis, K. K., Mintzer, M., Himmelfarb, C. R. D., Hayat, M. J., Rotman, S., & Allen, J. (2012). Targeted intervention improves knowledge but not self- care or readmissions in heart failure patients with mild cognitive impairment. *European Journal of Heart Failure*, 14(9), 1041-1049.

(15) Desai, N. R., Ross, J. S., Kwon, J. Y., Herrin, J., Dharmarajan, K., Bernheim, S. M., Krumholz, H. M., & Horwitz, L. I. (2016). Association between hospital penalty status under the hospital readmission reduction program and readmission rates for target and nontarget conditions. *JAMA*, 2016, 316(24), 2647-2656.

(16) Demiralp, B., He, F., & Koenig, L. (2018). Further evidence on the system-wide effects of the Hospital Readmissions Reduction Program. *Health Services Research*, 53(3), 1478-1497.

(17) Doran T., Maurer K., & Ryan A. M. (2016). Impact of provider incentives on quality and cost of health care. *Annual Review of Public Health*, 38, 449-465.

(18) Alimadadi, Elham., Abbasinia, M., Mohammadbeigi, A., & Abbasi, M. (2020). Risk factors of readmission after coronary artery bypass graft surgery: A case-control study. *Nursing Practice Today*, 7(4), 295-301.

(19) Fingar, K. R., Barrett, M. L., & Jiang, H. J. (2017). A comparison of all-cause 7-day and 30-day readmissions, 2014: statistical brief #230. Healthcare Cost and Utilization Project (HCUP).

(20) Finkelstein, A., Zhou, A., Taubman, S., & Doyle, J. (2020). Health care hotspotting -- a randomized, controlled trial. *New England Journal of Medicine*, 382(2), 152-162.

(21) Finn, K. M., Heffner, R., Chang, Y., Bazari, H., Hunt, D., Pickell, K., Berube, R., Raju, S., Farrell, E., Lyasere, C., Thompson, R., O'Malley, T., O'Donnell, W., & Karson, A. (2011). Improving the discharge process by embedding a discharge facilitator in a resident team. *Journal of Hospital Medicine*, 6(9), 494- 500.

(22) Ganong, P. & Jäger, S. (2018), "A Permutation Test for the Regression Kink Design", *Journal of the American Statistical Association*, 113:522, 494-504

(23) Goldsmith-Pinkham, P., Sorkin, I., & Swift, H. (2018). Bartik instruments: What, when, why, and how. (No. w24408). National Bureau of Economic Research.

(24) Graham, K. L., Auerbach, A. D., Schnipper, J. L., Flanders, S. A., Kim, C. S., Robinson, E. J., Ruhnke, G. W., Thomas, L. R., Kripalani, S., Vasilevskis, E. E., Fletcher, G. S., Sehgal, N. J., Lindenauer, P. K., Williams, M. V., Metlay, J. P., Davis, R. B., Yang, J., Marcantonio, E. R., & Herzig, S. J. (2018). Preventability of early versus late hospital readmissions in a national cohort of general medicine patients. *Annals of Internal Medicine*, 168(11), 766-774.

(25) Gupta, A., Allen, L. A., Bhatt, D. L., Cox, M., DeVore, A. D., Heidenreich, P. A., & Fonarow, G. C. (2017). Implementation of the hospital readmissions reduction program, readmissions, and mortality in heart failure. *Circulation*, 136(supp1-1), A15192-A15192.

(26) Gupta, A. (2021). Impacts of performance pay for hospitals: The readmissions reduction program. *American Economic Review*, 111(4), 1241-83.

(27) Ibrahim, A. M., Dimick, J. B., Sinha, S. S., Hollingsworth, J. M., Nuliyalu, U., & Ryan, A. M. (2018). Association of coded severity with readmission reduction after the hospital readmissions reduction program. *JAMA internal medicine*, 178(2), 290-292.

(28) Institute of Medicine. (2007). *Rewarding Provider Performance: Aligning Incentives in Medicare*. Washington, DC: The National Academies Press.

(29) David A. Jaeger & Theodore J. Joyce & Robert Kaestner, 2020. "A Cautionary Tale of Evaluating Identifying Assumptions: Did Reality TV Really Cause a Decline in Teenage Childbearing?," *Journal of Business & Economic Statistics*, vol 38(2), pages 317-326

(30) Joynt, K. E., Figueroa, J. F., Orav, E. J., & Jha, A. K. (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.

(31) Kansagara D., Englander, H., Salanitro, A., Kagen, D., Theobald, C., Freeman, M., & Kripalani, S. (2011). Risk prediction mo dels for hospital readmission: a systematic review. *JAMA*, 2011, 306(15): 1688–1698.

(32) Kripalani S, Theobald CN, Anctil B, Vasilevskis EE. Reducing hospital readmission rates: current strategies and future directions. *Annu Rev Med*. 2014;65:471-485. doi:10.1146/annurev-med-022613-090415

(33) Krumholz, H. M., Normand, S. T., Keenan, P. S., Desai, M. M., Lin, Z., Drye, E. E., Bhat, K. R., & Schreiner, G. C. (2008). Hospital 30-day acute myocardial infarction readmission measure: methodology. Report prepared for the Centers for Medicare & Medicaid Services.

(34) Landais, C. “Assessing the Welfare Effects of Unemployment Benefits Using the Regression Kink Design.” *American Economic Journal: Economic Policy*, 7(4):243–278 (2015).

(35) Lisk, C., & Stensland, J. (2018). Mandated Report: Effects of the Hospital Readmissions Reduction Program. In *Advising the Congress on Medicare Issues*.

(36) Mahmoudi, E., Kamdar, N., Kim, N., Gonzales, G., Singh, K., & Waljee, A. K. (2020). Use of electronic medical records in development and validation of risk prediction models of hospital readmission: systematic review. *BMJ*, 369: m958.

(37) Marusic, S., Gojo-Tomic, N., Erdeljic, V., Bacic-Vrca, V., Franic, M., Kirin, M., & Bozikov, V. (2013). The effect of pharmacotherapeutic counseling on readmissions and emergency department visits. *International Journal of Clinical Pharmacy*, 35(1), 37-44.

(38) McGarry, B. E., Blankley, A. A., & Li, Y. (2016). The impact of the Medicare hospital readmission reduction program in New York state. *Medical Care*, 54(2), 162-171.

(39) MedPAC. 2013. Report to the Congress: Medicare and the Health Care Delivery System; Chapter 4, Refining the Hospital Readmissions Reduction Program, pp. 91–110. Washington, DC: MedPAC. [accessed on May 28, 2014]. Available at http://www. medpac.gov/documents/reports/jun13_entirereport.pdf?sfvrsn=0

(40) Mellor, J., Daly, M., & Smith, M. (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.

(41) Mendelson, A., Kondo, K., Damberg, C., Low, A., Motuapuaka, M., Freeman, M., O'Neil, M., Relevo, R., & Kansagara, D. (2017). The effects of pay-for-performance programs on health, health care use, and processes of care: a systematic review. *Annals of Internal Medicine*, 166(5), 341-353.

(42) *NEJM Catalyst* (2018). Hospital Readmissions Reduction Program (HRRP). April 26.

(43) Ody, C., Msall, L., Dafny, L. S., Grabowski, D. C., & Cutler, D. M. (2019). Decreases in readmissions credited to Medicare's program to reduce hospital readmissions have been overstated. *Health Affairs*, 38(1), 36-43.

(44) Rau J. (2019). New round of Medicare readmission penalties hits 2583 hospitals. Kaiser Health News, Oct 1.

(45) Siracuse, B. L., & Chamberlain, R. S. (2016). A preoperative scale for determining surgical readmission risk after total hip replacement. *JAMA Surgery*, 151(8), 701- 709.

(46) Smith, L. N., Makam, A. N., Darden, D., Mayo, H., Das, S. R., Halm, E. A., & Nguyen, O. K. (2018). Acute myocardial infarction readmission risk prediction models: a systematic review of model performance. *Circulation: Cardiovascular Quality and Outcomes*, 11(1), e003885.

(47) Soumerai S., & Sullivan K. (2019). Ignoring evidence begot Medicare's dangerous hospital readmissions penalty. *STAT News*, May.

(48) Thompson, M. P., Waters, T. M., Kaplan, C. M., Cao, Y., & Bazzoli, G. J. (2017). Most hospitals received annual penalties for excess readmissions, but some fared better than others. *Health Affairs*, 36(5), 893-901.

(49) Van Walraven, C., Bennett, C., Jennings, A., Austin, P. C., & Forster, A. J. (2011). Proportion of hospital readmissions deemed avoidable: a systematic review. *CMAJ*, 183(7): E391-E402.

(50) Vest, J. R., Gamm, L. D., Oxford, B. A., Gonzalez, M. I., Slawson, K. M. (2010). Determinants of preventable readmissions in the United States: a systematic review. *Implement Sci*., 5, 88.

(51) Wasfy, J. H., Zigler, C. M., Choirat, C., Wang, Y., Dominici, F., & Yeh, R. W. (2017). Readmission rates after passage of the hospital readmissions reduction program: a pre–post analysis. *Annals of Internal Medicine*, 166(5), 324-331.

(52) Weinreich, M., Nguyen, O. K., Wang, D., Mayo, H., Mortensen, E. M., Halm, E. A., & Makam, A. N. (2016). Predicting the risk of readmission in pneumonia. a systematic review of model performance. *Annals of the American Thoracic Society*, 13(9), 1607-1614.

(53) Zingmond, D. S., Liang, L. J., Parikh, P., & Escarce, J. J. (2018). The impact of the Hospital Readmissions Reduction Program across insurance types in California. *Health services research*, 53(6), 4403-4415.

(54) Zuckerman, R. B., Sheingold, S. H., Orav, E. J., Ruhter, J., & Epstein, A. M. (2016). Readmissions, observation, and the hospital readmissions reduction program. *New England Journal of Medicine*, 374(16), 1543-1551.

Appendix A: Algebra Underlying Specification Test of DDD

The model is estimated on aggregate data at state (j) and year (t) level. For the treated group:

(1) $Y_{jt} = \alpha_t + \beta_j + \gamma\, Policy_{jt} + \lambda Z_{jt} + u_{jt} \qquad TREAT = 1$

Data on Z are unavailable (or only partly available). So observed model is:

(2) $Y_{jt} = \tilde{\alpha}_t + \tilde{\beta}_j + \tilde{\gamma}\, Policy_{jt} + \tilde{u}_{jt} \qquad TREAT = 1$

The coefficient on policy is biased by omitted variable.

$$
\begin{aligned}
&Y_{jt} = \tilde{\alpha}_t + \tilde{\beta}_t + \tilde{\gamma}\, Policy_{jt} + \tilde{u}_{jt} \qquad TREAT = 1 \\
\text{(2a)}\quad &\tilde{\gamma} = \gamma + \lambda\delta \\
&Policy_{jt} = \pi_j + r_t + \delta Z_{jt} + u_{jt}
\end{aligned}
$$

For the untreated group:

$$
\begin{aligned}
\text{(3)}\quad &Y_{jt} = \alpha'_t + \beta'_j + kPolicy_{jt} + \sigma Z_{jt} + \nu_{jt} \qquad TREAT = 0 \\
&k = 0
\end{aligned}
$$

Note that policy is included in this specification, but it has a 0 coefficient because policy has no effect on control group. Also note that I allow the effect of unobservable to be different in untreated and treated groups ($\sigma \neq \lambda$).

The observed model is:

$$
\begin{aligned}
&Y_{jt} = \alpha''_t + \beta''_j + k'Policy_{jt} + \nu_{jt} \qquad TREAT = 0 \\
\text{(3a)}\quad &k' = 0 + \sigma\delta \\
&Policy_{jt} = \pi_j + r_t + \delta Z_{jt} + u_{jt}
\end{aligned}
$$

Taking the difference (DDD) between treated (DD) and untreated (DD) groups and estimating the model yields:

$$
\begin{aligned}
&Y_{jt}(T=1) - Y_{jt}(T=0) = (\tilde{\alpha}_t - \alpha''_t) + (\tilde{\beta}_j - \beta''_j) + (\tilde{\gamma} - k')Policy_{jt} + \varpi_{jt} \\
\text{(4)}\quad &\tilde{\gamma} - k' = \gamma + (\lambda - \sigma)\delta \\
&\tilde{\gamma} - k' = \gamma \qquad if \quad (\lambda = \sigma)
\end{aligned}
$$

This is well known result that the DDD estimate is correct if the unobservable variable has same effect on Y for treated and untreated groups.

If we go back to first-difference specification, but use the true model, then we have:

(6) $Y_{jt}(T=1) - Y_{jt}(T=0) = (\tilde{\alpha}_t - \alpha'_t) + (\tilde{\beta}_j - \beta'_j) + \gamma Policy_{jt} + (\lambda - \sigma)Z_{jt} + \varpi_{jt}$

And if we rewrite (6) moving Y(T=0) to the right-hand side, then we have:

$$\text{(7)}\quad \begin{aligned} &Y_{jt}(T=1) = \rho Y_{jt}(T=0) + (\tilde{\alpha}_t - \alpha'_t) + (\tilde{\beta}_j - \beta'_j) + \gamma Policy_{jt} + (\lambda - \sigma) Z_{jt} + \varpi_{jt} \\ &\rho = 1 \end{aligned}$$

The observed version of this model is:

$$\begin{aligned} &Y_{jt}(T=1) = \ddot{\rho} Y_{jt}(T=0) + \pi_t + \varphi_j + \ddot{\gamma} Policy_{jt} + \varpi_{jt} \\ &Policy_{jt} = \pi_j + r_t + \delta Z_{jt} + cY_{jt}(T=0) + u_{jt} \\ &Policy_{jt} = \pi_j + r_t + \delta Z_{jt} + u_{jt}, \qquad c = 0 \\ \text{(8)}\quad &\ddot{\gamma} = \gamma + (\lambda - \sigma)\delta \\ &Y_{jt}(T=0) = \pi_j + r_t + \sigma Z_{jt} + kPolicy_{jt} + u_{jt} \\ &Y_{jt}(T=0) = \pi_j + r_t + \sigma Z_{jt} + u_{jt}, \quad k = 0, \quad \text{(see equation 3)} \\ &\ddot{\rho} = 1 + (\lambda - \sigma)\sigma \end{aligned}$$

The main result from (8) is that if the coefficient ($\ddot{\rho}$) on Y(T=0) is not equal to 1, then we reject the underlying assumption of the DDD model.

One limitation of this test of validity of the DDD model is that the bias term $(\lambda - \sigma)$ is scaled by σ, which is the effect of the unmeasured confounder on the outcome for the untreated group (*TREAT=0*). Consider an extreme case: $\sigma = 0$ and $(\lambda \neq \sigma)$. The DDD estimate is biased ($\text{bias is } \lambda\delta$), but the coefficient on $\ddot{\rho} = 1 + (\lambda - \sigma)\sigma$ is equal to one, which suggests no bias. Thus, rejecting ($\ddot{\rho} = 1$) is evidence that DDD is not valid, but failure to reject may not indicate that it is valid.

Table 1. Difference-in-differences (DD) Estimates of the Effect of HRRP Penalty on Readmissions, by HRRP Condition

	Condition-specific Penalty			Hospital-wide Penalty				
	AMI	HF	PN	AMI	HF	PN	All Targeted	All Non-Targeted
Penalty	-0.017***	-0.018***	-0.017***	-0.008***	-0.009***	-0.008***	-0.010***	-0.003***
	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.000)
Event-study Estimates								
Penalty X 2008	-0.009**	-0.004*	-0.015***	0.001	-0.001	-0.011***	-0.003*	-0.000
	(0.003)	(0.002)	(0.002)	(0.004)	(0.002)	(0.002)	(0.002)	(0.001)
Penalty X 2009	0.008**	0.007***	0.005*	0.007*	0.003	-0.003	0.002	0.001
	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)
Penalty X 2010	0.006	0.008***	0.003	0.009*	0.004*	0.000	0.003*	0.001
	(0.004)	(0.002)	(0.002)	(0.004)	(0.002)	(0.002)	(0.002)	(0.000)
Penalty X 2012	-0.014***	-0.013***	-0.018***	-0.003	-0.005*	-0.011***	-0.007***	-0.001*
	(0.003)	(0.002)	(0.002)	(0.004)	(0.002)	(0.002)	(0.001)	(0.000)
Penalty X 2013	-0.016***	-0.016***	-0.020***	-0.007	-0.009***	-0.012***	-0.010***	-0.003***
	(0.004)	(0.002)	(0.002)	(0.004)	(0.002)	(0.002)	(0.002)	(0.001)
Penalty X 2014	-0.016***	-0.019***	-0.020***	-0.003	-0.010***	-0.013***	-0.011***	-0.004***
	(0.004)	(0.002)	(0.002)	(0.004)	(0.002)	(0.002)	(0.002)	(0.001)
Mean Dep. Variable	0.182	0.215	0.162	0.182	0.215	0.162	0.195	0.181
Number of Obs.	15203	21401	21508	19147	22105	22245	22279	22279

Notes: The unit of observation is the hospital-year level and the dependent variable is 30-day readmissions. The regressions include patients admitted for each relevant condition between July 2008 and July 2014. AMI=acute myocardial infarction; HF=heart failure; and PN=pneumonia. Columns 1-3 present DD and event study estimates using an indicator for being penalized for the specific condition. Columns 4-8 present DD and event study estimates using an indicator of a hospital-wide instead of the condition-specific penalty. All regressions include hospital fixed effects, year fixed effects and a set of controls for patient characteristics that vary over time for each hospital (indicators for age under 65, age 65-70, age 70-75, age 80-85, age over 85, share Hispanic, share Black, share White, share other race, share female). Robust standard errors are constructed allowing for non-independence at the hospital level. Each regression is weighted by the number of patients admitted for the relevant condition (AMI, HF, PN) in the hospital year. * 0.05 < p-value<=0.10, ** p <= 0.05, *** p <= 0.01

Table 2. Difference-in-differences-in-differences (DDD) Estimates of the Effect of HRRP Penalty on Readmissions, by Condition

	AMI			HF			PN		
Condition-specific Penalty	-0.013*** (0.002)	-0.013*** (0.002)	-0.017*** (0.002)	-0.014*** (0.001)	-0.014*** (0.001)	-0.018*** (0.001)	-0.013*** (0.001)	-0.014** (0.001)	-0.017*** (0.001)
Readmissions Non-targeted		1.052*** (0.055)			1.025*** (0.029)			0.824*** (0.026)	
Readmissions GI			0.032* (0.013)			0.031*** (0.006)			0.021*** (0.006)
Mean Dep. Variable	0.182	0.182	0.182	0.215	0.215	0.215	0.162	0.162	0.162
Number of Obs.	15203	15203	15203	21401	21401	21401	21508	21508	21508
Hospital-wide Penalty	-0.005*** (0.002)	-0.005*** (0.002)	-0.008*** (0.002)	-0.006*** (0.001)	-0.006*** (0.001)	-0.009*** (0.001)	-0.005*** (0.001)	-0.005*** (0.001)	-0.008*** (0.001)
Readmissions Non-targeted		1.068*** (0.053)			1.054*** (0.029)			0.844*** (0.026)	
Readmissions GI			0.036** (0.012)			0.031*** (0.006)			0.022*** (0.006)
Mean Dep. Variable	0.182	0.182	0.182	0.215	0.215	0.215	0.162	0.162	0.162
Number of Obs.	19147	19147	19147	22105	22105	22105	22245	22245	22245

Notes: The unit of observation is the hospital-year level. In the first column of results for each condition, the dependent variable is the difference in readmissions between the condition indicated and readmissions for all non-targeted conditions. In the other two columns, the dependent variable is 30-day readmissions for that condition. The regressions include patients admitted for each relevant condition between July 2008 and July 2014. AMI=acute myocardial infarction; HF=heart failure; and PN=pneumonia. All regressions include hospital fixed effects, year fixed effects and a set of controls for patient characteristics that vary over time for each hospital (indicators for age under 65, age 65-70, age 70-75, age 80-85, age over 85, share Hispanic, share Black, share White, share other race, share female). Robust standard errors are constructed allowing for non-independence at the hospital level. Each regression is weighted by the number of patients admitted for the relevant condition (AMI, HF, PN) in the hospital year. * 0.05 < p-value<=0.10, ** p <= 0.05, *** p <= 0.01

Table 3. Difference-in-differences-in-differences (DDD) Estimates of the Effect of HRRP Penalty on Readmissions, by Condition
Event Study Design

	Condition-specific Penalty			Hospital Wide Penalty			
	AMI	HF	PN	AMI	HF	PN	All Targeted
Penalty X 2008	-0.010***	-0.004**	-0.015***	0.0009	-0.0002	-0.010***	-0.003**
	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.001)
Penalty X 2009	0.006**	0.005***	0.003*	0.006*	0.002	-0.004***	0.0005
	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.001)
Penalty X 2010	0.004	0.007***	0.002	0.008**	0.002	-0.001***	0.002
	(0.003)	(0.002)	(0.002)	(0.004)	(0.002)	(0.002)	(0.001)
Penalty X 2012	-0.013***	-0.010***	-0.015***	-0.002	-0.003*	-0.010***	-0.006***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Penalty X 2013	-0.014***	-0.011***	-0.016***	-0.004	-0.006***	-0.008***	-0.007***
	(0.003)	(0.002)	(0.002)	(0.004)	(0.002)	(0.002)	(0.001)
Penalty X 2014	-0.013***	-0.014***	-0.015***	0.0004	-0.005***	-0.008***	-0.006***
	(0.003)	(0.002)	(0.002)	(0.004)	(0.002)	(0.002)	(0.001)
Mean Dependent Variable	0.182	0.215	0.162	0.182	0.215	0.162	0.195
Number of Observations	15203	21401	21508	19147	22105	22245	22279

Notes: The unit of observation is the hospital-year level dependent variable is 30-day readmissions. In all regressions, the control conditions are all non-targeted readmissions. The regressions include patients admitted for each relevant condition between July 2008 and July 2014. AMI=acute myocardial infarction; HF=heart failure; and PN=pneumonia. All regressions include hospital fixed effects, year fixed effects and a set of controls for patient characteristics that vary over time for each hospital (indicators for age under 65, age 65-70, age 70-75, age 80-85, age over 85, share Hispanic, share Black, share White, share other race, share female). Robust standard errors are constructed allowing for non-independence at the hospital level. Each regression is weighted by the number of patients admitted for the relevant condition (AMI, HF, PN, Targeted Readmissions) in the hospital year. * 0.05 < p-value<=0.10, ** p <= 0.05, *** p <= 0.01

Table 4. OLS, Just Identified IV and Event-study IV Estimates of the Effect of HRRP Condition-specific Penalty on Condition Specific Readmissions 2008-2014

	Actual Penalty			Expected Penalty		
OLS (Difference-in-difference)	AMI	HF	PN	AMI	HF	PN
Condition-specific Penalty	-0.017***	-0.018***	-0.017***	-0.024***	-0.034***	-0.039***
	(0.002)	(0.0012)	(0 .001)	(0.002)	(0.001)	(0.001)
IV-- Just Identified Model						
Condition-specific Penalty	-0.057***	-0.029***	-0.024***	-0.041***	-0.023***	-0.030***
	(0.011)	(0.0037)	(0.003)	(0.006)	(0.003)	(0.004)
IV—Event-study IV						
Condition-specific Penalty X 2008	0.036***	0.041***	0.007	0.060***	0.043***	0.008
	(0.013)	(0.006)	(0.005)	(0.020)	(0.006)	(0.007)
Condition-specific Penalty X 2009	0.037***	0.016***	0.006	0.060***	0.017***	0.008
	(0.012)	(0.005)	(0.005)	(0.018)	(0.005)	(0.007)
Condition-specific Penalty X 2010	0.007	0.012**	0.003	0.015***	0.012***	0.005
	(0.012)	(0.005)	(0.005)	(0.021)	(0.006)	(0.008)
Condition-specific Penalty X 2012	-0.023*	0.002	-0.016***	-0.037	0.001	-0.020***
	(0.013)	(0.005)	(0.005)	(0.021)	(0.005)	(0.007)
Condition-specific Penalty X 2013	-0.034***	-0.010**	-0.021***	-0.056***	-0.012***	-0.026***
	(0.013)	(0.005)	(0.006)	(0.022)	(0.005)	(0.008)
Condition-specific Penalty X 2014	-0.038***	-0.025***	-0.022***	-0.053***	-0.027***	-0.028***
	(0.012)	(0.005)	(0.006)	(0.020)	(0.006)	(0.007)
Mean Dependent Variable	0.182	0.215	0.162	0.182	0.215	0.162
Number of Observations	15203	21099	21706	15203	21099	21706

Notes: The unit of observation is the hospital-year level. The regressions include patients admitted for each relevant condition between July 2008 and July 2014. AMI=acute myocardial infarction; HF=heart failure; and PN=pneumonia. All regressions include hospital fixed effects, year fixed effects and a set of controls for patient characteristics that vary over time for each hospital (indicators for age under 65, age 65-70, age 70-75, age 80-85, age over 85, share Hispanic, share Black, share White, share other race, share female). The instrument(s) for IV estimates is (are) the interaction(s) between 2006-07 condition-specific readmissions and indicator(s) for post-HRRP period (year). Robust standard errors are constructed allowing for non-independence at the hospital level. Each regression is weighted by the number of patients admitted for the relevant condition (AMI, HF, PN) in the hospital year. * 0.05 < p-value<=0.10, ** p <= 0.05, *** p <= 0.01

Table 5. Just Identified and Over Identified IV Estimates of the Effect of HRRP of HRRP Condition-specific Penalty on Condition Specific Readmissions 2008-2014

	Actual Penalty			Expected Penalty		
IV-- Just Identified Model	AMI	HF	PN	AMI	HF	PN
Condition-specific Penalty	-0.057***	-0.029***	-0.024***	-0.041***	-0.023***	-0.030***
	(0.011)	(0.0037)	(0.003)	(0.006)	(0.003)	(0.004)
IV—Over Identified Model						
Condition-specific Penalty	-0.060***	-0.032***	-0.025***	-0.108***	-0.063***	-0.080***
	(0.013)	(0.005)	(0.004)	(0.018)	(0.009)	(0.013)
Share Black 2006-2007 X Post	0.027***	0.006	0.006	0.017+	0.004	0.002
	(0.014)	(0.004)	(0.006)	(0.010	(0.004)	(0.006)
Share Hispanic 2006-2007 X Post	0.024	0.005	-0.002	0.033	-0.005	-0.011
	(0.026)	(0.020)	(0.011)	(0.024)	(0.016)	(0.013)
Share Dual Eligible 2006-2007 X Post	0.016	0.009	0.008	0.012	0.015**	0.024***
	(0.015)	(0.007)	(0.005)	(0.012)	(0.007)	(0.007)
Medium Group Diabetes 2006-2007 X Post	-0.002	0.004***	0.002	-0.002	0.003***	<0.000
	(0.003)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)
High Group Diabetes 2006-2007 X Post	<0.000	0.006	0.001	-0.002	0.005***	0.001
	(0.004)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)
Medium Group Pneumonia 2006-2007 X Post	<0.000	<0.000	<0.000	0.0004	<0.000	<0.000
	(0.003)	(0.001)	(0.001)	(0.002)	(0.014)	(0.001)
High Group Pneumonia 2006-2007 X Post	-0.003	<0.000	<0.000	-0.0008	0.002	<0.000
	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)
Teaching 2006-2007 X Post	0.001	-0.001	-0.002	0.0006	-0.0005	-0.002
	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Medium Group Bed Size 2006-2007 X Post	0.013***	0.002	-0.002	0.010**	-0.001	-0.006***
	(0.005)	(0.002)	(0.002)	(0.004)	(0.002)	(0.001)
High Group Bed Size 2006-2007 X Post	0.016***	0.005**	-0.004**	0.006	-0.003	-0.013***
	(0.005)	(0.002)	(0.002)	(0.004)	(0.002)	(0.002)
Mean Dependent Variable	0.182	0.215	0.162	0.182	0.215	0.162
Number of Observations	15203	21099	21706	15203	21099	21706

Notes: The unit of observation is the hospital-year level. The regressions include patients admitted for each relevant condition between July 2008 and July 2014. AMI=acute myocardial infarction; HF=heart failure; and PN=pneumonia. All regressions include hospital fixed effects, year fixed effects and a set of controls for patient characteristics that vary over time for each hospital (indicators for age under 65, age 65-70, age 70-75, age 80-85, age over 85, share Hispanic, share Black, share White, share other race, share female). The instrument(s) for IV estimates is (are) the interaction(s) between 2006-07 condition-specific readmissions and indicator(s) for post-HRRP period (year). Also added to model are interactions of baseline 2006-2007 characteristics (share Black, share Hispanic, share dual eligible, share with diabetes comorbidity, share with pneumonia comorbidity, teaching status, and bed count) and post-HRRP indicator. Robust standard errors are constructed allowing for non-independence at the hospital level. Each regression is weighted by the number of patients admitted for the relevant condition (AMI, HF, PN) in the hospital year. * 0.05 < p-value<=0.10, ** p <= 0.05, *** p <= 0.01

Table 6. Estimates of the Effect of the HRRP Excess Readmissions Ratio on the Round 1 HRRP Penalty

	AMI	HF	PN
Excess Readmissions Ratio 2009-11	<0.000***	<0.000***	<0.000***
	(<0.000)	(<0.000)	(<0.000)
Excess Readmissions Ratio*Penalty Threshold	0.031***	0.044***	0.035***
	(0.000)	(0.002)	(0.003)
Number of Observations	640	708	720
Mean Dependent Variable	0.0005	0.046	0.040

Notes – Each column represents a separate regression. For each condition (AMI, Heart Failure, Pneumonia), the sample of hospitals differs and includes hospitals that were penalized because of the condition (e.g., AMI) and hospitals who were not penalized for any condition. Hospitals with less than 50 cases throughout the three-year performance period (June 2008-July 2011) are excluded. Coefficient estimates show the effect of a 0.01 change in the excess readmission ratio and the effect is allowed to differ before and after the HRRP penalty threshold.
* 0.05 < p-value<=0.10, ** p <= 0.05, *** p <= 0.01

Table 7. Reduced Form Regression Kink Estimates of the Effect of the HRRP on Readmissions Pre- and Post-HRRP Round 1 Penalty

Pre-HRRP 2010	AMI	HF	PN
Excess Readmissions Ratio 2009-11	0.002***	0.004***	0.003***
	(0.0005)	(0.0002)	(0.0002)
Excess Readmissions Ratio*Penalty Threshold	0.002	0.001	-0.0004
	(0.001)	(0.001)	(0.0006)
Pre-HRRP 2011			
Excess Readmissions Ratio 2009-11	0.002***	0.004***	0.003***
	(0.0006)	(0.0002)	(0.0003)
Excess Readmissions Ratio*Penalty Threshold	0.002	-0.001	0.0002
	(0.002)	(0.001)	(0.0006)
Post-HRRP 2012			
Excess Readmissions Ratio 2009-11	0.0003	0.002***	0.001***
	(0.0006)	(0.0003)	(0.0003)
Excess Readmissions Ratio*Penalty Threshold	0.001	-0.002	-0.0004
	(0.002)	(0.001)	(0.0007)
Post-HRRP 2013			
Excess Readmissions Ratio 2009-11	0.0009	0.001***	0.002***
	(0.0005)	(0.0003)	(0.0003)
Excess Readmissions Ratio*Penalty Threshold	0.0003	-0.001	-0.001
	(0.001)	(0.001)	(0.0007)
Post-HRRP 2014			
Excess Readmissions Ratio 2009-11	0.0006	0.001***	0.002***
	(0.0006)	(0.0003)	(0.0003)
Excess Readmissions Ratio*Penalty Threshold	0.0006	-0.002*	-0.001
	(0.001)	(0.001)	(0.001)
Mean Dependent Variable	0.178	0.202	0.145
Number of Observations	640	708	720

Notes – Each column represents a separate regression. For each condition (AMI, Heart Failure, Pneumonia), the sample of hospitals differs and includes hospitals that were penalized because of the condition (e.g., AMI) and hospitals who were not penalized at all. Hospitals with less than 50 cases throughout the three - year performance period (June 2008-July 2011) are excluded. Coefficient estimates show the effect of a 0.01 change in the excess readmission ratio and the effect is allowed to differ before and after the HRRP penalty threshold.
* 0.05 < p-value<=0.10, ** p <= 0.05, *** p <= 0.01

Table 8. IV Estimates of the Effect of the HRRP Round 1 Penalty on Readmissions

Pre-HRRP 2010	AMI	HF	PN
Penalty	0.051	0.017	-0.012
	(0.044)	(0.021)	(0.017)
Pre-HRRP 2011			
Penalty	0.055	-0.023	0.003
	(0.051)	(0.017)	(0.020)
Post-HRRP 2012			
Penalty	0.064	-0.035	-0.011
	(0.050)	(0.024)	(0.020)
Post-HRRP 2013			
Penalty	0.012	-0.028	-0.029
	(0.046)	(0.022)	(0.021)
Post-HRRP 2014			
Penalty	0.024	-0.043*	-0.010
	(0.046)	(0.023)	(0.022)
Mean Dependent Variable	0.178	0.202	0.145
Number of Observations	640	708	720

Notes – Each column represents a separate regression. For each condition (AMI, Heart Failure, Pneumonia), the sample of hospitals differs and includes hospitals that were penalized because of the condition (e.g., AMI) and hospitals who were not penalized at all. Hospitals with less than 50 cases throughout the three - year performance period (June 2008-July 2011) are excluded. The instrument for the Round 1 Penalty is the interaction between the excess readmission ratio and indicator for being above the HRRP penalty threshold (see Table 6). Coefficient estimates show the effect of a 1% (max) in the HRRP penalty.
* 0.05 < p-value<=0.10, ** p <= 0.05, *** p <= 0.01

Appendix Table 1. OLS, Just Identified IV and Event-study IV Estimates of the Effect of HRRP Hospital-wide Penalty on Condition Specific Readmissions 2008-2014

	Actual Penalty					Expected Penalty				
OLS (Difference-in-difference)	AMI	HF	PN	All Targeted	Non-Targeted	AMI	HF	PN	All Targeted	Non-Targeted
Hospital-wide Penalty	-0.008***	-0.009***	-0.008***	-0.009***	-0.003***	-0.034***	-0.028***	-0.010***	-0.026***	-0.031***
	(0.002)	(0.0014)	(0.0012)	(0.001)	(0.0006)	(0.005)	(0.004)	(0.004)	(0.003)	(0.002)
IV-- Just Identified Model										
Hospital-wide Penalty	-0.045***	-0.039***	-0.013***	-0.031***	-0.037***	-0.087***	-0.038***	-0.064***	-0.070***	-0.041***
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.014)	(0.005)	(0.010)	(0.007)	(0.042)
IV—Event-study IV										
Hospital-wide Penalty X 2008	0.044***	0.041***	0.009	0.039***	0.033***	0.058***	0.096***	0.018	0.075***	.036***
	(0.017)	(0.006)	(0.007)	(0.005)	(0.003)	(0.019)	(0.014)	(0.015)	(0.008)	(0.003)
Hospital-wide Penalty X 2009	0.045***	0.016***	0.01	0.020***	0.021***	0.058***	0.037***	0.018	0.039***	0.024***
	(0.015)	(0.005)	(0.008)	(0.004)	(0.022)	(0.017)	(0.0122)	(0.016)	(0.008)	(0.002)
Hospital-wide Penalty X 2010	0.010	0.011***	0.005	0.012***	0.008***	-.013	0.025***	0.009	0.021***	0.009***
	(0.015)	(0.005)	(0.008)	(0.004)	(0.002)	(0.020)	(0.012)	(0.015)	(0.008)	(0.002)
Hospital-wide Penalty X 2012	-0.029+	0.001	-0.024***	-0.017***	-0.008***	-0.036*	0.002	-0.044***	-0.032***	-0.009***
	(0.017)	(0.005)	(0.008)	(0.004)	(0.002)	(0.020)	(0.012)	(0.015)	(0.008)	(0.002)
Hospital-wide Penalty X 2013	-0.041***	-0.010***	-0.030***	-0.021***	-0.026***	-0.053***	-0.026***	-0.057***	-0.041***	-0.029***
	(0.018)	(0.005)	(0.009)	(0.005)	(0.003)	(0.021)	(0.012)	(0.016)	(0.008)	(0.003)
Hospital-wide Penalty X 2014	-0.043***	-0.025***	-0.032***	-0.032***	-0.031***	-0.050***	-0.061***	-0.060***	-0.063***	-0.035***
	(0.017)	(0.005)	(0.009)	(0.005)	(0.004)	(0.020)	(0.013)	(0.017)	(0.009)	(0.003)
Mean Dependent Variable	0.182	0.215	0.162	0.195	0.181	0.182	0.215	0.162	0.195	0.181
Number of Observations	15203	21099	21706	22279	22279	19150	21768	21731	22279	22279

Notes: The unit of observation is the hospital-year level. The regressions include patients admitted for each relevant condition between July 2008 and July 2014. AMI=acute myocardial infarction; HF=heart failure; and PN=pneumonia. All regressions include hospital fixed effects, year fixed effects and a set of controls for patient characteristics that vary over time for each hospital (indicators for age under 65, age 65-70, age 70-75, age 80-85, age over 85, share Hispanic, share Black, share White, share other race, share female). The instrument(s) for IV estimates is (are) the interaction(s) between 2006-07 hospital-wide readmissions and indicator(s) for post-HRRP period (year). Robust standard errors are constructed allowing for non-independence at the hospital level. Each regression is weighted by the number of patients admitted for the relevant condition (AMI, HF, PN) in the hospital year. * 0.05 < p-value<=0.10, ** p <= 0.05, *** p <= 0.01

Appendix Table 2. Just Identified and Over Identified IV Estimates of the Effect of HRRP of HRRP Hospital-wide Penalty on Condition Specific Readmissions 2008-2014

	Actual Penalty					Expected Penalty				
IV-- Just Identified Model	AMI	HF	PN	All Targeted	Non-Targeted	AMI	HF	PN	All Targeted	Non-Targeted
Hospital-wide Penalty	-0.045***	-0.039***	-0.013***	-0.031***	-0.037***	-0.087***	-0.038***	-0.064***	-0.070***	-0.041***
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.014)	(0.005)	(0.010)	(0.007)	(0.042)
IV—Over Identified Model										
Hospital-wide Penalty	-0.073***	-0.048***	-0.037***	-0.048***	-0.029***	-0.038***	-0.043***	-0.021***	-0.051***	-0.031***
	(0.020)	(0.008)	(0.007)	(0.006)	(0.003)	(0.007)	(0.006)	(0.006)	(0.004)	(0.002)
Share Black 2006-2007 X Post	0.013	-0.003	-0.003	-0.004	0.0001	-0.001	-0.0007	-0.003	0.006**	0.006***
	(0.012)	(0.005)	(0.006)	(0.005)	(0.003)	(0.01)	(0.004)	(0.006)	(0.003)	(0.002)
Share Hispanic 2006-2007 X Post	0.005	-0.007	-0.001	-0.012	-0.001	0.023	0.005	0.0003	<0.000	0.009
	(0.03)	(0.020)	(0.013)	(0.013)	(0.008)	(0.024)	(0.015)	(0.012)	(0.01)	(0.006)
Share Dual Eligible 2006-2007 X Post	0.026	0.020***	0.021***	0.025***	0.009**	-0.005	0.003	0.007	0.012***	0.0002
	(0.020)	(0.010)	(0.007)	(0.007)	(0.004)	(0.011)	(0.006)	(0.008)	(0.004)	(0.002)
Medium Group Diabetes 2006-2007 X Post	<0.000	0.005***	0.002	0.0015	0.002+	-0.006***	0.004***	0.0003	0.001	0.001**
	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.0006)
High Group Diabetes 2006-2007 X Post	<0.000	0.007***	0.002	0.003	0.002**	-0.003	0.004***	<0.000	0.002	0.002***
	(0.004)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.0008)
Medium Group Pneumonia 2006-2007 X Post	<0.000	<0.000	<0.000	<0.000	-0.001	-0.002	-0.0009	-0.002	<0.000	-0.001***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.0006)
High Group Pneumonia 2006-2007 X Post	<0.000	<0.000	-0.002	<0.000	-0.002+	-0.003	0.0002	-0.002	<0.000	-0.002***
	(0.004)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.016)	(0.001)	(0.0007)
Teaching 2006-2007 X Post	0.001	<0.000	-0.001	<0.000	<0.000	-0.001	-0.0007	-0.002+	-0.001	-0.001
	(0.003)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.0006)
Medium Group Bed Size 2006-2007 X Post	0.012**	0.007***	<0.000	0.002	0.004***	-0.0002	-0.001	-0.008	<0.000	0.003***
	(0.005)	(0.002)	(0.002)	(0.002)	(0.001)	(0.003)	(0.002)	(0.001)	(0.001)	(0.0007)
High Group Bed Size 2006-2007 X Post	0.0131**	0.007***	-0.003	0.001	0.005***	0.003	-0.001	-0.003	<0.000	0.004***
	(0.005)	(0.003)	(0.002)	(0.002)	(0.001)	(0.003)	(0.002)	(0.006)	(0.001)	(0.0008)
Mean Dependent Variable	0.182	0.215	0.162	0.195	0.181	0.182	0.215	0.162	0.195	0.181
Number of Observations	19150	21768	21731	22279	22279	19150	21768	21731	22279	22279

Notes: The unit of observation is the hospital-year level. The regressions include patients admitted for each relevant condition between July 2008 and July 2014. AMI=acute myocardial infarction; HF=heart failure; and PN=pneumonia. All regressions include hospital fixed effects, year fixed effects and a set of controls for patient characteristics that vary over time for each hospital (indicators for age under 65, age 65-70, age 70-75, age 80-85, age over 85, share Hispanic, share Black, share White, share other race, share female). The instrument(s) for IV estimates is (are) the interaction(s) between 2006-07 hospital-wide readmissions and indicator(s) for post-HRRP period (year). Also added to model are interactions of baseline 2006-2007 characteristics (share Black, share Hispanic, share dual eligible, share with diabetes comorbidity, share with pneumonia comorbidity, teaching status, and bed count) and post-HRRP indicator. Robust standard errors are constructed allowing for non-independence at the hospital level. Each regression is weighted by the number of patients admitted for the relevant condition (AMI, HF, PN) in the hospital year. * 0.05 < p-value<=0.10, ** p <= 0.05, *** p <= 0.01

Appendix Table 3. Reduced Form Regression Kink Estimates of the Effect of the HRRP on Readmissions
Pre- and Post-HRRP Round 1 Penalty with and without Additional Controls

Pre-HRRP 2010	AMI		HF		PN	
Excess Readmissions Ratio 2009-11	0.002***	0.002***	0.004***	0.004***	0.003***	0.003***
	(0.0005)	(0.0006)	(0.0002)	(0.0003)	(0.0002)	(0.0003)
Excess Readmissions Ratio*Penalty Threshold	0.002	0.002	0.001	0.001	-0.0004	-0.0004
	(0.001)	(0.002)	(0.001)	(0.001)	(0.0006)	(0.0007)
Pre-HRRP 2011						
Excess Readmissions Ratio 2009-11	0.002***	0.002***	0.004***	0.004***	0.003***	0.003***
	(0.0006)	(0.0007)	(0.0002)	(0.0003)	(0.0003)	(0.0003)
Excess Readmissions Ratio*Penalty Threshold	0.002	0.002	-0.001	-0.001	0.0002	0.0002
	(0.002)	(0.002)	(0.001)	(0.001)	(0.0006)	(0.0006)
Post-HRRP 2012						
Excess Readmissions Ratio 2009-11	0.0003	0.0005	0.002***	0.002***	0.001***	0.001***
	(0.0006)	(0.0007)	(0.0003)	(0.0004)	(0.0003)	(0.0003)
Excess Readmissions Ratio*Penalty Threshold	0.001	0.001	-0.002	-0.002	-0.0004	-0.0002
	(0.002)	(0.002)	(0.001)	(0.001)	(0.0007)	(0.0007)
Post-HRRP 2013						
Excess Readmissions Ratio 2009-11	0.0009	0.001	0.001***	0.001***	0.002***	0.002***
	(0.0005)	(0.0006)	(0.0003)	(0.004)	(0.0003)	(0.0003)
Excess Readmissions Ratio*Penalty Threshold	0.0003	-0.0004	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.002)	(0.001)	(0.001)	(0.0007)	(0.0007)
Post-HRRP 2014						
Excess Readmissions Ratio 2009-11	0.0006	0.001	0.001***	0.001***	0.002***	0.001***
	(0.0006)	(0.0008)	(0.0003)	(0.0004)	(0.0003)	(0.0004)
Excess Readmissions Ratio*Penalty Threshold	0.0006	-0.0005	-0.002*	-0.002*	-0.001	-0.003
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Additional Controls	No	Yes	No	Yes	No	Yes
Mean Dependent Variable	0.178	0.178	0.202	0.202	0.145	0.145
Number of Observations	640	640	708	708	720	720

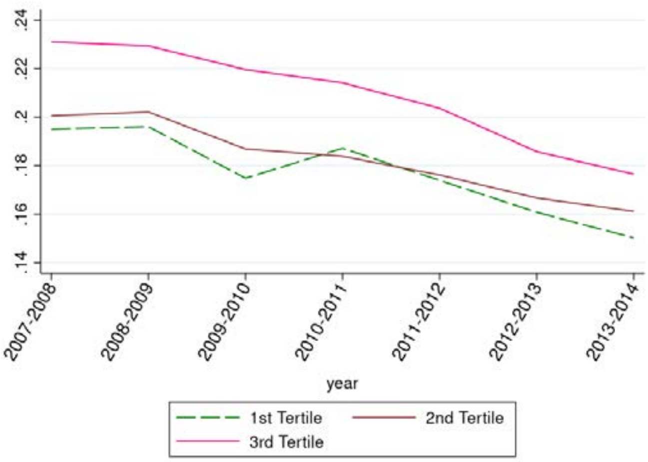
Notes – Each column represents a separate regression. For each condition (AMI, Heart Failure, Pneumonia), the sample of hospitals differs and includes hospitals that were penalized because of the condition (e.g., AMI) and hospitals who were not penalized at all. For each condition we report estimates from a model with a modest level of hospital baseline (2007) level controls (see table 7); share in each age bracket (5 age brackets), race/ethnicity (share white, share Hispanic, share black) and gender (share female). In a second model we add the hospital baseline (2007) level share of patients with pneumonia comorbidities (tertiles), share of patients with diabetes comorbidities (tertiles), share dual eligible (in tertiles), and number of beds (in tertiles). Hospitals with less than 50 cases throughout the three - year performance period (June 2008-July 2011, i.e., 2009-2011) are excluded. Coefficient estimates show the effect of a 0.01 change in the excess readmission ratio and the effect is allowed to differ before and after the HRRP penalty threshold. * 0.05 < p-value<=0.10, ** p <= 0.05, *** p <= 0.01

Figure 1. Medicare Hospital 30-Day Readmission Rates for AMI, Heart Failure, Pneumonia and All Targeted Conditions between July 2006 to July 2014 by the Condition Specific Penalty Status in the First Year of the HRRP

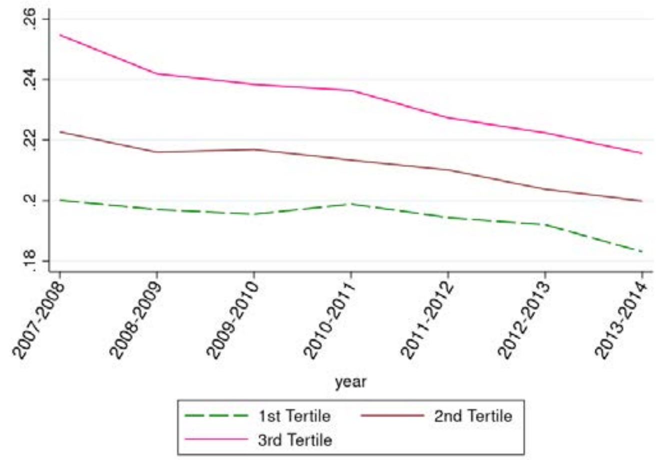
Notes – The unit of observation is the HRRP year (e.g.: July 2006 to July 2007). We plotted 30-day readmissions for AMI (top left), Heart Failure (top right), and Pneumonia (bottom right) and for all targeted conditions (bottom left) by the penalty status in the first round of the HRRP.

Figure 2. Medicare Hospital 30-Day Readmission Rates for AMI, Heart Failure, Pneumonia and All Targeted Conditions between July 2007 to July 2014 by the Baseline Condition Specific Readmissions Tertile in year 2006-2007

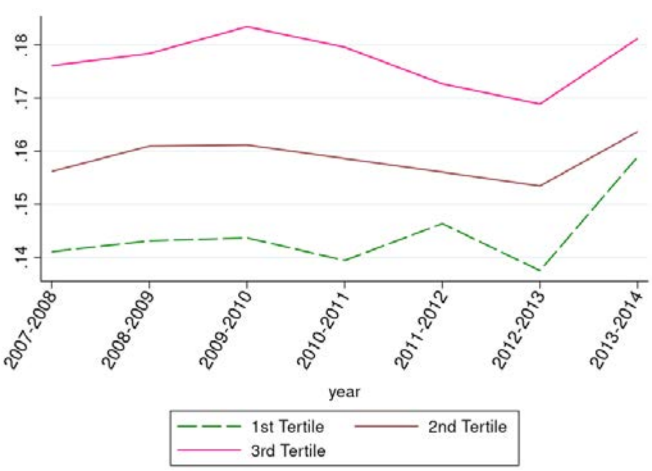
AMI 30-Day Readmission Rates



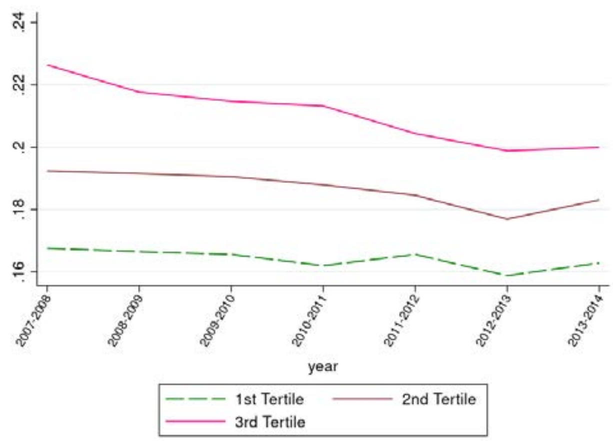
Heart Failure 30-Day Readmission Rates



Pneumonia 30-Day Readmission Rates

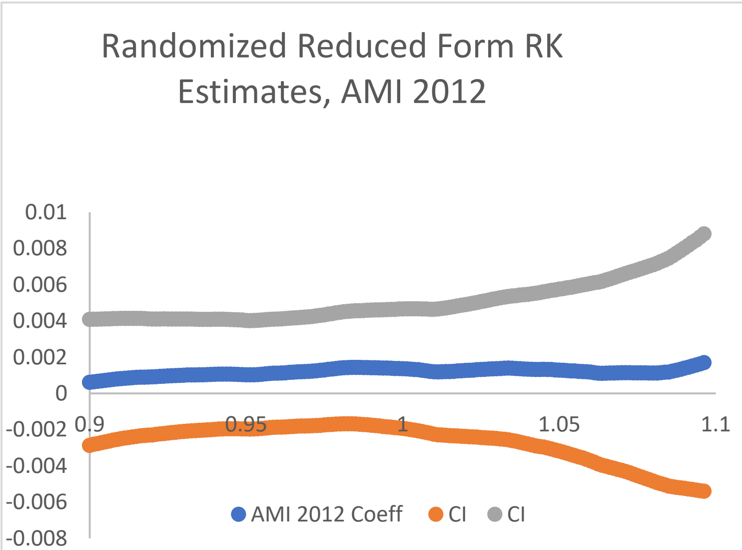


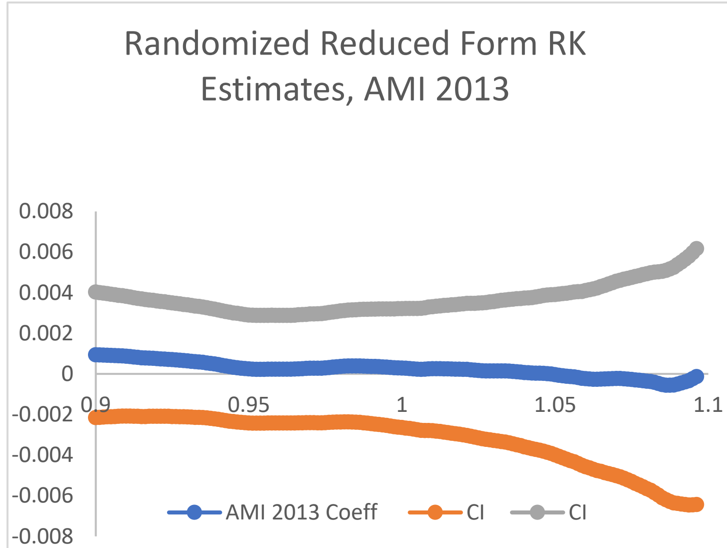
All Targeted Conditions 30-Day Readmission Rates

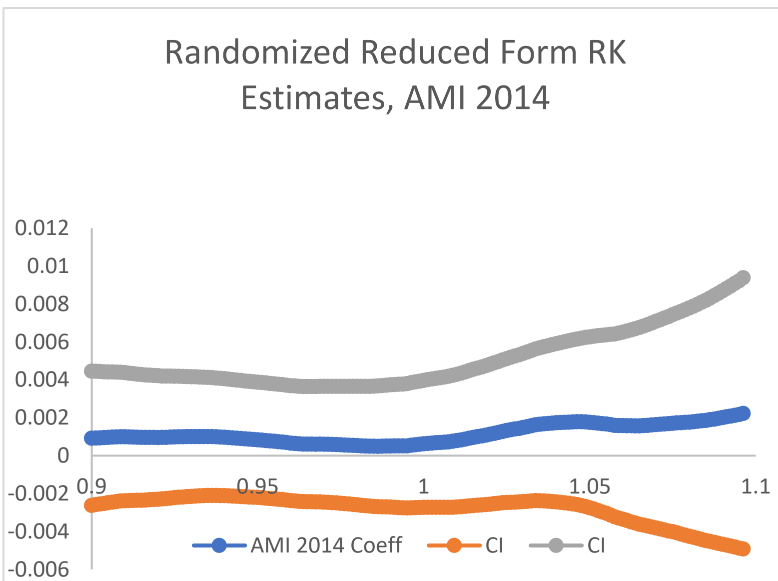


Notes – The unit of observation is the HRRP year (e.g.: July 2007 to July 2008). We plotted 30-day readmissions for AMI (top left), Heart Failure (top right), and Pneumonia (bottom right) and for all targeted conditions (bottom left) by tertiles of the respective baseline 30-day readmissions in 2006-2007.

Figure 3. Randomized RK Estimates 2012-2014 for AMI

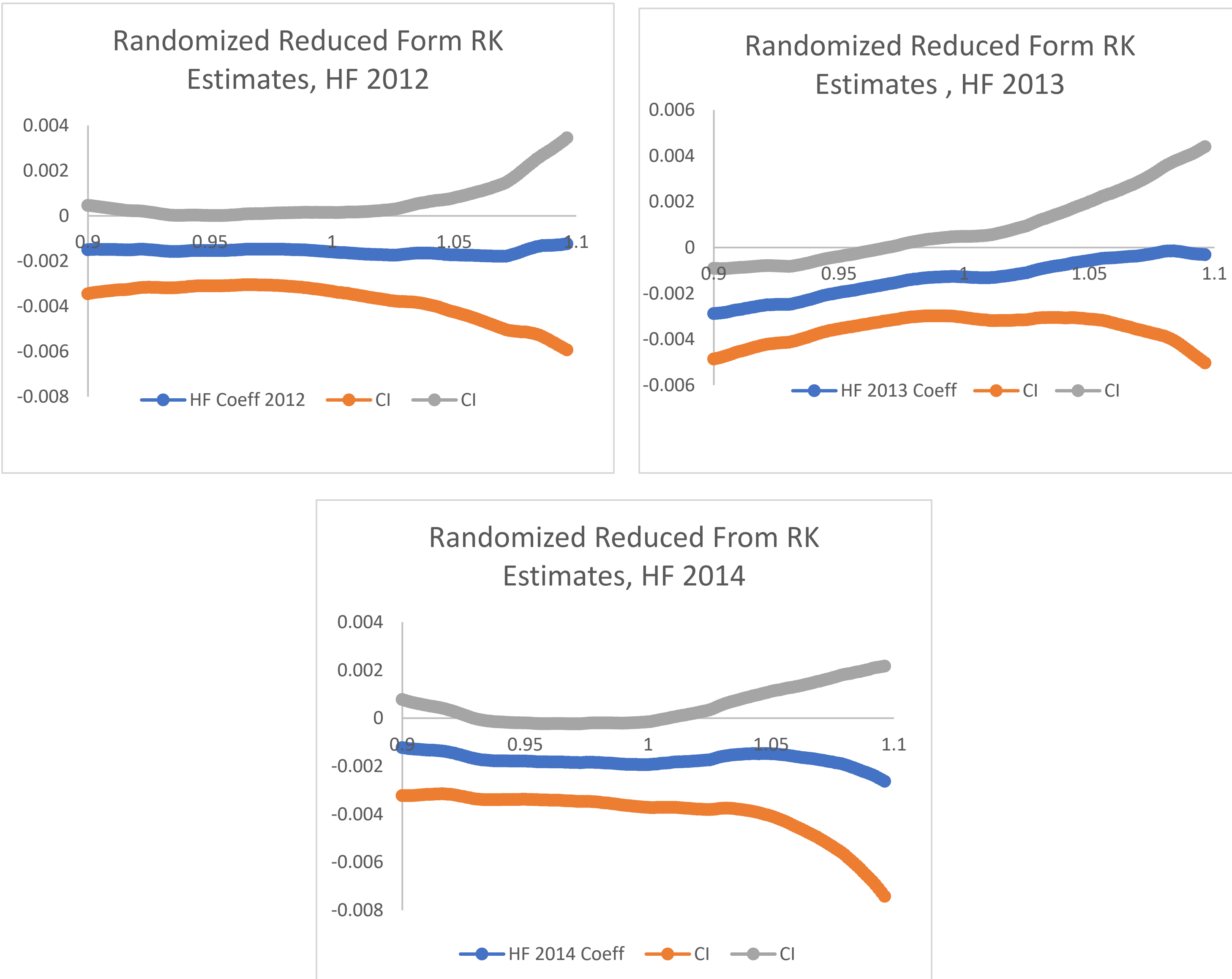






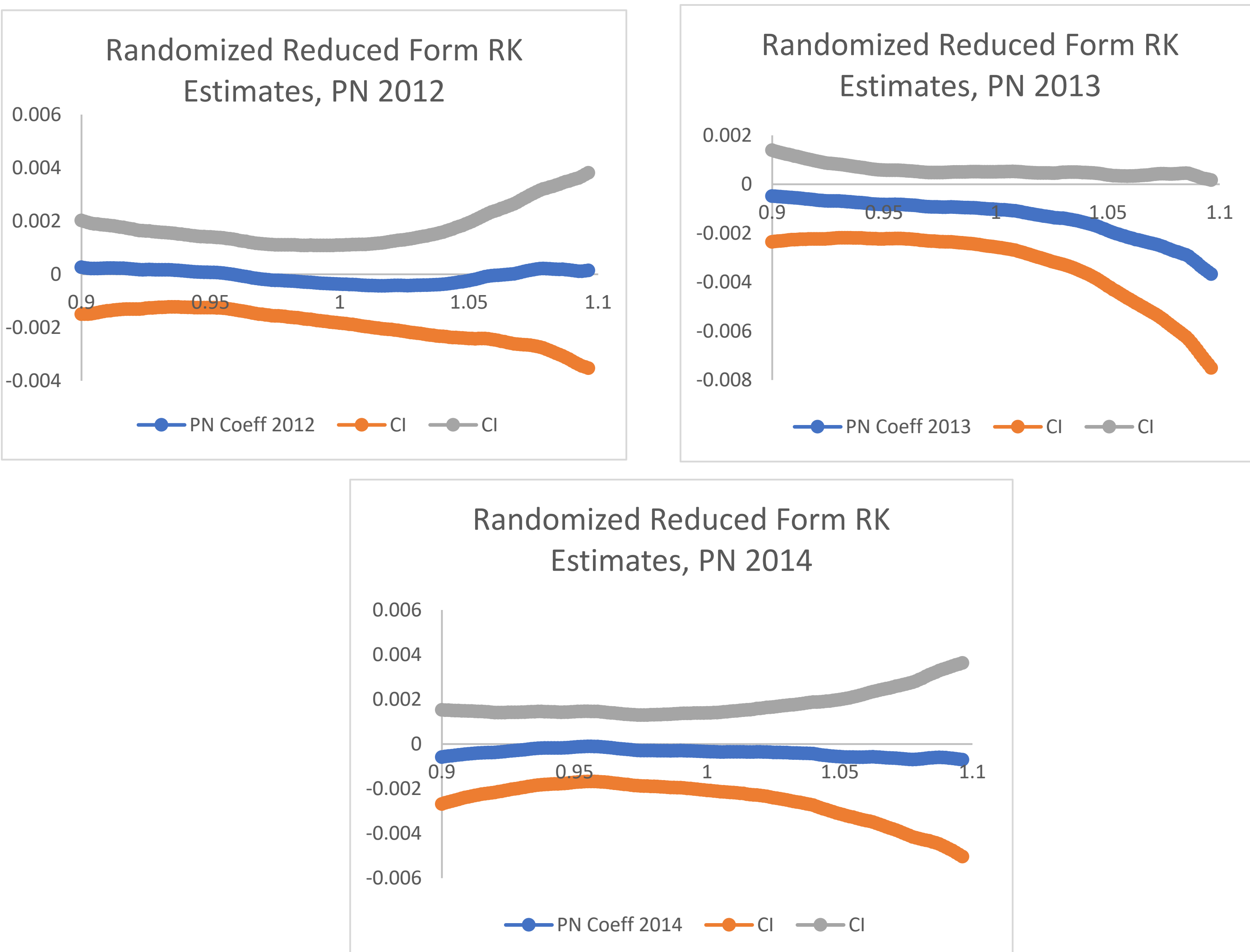
Notes: Each graph shows 158 estimates of the coefficient on the interaction between the running variable (Excess Readmissions Ratio) and the "kink" threshold for 158 randomly assigned thresholds between the values 0.9 to 1.096 of the excess readmissions ratio.

Figure 4. Randomized RK Estimates 2012-2014 for Heart Failure



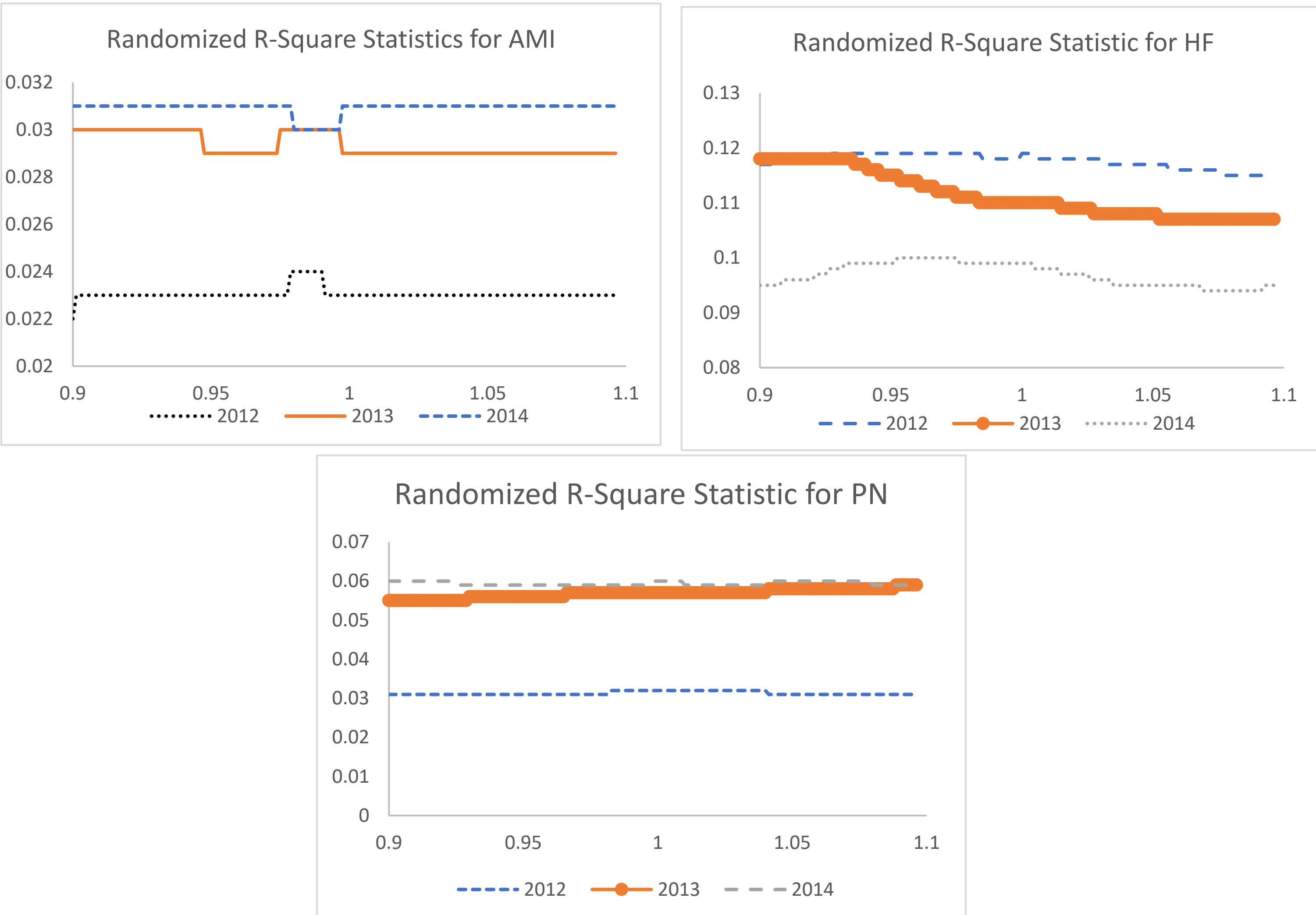
Notes: Each graph shows 158 estimates of the coefficient on the interaction between the running variable (Excess Readmissions Ratio) and the "kink" threshold for 158 randomly assigned thresholds between the values 0.9 to 1.096 of the excess readmissions ratio.

Figure 5. Randomized RK Estimates 2012-2014 for Pneumonia



Notes: Each graph shows 158 estimates of the coefficient on the interaction between the running variable (Excess Readmissions Ratio) and the "kink" threshold for 158 randomly assigned thresholds between the values 0.9 to 1.096 of the excess readmissions ratio.

Figure 6. R-square Statistics from Randomized RK Models, by Condition



Notes: Each graph shows 158 adjusted r-square estimates from the reduced form regression kink model for randomly assigned kink thresholds between values 0.9 to 1.096 of the excess readmissions ratio.